A Technical FAQ

Frequently Asked Questions About Voice and Video over IP Networks

January 2003

Saqib Jang, Margalla Communications

E. Brent Kelly, Wainhouse Research

Andrew W. Davis, Wainhouse Research

Wainhouse Research

112 Sumner Road
Brookline, MA 02146
www.wainhouse.com

1. What are the benefits of IP vs. ISDN for business-quality videoconferencing?

The business case for IP vs. ISDN-based videoconferencing spans quality, cost, management, efficiency, reliability, and scalability areas.

ISDN is usually inexpensive to own, but it is expensive to use. Besides an initial capital outlay to provision select conference rooms with ISDN connectivity, there are few additional costs required to begin videoconferencing using ISDN. A standard ISDN business-level videoconferencing call at 384 Kbps requires the bonding together of 6 ISDN channels; higher call speeds require the bonding of additional channels. Enterprises pay for ISDN on a per-minute-per-B-channel basis (often based on distance as well), making the use of the equipment costly. TV quality video at 768 Kbps on an ISDN system quickly becomes prohibitive in cost. These expensive ISDN usage fees often prohibit deep adoption of ISDN videoconferencing within an organization or enterprise..

The availability of flat-rate pricing for IP videoconferencing, on the other hand, allows calls at bandwidths too expensive for ISDN, including some IP calls up to 2 Mbps and beyond. These high bit rate calls enable higher quality audio and video communications. Because IP is so affordable and due to the pervasiveness of IP network connections, IP videoconferencing endpoints can be deployed across the enterprise economically. Furthermore, since IP systems do not bond channels together like ISDN system do, very high call reliability rivaling the POTS network can be achieved. ISDN has proven itself unreliable over the years due to the channel bonding problem: if one of the bonded channels is dropped during a call, often the entire call goes down. Most companies using ISDN are delighted to achieve a 92-94% success rate while those using IP videoconferencing often achieve greater than 99% reliability..

IP videoconferencing also permits significant management benefits. IP based video systems are always connected to the packet-switched network. This constant connectivity allows these systems to be remotely controlled and managed from a central, remote location. Large-scale conferencing environments often use an IP-based software product, called a gatekeeper, to control and track the usage of their videoconferencing systems, enabling improved measurement of ROI and convenient billing mechanisms.

One of the primary advantages of deploying IP based videoconferencing is the ability to use an organization's existing data network as the means of transport. This is called "converged networking". Converged networking can result in both cost savings and efficiency enhancements because only one network is deployed, maintained, and managed. Furthermore, since IP connections are already nearly everywhere – to every enterprise conference room and to every enterprise desktop, scaling voice and video over IP applications is easy because the network is already deployed, debugged, up and running. ISDN requires a separate network infrastructure and a separate management team, and will usually be limited to niche deployments within the enterprise.

2. What network protocols are used for IP videoconferencing services?

The two most important protocols are H.323 and Session Initiation Protocol (SIP).

H.323. H.323 is an ITU umbrella standard describing a family of protocols used to perform call control for multimedia communication on packet networks. The most important protocols used to set up, manage, and tear down calls are H.225 and H.245. H.225 is used to perform call control, and H.245 is used to perform call management.

In the most basic use of H.323v1 to set up a call, an endpoint initiates an H.225 exchange on a TCP well-known port with another endpoint. This exchange uses the Q.931 signaling protocol. Once a call has been established using Q.931 procedures, the H.245 call management phase of the call is begun. H.245 negotiations take place on a separate channel from the one used for H.225 call setup (although with the use of H.245 tunneling, H.245 messages can be encapsulated in Q.931 messages on existing H.225 channels), and the H.245 channel is dynamically allocated during the H.225 phase. The port number to be used for H.245 negotiation is not known in advance. The media channels (those used to transport voice and video) are similarly dynamically allocated, this time using the H.245 OpenLogicalChannel procedure.

Note that H.245 channels are unidirectional. In a minimal situation with direct call signaling between endpoints and the use of one bi-directional voice channel, for each call there will be a minimum of five channels (one H.225 channel, one H.245 channel, and one shared voice channel). Three of these will be on dynamically allocated ports. Business-quality IP video communication between two H.323 end-points typically requires in excess of 380 kbps data rates for each unidirectional media channel or aggregate data rates of over 750 kbps.

SIP. Session Initiation Protocol (SIP) is an Internet Multimedia Architecture established by the Internet Engineering Task Force (IETF – www.ietf.org).., SIP may be used for Voice over IP (VoIP), video conferencing, instant messaging, and is being planned for use in 3G wireless applications as well as new converged data and voice applications. It is an application layer signaling protocol used to establish, modify, and terminate multimedia sessions and is part of the. SIP applications include voice, video, gaming, instant messaging, presence, call control, etc.

In the spirit of other Internet based applications, SIP relies on a number of other computer communications standards including Session Description Protocol (SDP), Real-Time Protocol (RTP), TCP, UDP, and so on. SIP messages are based on HTTP protocol and have a similar text-based structure.

SIP uses Uniform Resource Indicators (URIs) which are a more general form of world-wide web Uniform Resource Locators (URLs). There are a number of URI forms, including user@domain, domain, user@ip-addr, telephone-number@domain. SIP messages can also use other URIs, such as the telephone URL (as defined in IETF RFC2806)

Generally, the SIP components are defined as user agents, proxies, redirect servers, and registrars: user agents are much like an endpoint in H.323 and may be telephones, video units, PDAs, etc. SIP communicates between these four components using a request – response data model. Messages between components are initiated when one component sends a request message (called a method) to second component. Responses consist of a numerical code and a textual "reason". To initiate a session, one SIP device sends an "invite" message to another SIP device. SDP is carried in the SIP message to describe the media streams and RTP is used to exchange real-time media streams.

3. What are the main IP videoconferencing deployment issues?

The main technical issues surrounding IP videoconferencing deployment include the following:

- **Quality of Service**. Quality of service (QoS) is a network term that specifies a guaranteed throughput level. In layman's terms, it means that the network must be designed so that the voice and video data are transmitted through the network with a minimum of delay and loss. The network must be carefully evaluated to insure that it will be able to transmit voice and video data properly. Often components in the network must be upgraded or additional routers, switches, or "packet shaping" devices may be required.

- **Overlay Network vs. Converged Network Architecture**. Enterprises may not want to put voice and video data in competition with mission critical data applications such as market or manufacturing data running across the same network. Consequently, a separate QoS enabled "overlay" network may be deployed for voice and video applications.

- **Security**. Most enterprise networks employ firewalls and network address translation (NAT) in an effort to prevent hackers or unauthorized persons to have access to the data one network. Voice and video over IP are not NAT and firewall friendly. Organizations will need to consider how they will securely traverse the corporate firewalls or NAT system be modified, re-configured, or upgraded to allow IP-based videoconferencing traffic?

- **Bandwidth Over the WAN**. IP data connections must be available at the locations where the enterprise needs to use video. These will typically be available from service providers (or

telephone companies) in the locations of interest; however, if they are not, organizations will need to consider what other alternatives are available and what bandwidth will be required. In general, satellites communications will not offer sufficient quality of service for IP videoconferencing due to excessive latency.

- **Multipoint Bridging Capability**. Organizations will need to consider whether more than two parties will need to participate in a video call. If so, some type of video multipoint bridging capability will be necessary. The MCU may be purchased and managed internally or all bridging functions may be outsourced to a service provider. If an internal MCU is to be purchased, then the magnitude of the initial investment increases and internal staff will need to be allocated to manage the video bridge.

- **IP-ISDN Gateway Needs**. As the transition to IP will not be complete for some years to come, organizations using IP video systems will likely need to communicate others using ISDN. Organizations will need to consider how many gateways are needed and how many ISDN lines should be provisioned for each. Rather than owning the gateway with its associated capital outlay and management costs, an enterprise may utilize a gateway owned and managed by a service provider.

- **Additional IT Resource Requirements**: Network maintenance and support resources must be willing, capable, and available to support a converged network carrying data, video, and voice traffic? Organizations should consider if additional resources will be required to manage new video-centric devices on the network.

4. What are quality-of-service (QoS) requirements for IP-based voice and video?

Real-time IP applications, such as videoconferencing and voice-over-IP are much more sensitive to network quality of service vis-à-vis store-and-forward-type of data applications, such as e-mail and file transfer. Quality of Service (QoS) refers to intelligence in the network to grant appropriate network performance to satisfy an application's requirements. For multimedia over IP networks, the goal is to preserve both the mission-critical data in the presence of multimedia voice and video and to preserve the voice and video quality in the presence of bursty data traffic.

Four parameters are generally used to describe quality of service: latency or delay, the amount of time it takes a packet to transverse the network; jitter, the variation in delay from packet to packet; bandwidth, the data rate that can be supported on the network; and packet loss, the per cent of packets that do not make it to their destination for various reasons.

- *End-to-end latency*. End-to-end latency refers to the total transit time for packets in a data stream to arrive at the remote endpoint. The upper bound for latency for H.323 voice and video packets should not be more than 125-150 milliseconds. The average packet size for video packets is usually large (800-1500 bytes) while audio packet sizes are generally small (480 bytes or less). This means that the average latency for an audio packet may be less than that for a video packet as intervening routers/switches typically prioritize smaller over larger packets when encountering network congestion. In addition, an H.323 video call actually represents four streams – each station sends and receives audio and video. The difference in latency of the streams will manifest itself as additional delay (both H.323 and SIP convey sufficient information to lip-synch the various streams).
- *Jitter or variability of delay.* This refers to the variability of latencies for packets within a given data stream and should not exceed 20 - 50 milliseconds. An example would be a data stream in a 30 FPS H.323 session that has an average transit time of 115 milliseconds. If a single packet encountered a jitter of 145 milliseconds or more (relative to a prior packet), an underun condition may occur at the receiving endpoint, potentially causing either blocky, jerky video or undesirable audio. Too much jitter can cause inter-stream latencies which as discussed next.

- *Inter-stream latency*. This refers to the relative latencies that can be encountered between the audio and video data streams and is based on how the relative average transit time for the given streams, at any given point, vary from each other. In this case the relative latency variations are not symmetrical. This is due to the fact that the human brain already compensates for audio latency relative to video. Due to this fact, an audio stream that starts arriving at an endpoint 30 milliseconds ahead of its video stream counterpart(s) will produce detectable lip-synchronization problems for most participants. An audio stream that arrives later than its associated video stream data has a slightly higher tolerance of 40 milliseconds before the loss of audio and video synchronization becomes generally detectable.

- *Packet loss.* This term refers to the loss or desequencing of data packets in a real-time audio/video data stream. A packet loss rate of 1% produces roughly a loss of one fast video update per second for a video stream producing jerky video. Lost audio packets produce choppy, broken audio. Since audio operates with smaller packets at a lower bandwidth, in general, it is usually less likely to encounter packet loss, but an audio stream is not immune from the effects of packet loss. A 2% packet loss rate starts to render the video stream generally unusable, though audio may be minimally acceptable. Consistent packet loss above 2% is definitely unacceptable for H.323 videoconferencing unless some type of packet loss correction algorithm is used between the endpoints. Packet loss in the 1-2% should still be considered a poor network environment and the cause of this type of consistent, significant packet loss should be resolved.

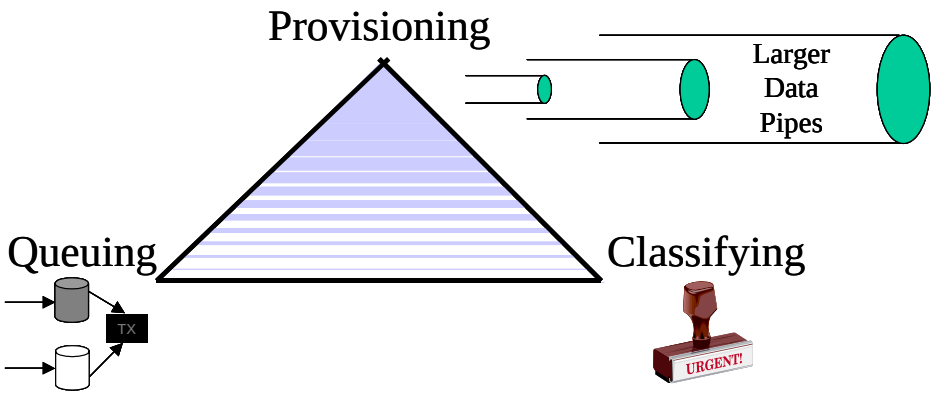

Figure 1 Three tools for network quality of service

Three types of tools or solutions are available to the network engineer to build quality of service into the network system. 1) Provisioning means providing adequate bandwidth for all voice, video, and data applications that traverse a common network. By using a 100 Mbps Ethernet network for example instead of a 10Mbps network, the network is more likely to support multimedia traffic together with data. Note that IP networks typically have significant packet overhead. For example, a 384 Kbps video call actually requires about 10% additional bandwidth for IP overhead; furthermore, when going from IP to ATM or frame relay, an additional 10% of the call bandwidth should be allocated for encapsulation. Hence, a 384 Kbps IP call traversing an ATM backbone may required as much as 460 Kbps of bandwidth. 2) Classifying means giving packets a classification based on their priority. Voice packets would be given the highest priority since they are very delay and jitter sensitive, even though they are not particularly bandwidth challenging. Video packets might be given a slightly lower priority; and email packets, for example, given the lowest priority. There are many different classification schemes possible, including some that are in the process of being standardized. One common scheme is to give VoIP packets an IP precedence of of 5 and videoconferencing applications an IP precedence of 4. 3) Queuing refers to a process that takes place in the routers and switches whereby different queues or buffers are established for the different packet classifications. One of the buffers, for example, might be a delay and drop sensitive buffer designed to handle voice and/or video packets. Many queuing schemes are available for implementation.

Solving QoS over IP networks for multimedia conferencing is a two-phase problem:

1. Guarantee QoS within a specific, controlled enterprise intranet or service provider network.
2. Guarantee QoS across the hand-off (peering) points between the networks. The public Internet presents this second challenge to the extreme.

Four major QoS initiatives are RSVP (resource ReSerVation Protocol), IP Precedence, and Differentiated Services (DiffServ) from the IETF, and 802.1p from the IEEE.

Improved quality of service through use of standard mechanisms, such as DiffServ and MPLS, is the key factor behind the promise of broad-based use of interactive business-quality IP video. The underlying requirement is for the IP video infrastructure to enable end-to-end prioritized processing and delivery of video traffic between subscriber networks and carrier core networks. This requires prioritized treatment of video traffic over the "last mile" access network through the metro network through the carrier networks.

While DiffServ is gaining broad support to enable "soft" QoS through prioritization of processing of traffic by service provider routers, MPLS is being deployed in service provider networks as an adjunct to enable the fine-grained delivery of a number of value-added services.

5. Why do firewalls cause problems with H.323 and SIP?

Because of the heavy use of dynamically allocated ports within H.323 and SIP, it is not possible to preconfigure firewalls to allow multimedia traffic without opening up large numbers of ports in the firewall. Microsoft's web site has a page on configuring firewalls for use with NetMeeting, which is H.323-based, and they recommend this: "To establish outbound NetMeeting connections through a firewall, the firewall must be configured to do the following:

- Pass through primary TCP connections on ports 389, 522, 1503, 1720, and 1731.
- Pass through secondary TCP and UDP connections on dynamically assigned ports (1024-65535)."

Needless to say, this represents a somewhat more lax firewall policy than would be acceptable at many sites, and it does not address the problem of receiving incoming calls.

Both SIP and H.323 use Real-Time Protocol (RTP) for media delivery while only H.323 uses a dynamic-port based call signaling protocol (H.245)[1]. The issue with firewall traversal arises because both H.245 and RTP do not use fixed layer 4 port numbers but rather can use any port in the range 1024 through 65534.

The second problem with firewalls for the H.323 and SIP protocols is that they carry interactive voice and video, which are very sensitive to delay. The traditional network security approach to the dynamic port allocation problem is to use a H.323 or SIP application layer gateway (ALG) or proxy. This is a software component of a firewall that actually takes part in the protocol. In the H.323 context, a proxy would take part in the H.323 conversations, terminating the call on the firewall and creating a second call to the final destination, and finally plug-boarding the two calls together. The enterprise firewall with an H.323 proxy would then handle all H.323 call setup/teardown work, as well as moving all the video traffic for all the endpoints attempting to communicate across the firewall. This presents significant security and performance/scalability challenges for both H.323 as well as all data traffic traversing the enterprise firewall.

- H.323 and SIP support within the firewall (which is in addition to the traditional role of firewalls in securing common protocols such as HTTP and FTP) makes firewall design complex and (by definition) more vulnerable to attack.
- H.323 and SIP support (especially as individual video sessions move to broadband 1Mbps+ business quality data rates) has the potential of degrading overall firewall performance and scalability.

[1] There is no equivalent to H.245 in SIP.

6. What is the NAT problem with H.323 and SIP?

Network Address Translation (NAT) is a method by which IP addresses are mapped from one realm to another, in an attempt to provide transparent routing to hosts. Traditionally, NAT devices are used to connect an isolated address realm with private unregistered addresses to an external realm with globally unique registered addresses. NAT is generally used for two purposes: 1) as a mechanism to work around the problem of IPv4 address space depletion, and 2) for security purposes (to hide hosts at an unroutable address). NAT works by having a NAT device, often implemented as part of a firewall application, rewrite IP addresses in the packet header as packets pass through the NAT. The NAT maintains a table of mappings between IP addresses and port numbers.

The problem with NAT from an H.323 perspective is that H.225 and H.245 make heavy use of embedded IP addresses. If NAT is being used, addresses in the protocol stream will be the addresses in the private address space (behind the NAT), rather than the public address allocated by the NAT. For example, a host may have its address in a private address space, 172.16.0.81, which when traversing a NAT is translated to 207.127.234.239. When that host attempts to place a call, the "calling party" information element in the H.225 signaling stream will contain the private, non-routable address (172.16.0.81), and attempts to make an H.225 connection back to that address will fail.

Because SIP signaling messages include IP addresses within the data segment of IP packets, NAT devices will also break SIP unless they are somehow made “SIP aware”. Compounding the problem is the fact that the source and destination of SIP signaling messages may not have any direct relationship to the source and destination of media streams.

7. What NAT/Firewall solutions are available today?

One obvious way for organizations to overcome the Firewall/NAT problem is to avoid using them. For most organizations, the security risks of this solution are too great; furthermore, obtaining enough routable IP addresses for the entire organization may prove difficult and expensive. There are, however, a number of organizations, particularly among educational institutions, that have little firewall protection, and they often do not use NAT.

Rather than have any concern with employing IP communications outside the LAN, organizations can use a gateway to convert from IP voice and video on the LAN to PSTN voice and video over the public circuit-switched network. Use of a gateway eliminates the concern for network firewall traversal because no data packets cross the firewall. It also overcomes the NAT issue because all calls made to endpoints on the LAN are routable, and calls coming into the LAN through the gateway are routable. Today most IP telephones use a gateway to communicate with non-IP telephones both within and without an organization. Gateway approaches are local solutions, however, that require all locations participating in the call to have a corresponding gateway behind the last layer of NAT and firewall they have deployed. Using PSTN gateways also removes the converged network cost savings and mobile use benefits an all-IP solution provides.

SIP or H.323 proxies[2] can be used to negotiate the NAT or both the NAT and the firewall depending upon how they are configured. Proxies act like a gateway, but instead of converting from one IP communications protocol to another, the same protocol is used on both sides of the proxy. A proxy has knowledge of both the public and private IP networks and makes the IP call effectively look like two separate calls: one from the endpoint in the private network to the proxy and a second call from the proxy to the endpoint in the public network. Internally, the proxy puts these two calls together thus resolving the NAT issue.

[2] The term “proxy” as used here refers to the protocol gateway element and not to the specific SIP address resolution component.

Proxies may have several different configurations in the network including being built into a gatekeeper or firewall, but they always require a gatekeeper for H.323 or a SIP registrar in order to resolve where to properly route the voice and video data packets. In some cases NAT is deployed at multiple locations along the network path: multiple points within the enterprise, and even within the external network at the ISP. For a proxy solution to work, it needs to be deployed at every NAT.

Application level gateways (ALG) are firewalls that are programmed to understand specific IP protocols, like H.323 and SIP. Rather than simply looking at packet header information to determine if packets can or cannot pass, ALGs go deeper by parsing the data in the packet payload. H.323 and SIP both put critical control information in the payload, such as which data ports the voice or video endpoint is expecting to use to receive the voice and video data from the other endpoint in the call. By understanding which ports need opening, the firewall dynamically opens only those ports needed by the application, leaving all others securely closed. This technique of opening small numbers of ports in the firewall dynamically is called "pinholing."

ALGs require a proxy if a NAT is being used to hide internal addresses. Some firewall manufacturers build the Proxy into the ALG, but it must be there to negotiate the NAT. As firewalls are mission-critical components for most enterprise networks, adding an ALG may prove difficult in some organizations because it requires both physical and political access to the firewall. ALGs are not consistently implemented from vendor to vendor; for example, at least one vendor's ALG does not allow T.120 data sharing. ALGs can also affect network performance, placing a heaver load on the firewall due to the parsing of the packet payloads. Moreover, if there are multiple levels of firewall/NAT combinations, each firewall/NAT in the call path must be upgraded to support ALG functionality.

Some organizations overcome NAT and firewall traversal issues by placing a multipoint control unit (MCU) in what is known as the demilitarized zone, or DMZ. The DMZ usually sits between the Internet and an internal network's firewall. Organizations that want to host their own Internet services, such as web servers, ftp servers, email servers, and domain name servers, without allowing unauthorized access to their private networks, place these servers in the DMZ. An MCU can be configured in the DMZ with two network interface cards such that one card provides access to the private network, and the other card gives access to the public Internet. One of the big disadvantages to this solution is that it uses ports on the MCU unnecessarily if the call is only point-to-point. As with a proxy or an ALG, if there are multiple NATs in the network, an MCU would need to be placed across each individual NAT. This solution does not scale very well either.

A semi-tunnel/transparent traversal method patented by Ridgeway Systems & Software borrows concepts from the proxy method discussed above to "funnel" data traffic through two "well known ports[3]" in the firewall. In this solution, a Ridgeway server in the DMZ plays the role of a full proxy. A Ridgeway client inside the firewall, in the private address space, also acts as a proxy in that it substitutes its own address and port numbers within packets sent to and received from the endpoints.

When the client starts it creates a single connection to port 2776 on the server for control and status information. It then listens for H.323 gatekeeper registrations and inquiries or SIP proxy/registrar registrations and inquires. As an IP endpoint or a SIP agent boots up, it send registration information, typically consisting of a telephone number and/or an email address, to the gatekeeper or registrar through the client/server connection; the server then allocates each registering endpoint a unique port on the server IP address and registers that endpoint with the gatekeeper. When an endpoint or user agent makes a call to another endpoint or user agent outside the firewall, all data packets are routed through the client to the server and from the server through the client back to the endpoint or user agent. As the call is setup, the client insures that all needed voice and video connections through the firewall are opened in an outbound direction. Voice and video data then flow in both directions through the firewall via these open ports.

[3] These ports, numbers 2776 and 2777 have been assigned to Ridgeway Systems by the Internet Assigned Numbers Authority (www.iana.org).

8. What are emerging standards for media protocol NAT/firewall traversal?

Middlebox Communication (MIDCOM)[4] is an emerging scheme very similar to the ALG method; however, in the MIDCOM approach to firewall and NAT traversal, protocol intelligence is not built into the firewall. Instead, the SIP or H.323 protocol knowledge is built into a different device, called a "midcom agent," that tells the firewall which ports to open for a given voice or video call. The primary advantage of this technique, in principle, is that firewalls do not have to be continually upgraded as protocols change or as they come in and out of fashion. Other advantages include scalability and reliability. Its disadvantages are similar to those of ALG. Additionally, the firewall would require an initial "forklift" upgrade to implement the MIDCOM strategy. This method is still under development in an Internet Engineering Task Force working group.

Another solution for NAT traversal is Simple Traversal of UDP through NAT[5] (STUN), and is also being standardized in the IETF midcom working group. STUN allows certain UDP-based applications, such as video and audio, to traverse certain types of NATs. Unlike many of the other solutions discussed here, STUN does not require an enterprise to deploy additional servers, nor does it require changes to the firewall or NAT. Unlike the Ridgeway solution, it does not require media to be routed through network intermediaries, increasing latency. In STUN, the client sends a message to a STUN server, resident on the public network, to learn a publicly routable IP address and port it can use to receive media. STUN servers are only slightly more complicated than echo servers and they scale well.

The drawback to STUN is its limited applicability. It only works for UDP, and it only works for certain types of NATs, known as "full-cone" NAT. Although most residential NATs (also known as home routers, cable modem routers, etc., and made by manufacturers like Netgear and Linksys) are full-cone, many enterprise NATs are not, so STUN may not work in these cases. STUN also requires support in the clients, and few endpoints support STUN today..

9. What are enterprise network policy-related challenges for H.323-and SIP-based video usage?

Use of high data rate applications, such as H.323/SIP-based business-quality video, has the potential of significantly impacting available LAN capacity for data traffic. Even with the on-going migration of corporate LANs to gigabit backbones and 100BT switched subnets, uncontrolled usage of interactive video services has the potential of severely reducing response times for business applications. Thus, the H.323 video delivery infrastructure is required to permit corporations to implement fine-grained controls regarding who can use interactive video and under what conditions. Specifically, corporations require the ability to control:

- who (by user and IP address) can use IP videoconferencing services
- can specific users and end-points only receive and/or initiate calls
- what types of codecs can specific end-points/users use for calls
- maximum aggregate video traffic throughput coming into/exiting an enterprise network
- the above by time of day

10. What are the potential pitfalls of managed (i.e. CSP-based) H.323 security (i.e. firewall, NAT, and policy) services?

The complex security, connectivity, performance, and policy implications of inter-enterprise H.323-based interactive video communication (as discussed above) make it very attractive for H.323 security services to be offered as a component of managed service provider-based H.323 video services. The benefit for corporations when using a managed H.323 services model is that they can off-load complex service deployment issues to conferencing service providers (CSPs), while the benefit for CSPs is that they can offer higher-margin, differentiated H.323 video services. One option that CSPs have for implementing

[4] http://www.ietf.org/html.charters/midcom-charter.html
[5] The latest STUN protocol draft is at http://www.ietf.org/internet-drafts/draft-ietf-midcom-stun-05.txt

such services is to utilize CPE firewalls, NATs, and policy managers for offering such services. Although the CPE model offers incremental revenue opportunities, it has the following deficiencies that constrain CSPs' profit margin.

- H.323 video services service rollouts require CSPs to deploy specialized equipment at every subscriber site, creating an expensive up-front capital investment as well as significant operational expenses. This results in service delivery delays, increased customer start-up costs and decreased SP margins.
- Every extra site leads to an incremental growth in the capital equipment pool because these resources are not shared, quite unlike traditional network transport equipment environments.
- Each new site adds another element that needs to be remotely managed.

One intermediate type of managed H.323 video services offering that could be deployed as a consequence of the above is to fashion H.323 security services into a hub and spoke configuration. In other words, CPE H.323 firewall/NAT/policy services are shared centrally at a regional office or headquarters (hub) level by a number of remote sites (spokes) that backhaul their IP video traffic. In this approach, only the hub sites have direct Internet access for video communication; the rest require one or two hops to ultimately get out to the Internet. While enterprise subscribers may be able to live with this hub and spoke solution for non real-time data traffic, such a configuration will not provide the performance required for business-quality inter-enterprise IP video communication. Moreover, subscribers expect secure direct interactive video communication paths for the extranet initiatives they are focused on developing with their strategic partners.

11. What are the potential issues with H.323 *gatekeeper* deployment?

H.323 gatekeepers are the heart and soul of a serious IP videoconferencing implementation. Gatekeepers have two main functions: the first is to enable host resolution by maintaining a mapping of users' nick names or telephone numbers and their IP addresses. The second is bandwidth management, which allows an IT manager to put a cap on the amount of network bandwidth available for H.323 conferencing. When implementing an H.323 system, the following issues with gatekeeper use will need to be addressed:

- A gatekeeper with its equipment and management costs is required at every location where there needs to be a usage control point.
- It is the responsibility of the system administrator to keep the list of peer gatekeepers (required for inter-zone communication) up to date. Newer gatekeeper mechanism, however, simplify this task.
- Calls to external enterprises may require the fixed public IP address of the gatekeeper and the endpoint naming convention to be shared with other organizations. This may be a security risk if the gatekeeper is running on a router-class system.

Acknowledgement
We appreciate Jonathan Rosenberg's (DynamicSoft) review of this FAQ. Jonathan is the former chair of the IETF SIP Working Group and is the co-author of the SIP protocol standard and is co-author of the Midcom STUN protocol draft.