A POLYCOM WHITEPAPER

Unified Communications Drive Protocol Convergence

November 2011

POLYCOM®

Introduction

With the emergence of the Unified Communications[1] (UC) concept, enterprises, service providers and other organizations started morphing their voice, video, instant messaging, and presence systems into one. This trend has created an interesting technical challenge.

Video networks today are mostly based on the ITU-T H.323[2] protocol while many of the network elements also support the Session Initiation Protocol (SIP).[3]

Telephony call control servers have started the migration from proprietary protocols to standard SIP, and there are already a large number of SIP standard-based implementations, some of them open source. Even the remaining proprietary IP-PBX systems on the market provide some level of SIP interoperability and allow third-party equipment to connect to the IP-PBX.

Many presence and instant messaging systems support SIP via the SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE)[4] protocol.

The technical challenge that UC poses is how to connect all of these elements into one system that provides the full range of services to users. Based on the current state of the networking technology, SIP is the most functional common denominator that could interconnect the different applications within the organization. SIP also meets the requirements for scalable distributed visual communications, and has already been deployed in certain scenarios.

What are the similarities and differences between SIP and H.323? Which UC scenarios require the use of SIP and when is H.323 more appropriate? This paper identifies the environments in which the use of each protocol is practical, and then compares H.323 and SIP, as they apply to visual communication. Since in many deployments there will be a mix between SIP and H.323, the paper discusses interoperability between SIP and H.323 and identifies approaches for smoother migration.

How Are Unified Communications Driving Protocol Convergence?

The term Unified Communications (UC) refers to a trend in business to simplify and integrate all forms of communications. And while different vendors and analysts apply the definition quite differently, several functions are always mentioned in discussions about UC: voice over IP, video desktop and conferencing, instant messaging, presence, and central directory.

Visual communication, including telepresence and video conferencing, has been a stand-alone application for years. With the migration from ISDN to IP networks, visual communication started using the same IP network resources as other applications—e-mail, Web, Voice over IP—but video continued to be an overlay application. Video also required separate management tools and directories, fully independent from the rest of the IT infrastructure and, in general, hardly connected to it.

Then the UC concept emerged, and enterprises, service providers, and other organizations started morphing their voice, video, and data communication systems into one. Figure 1 is a graphical representation of this trend.

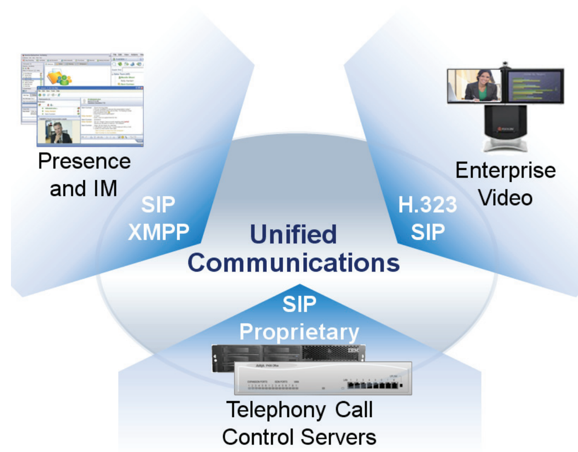

Figure 1: Trend towards Unified Collaboration

Telephony systems in the enterprise support proprietary protocols and SIP but generally not H.323. Instant messaging platforms are based on either XMPP[5] or SIP.

Enterprise video has traditionally used the H.323 protocol and only recently added SIP support in video network elements such as endpoints and conference servers. For example, Polycom® HDX™ endpoints have a dual protocol stack and can place and receive calls through both the H.323 and SIP protocol. Conference servers, such as the Polycom RMX 2000® and Polycom RMX™ 4000, support SIP, H.323, and H.320 protocols.

SIP is the most functional common denominator that could interconnect the different applications within the organization, and promises converting the currently separate systems into a UC solution. Therefore, we see increased support of SIP functionality in both telephony call control servers and in enterprise video.

Where Does Using SIP For Visual Communication Make Business Sense?

Let's look at several scenarios and identify the ones that benefit from using SIP instead of H.323.

Overlay SIP Visual Communication Network

SIP could be used to build an entire overlay visual communication network, similar to the ones based on H.323 today. SIP can deliver interoperability across products and vendors. The so called 'SIPit' events are a good place for interoperability testing of new products and there are efforts in the SIP community to create automated online test tools that allow testing of new software releases to make sure that interoperability is preserved.

So does it make sense to migrate enterprise video to SIP for the interoperability benefit? The answer is no; H.323 is a well-established protocol and interoperability across products and vendors is guaranteed. The International Multimedia Telecommunications Consortium (IMTC)[6] organizes regular interoperability tests, and all major players in the market participate. In particular, video-specific functions, such as content sharing and Far-End Camera Control, are very well understood in the H.323 community and very new to the SIP world.

Desktop Video

SIP is a scalable protocol and could be used for desktop video. A SIP-based desktop video solution must, however, interwork with the H.323 installed base and this requires a lot of additional infrastructure boxes in the network: signaling gateways, SIP servers, and a mix of video border proxies for both protocols. In addition, separate management mechanisms have to be deployed to accommodate the SIP part of the network.

Using H.323 for desktop video keeps the network simple. The soft clients become regular endpoints registered to the gatekeeper, equal to any room systems, which makes the transition from room video to desktop video very simple. Polycom CMATM Desktop is an H.323 soft client that is centrally managed by the CMA server. It delivers excellent audio and video quality while natively connecting to the H.323 video equipment in the enterprise. It also supports instant messaging and presence, and can integrate with the corporate directory through the CMA server.

SIP is becoming important in both the enterprise and service provider market, and is absolutely necessary for integrations with communication systems from Polycom partners. Therefore, the RealPresence Mobile soft client, released in October 2011, supports both H.323 and SIP and can be used in both unmanaged and managed (via CMA) environments. RealPresence Mobile is initially available for Apple iPad, Samsung Galaxy, and Motorola Xoom; it can be downloaded from Apple iTunes App Store and from Android Market.

Video-Enabled Hosted Communication

Hosted communication systems, such as BroadSoft BroadWorks, provide Centrex-like functionality in IP networks and usually have IP telephones as endpoints. When these systems were developed in the early 2000s, there was fierce competition between the functional SIP protocol and the stimulus MGCP[7] protocol in the hosted communication market. The difference between functional and stimulus protocols is discussed in the white paper "Scalable Infrastructure for Distributed Video."[8] SIP won the competition and is now supported in all platforms used in hosted communications.

Vendors in this space who are looking for ways to enrich the user experience and better compete with on-premise systems see video as an excellent way to achieve differentiation. Since H.323 was never actually deployed in the hosted communication space, SIP is the only standard for integration, and video endpoints and conference servers must support SIP to become part of the hosted communication solution.

However, hosted communication (a.k.a. IP Centrex) follows the PBX user interface paradigm (hold, transfer, park, pickup, multiple line appearances) while enterprise video developed around the concept of conferencing (meetings, rooms, participants). Merging the two, therefore, will not be easy, and existing video endpoints will need to borrow some functions from the telephony world to meet user expectations in this space.

For example, users of hosted communications expect the ability to configure lines (as in 'multiple line appearances') in their devices, which is a concept completely strange to the conferencing world. Therefore, a complementary approach to delivering video in the hosted communication space is one that will add video capabilities to IP telephones. For example, the Polycom VVX™ 1500 business media phone provides a user interface that is familiar to users of hosted communication, including the ability to configure lines.

Polycom works closely with vendors in the hosted communication space—most prominently BroadSoft—to integrate voice and video elements into hosted UC solutions.

Integration with Microsoft Lync Server 2010

Today, Lync is mostly deployed for instant messaging and presence, however, additional voice and video functionality makes it a compelling UC platform for organizations that standardize on Microsoft.

Lync is based on the SIP protocol, and any integration with Lync requires SIP support as well as authentication against the Microsoft network.

Polycom is working closely with Microsoft to integrate voice and video components in the Lync environment.

Integration with IBM Sametime

IBM Sametime is in a similar situation as Microsoft Lync. It is based on the SIP protocol and is well-positioned as an UC platform for organizations. While mostly used for instant messaging (IM) and presence, it is also moving towards adding advanced voice and video capabilities.

Polycom is working closely with IBM to integrate voice and video components into the Sametime environment.

Integration with IP-PBXs

Early versions of leading IP-PBX systems, such as Avaya Communication Manager, supported basic H.323 and allowed the registering of H.323 clients. However, as SIP gained momentum in the early 2000s, IP-PBX systems switched to SIP while H.323 support was completely dropped or was not updated to the latest H.323 versions. Since most IP-PBX systems in the market today support SIP (and do not support H.323), SIP is often the preferred way to integrate with IP-PBX systems.

Polycom works closely with vendors in this space to integrate voice and video components into their environments using the SIP protocol.

In summation, integration with UC platforms—both on-premise and hosted—always requires SIP while creating a complete overlay SIP network or using SIP for desktop video does not bring any business benefits. Therefore, both protocols will need to coexist for long time to come.

Next, we will make a "technical deep-dive" and compare the capabilities of H.323 and SIP, as they relate to visual communication.

The H.323 Protocol

The H.323 protocol was developed by the International Telecommunication Union (ITU)—an international standardization body based in Geneva, Switzerland. H.323 is an umbrella signaling protocol, that is, it refers to a set of other protocols, such as H.225 and H.245, which is known as "the H.323 family of protocols."

H.323 was originally defined for multimedia communications and perfectly fits the video conferencing application because it had from the very beginning mechanisms for audio and video call setup. It also has the so-called capability exchange procedure (often referred to as CAPS) that is very important for finding communication parameters acceptable for both communication sides, as well as a master-slave determination mechanism that is very useful when MCUs are involved in the communication.

H.323 is optimized for machine-to-machine communication. It uses ASN.1 notation/encoding, and the H.323 messages are encoded using the Basic Encoding Rules (BER). This means that very few people can actually read captured H.323 messages.

H.323 Elements and Call Flow

H.323 defines H.323 Terminals which can initiate or receive calls and H.323 Gatekeepers which register H.323 terminals, provide call admission control, and call routing. Gatekeepers can be very simple or very complex—depending on how many of the optional functions in H.323 they implement. H.323 also defines Gateways to other networks, for example, H.320/ISDN. While gateways are optional in H.323, they play an important role in multi-protocol environments, for instance, H.320-H.323 or H.323-SIP.

Figure 2 looks at the interaction of the two critical and mandatory elements in the H.323 network: Terminals and Gatekeeper.

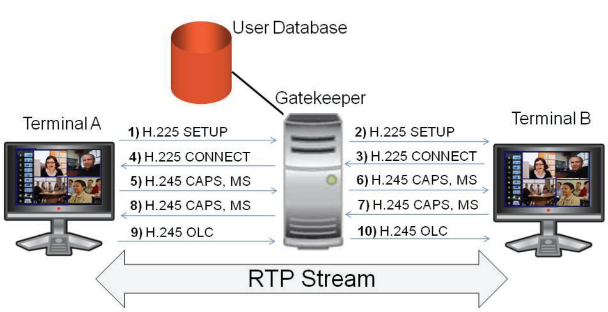

Figure 2: H.323 Basic Call Flow

H.323 describes the call setup procedure, and refers to the H.225 and H.245 protocols for signaling message formats and some additional functions. The signaling messages are described in H.225. The H.225 SETUP message includes information about the source, that is, who is sending the message (in Figure 2, this is Terminal A) and about the destination (Terminal B). The Gatekeeper then uses this information to allocate the destination (Terminal B).

After receiving the SETUP message, Terminal B stores the information about the request (IP addresses, port numbers, and so on), and sends back the CONNECT message. The most important information in the CONNECT message is about the setup of an H.245 control channel, which is used for three main functions: capability exchange (CAPS), master-slave determination (MS), and opening logical channels (OLC), that is, creating media streams for audio, video, and content.

H.245 Terminal Capability Exchange is a procedure for exchanging preferred audio and video codecs and settings between the two H.323 terminals. For example, Terminal A may suggest H.264 or H.263 video and Siren22 or G.722.1 audio, and Terminal B may respond that it only supports H.263 and G.722.1. Once both sides agree on common parameters the "conversation" moves to its next phase—H.245 Master Slave Determination—which is useful for avoiding conflicts during call control operations. H.245 Master Slave Determination is very important when an H.323 Terminal connects to an MCU (the MCU is the master), and when one MCU connects to another MCU through a so-called "cascading"—in this case one of the MCUs has to be the master.

After capabilities have been exchanged and connection master determined, the H.245 Open Logical Channel Request procedure creates media channels (voice, video, or content/data) between the communication parties. Note that these channels are always created in pairs, that is, the video channel from Terminal A to Terminal B is different and separate from the video channel from Terminal B to Terminal A. Therefore, communication can be asymmetric: Terminal A can send high quality video to B, and receive lower quality video from B, and vice versa.

H.245 control channel is also used to transmit the Flow Control command, which is used by the receiver to set an upper limit for the transmitter bit rate on any logical channel, and the Fast Update command, which is used by the receiver to request resending video frames that were lost in the transmission.

Audio streams and video streams are transmitted via the Real Time Protocol (RTP)[9]. For each RTP stream there is an associated Real Time Control Protocol (RTCP) channel which is used to periodically transmit control packets to participants in a multimedia session. The primary function of RTCP is to provide feedback on the quality of service being provided by RTP.

H.323 for Enterprise Video

H.323 has been widely deployed in visual communication equipment. The H.323 Terminal function is implemented in video endpoints such as Polycom HDX, QDX™, and CMA Desktop. The H.323 Gatekeeper function is implemented in communication servers such as Polycom CMA 5000. The H.323 MCU function is implemented in conference servers such as the Polycom RMX 1000, 2000, and 4000 servers.

H.323-based systems support the following important functions.

Multipoint conferencing is the ability to connect multiple participants in a conference call. Multipoint conferencing is very natural in H.323 because every call in H.323 (including point-to-point calls) is defined as a "conference." It is therefore assumed from the start that parties will be added to the conference.

DTMF tones are used for conference management during a multipoint call, for example, the user can enter DTMF tones on a phone or video endpoint and activate functions on the MCU during the conference. This capability is especially important for the "chairperson" of the conference. While there are better alternatives to control conferences on the MCU from a video endpoint, DTMF tones are still widely used because conference participants may be connecting via voice devices with very simple user interface.

Dual Video Streams allows a "presentation" (sometimes also called "content") audio-video stream to be created in parallel to the primary "live" audio-video stream. This second stream is used to share any type of content: slides, spreadsheets, X-rays, video clips, to name a few. ITU-T H.239 standardizes the functionality.

Video channel control is embedded in H.245. The protocol allows sending messages such as "Flow Control" from the receiver of live and presentation streams back to the sender of these streams, and telling the sender to modify the bit rate, usually to reduce the bit rate when the receiver detects high packet loss. By sending a "Fast Update" message the receiver asks the sender to resend a full or intra video frame(s), usually when a video frame is lost in transmission.

Far End Camera Control (FECC) allows the near end of the video call to control the camera at the far end: zoom, pan (move the camera left and right), and tilt (move the camera up and down). FECC is implemented through two ITU standards: H.281 defines the binary data that is transmitted between the two terminals to control the camera while H.224 defines the format of the frames that carry the binary data.

H.323 has its own set of security mechanisms. Early implementations used DES and 3DES encryption, while the latest generation of equipment supports the state-of-the art Advanced Encryption Standard (AES) adopted by the US Government and many security-conscious organizations.

H.323 also includes mechanisms for traversing firewalls and Network Address Translation (NAT). They are described in H.460.17, H.460.18, and H.460.19 standards. Firewalls and NATs are major hurdles to audio and video communication in IP networks because they block traffic from passing from one IP network to another.

The SIP Protocol

SIP was developed by the Internet Engineering Task Force (IETF), an organization that sets the technical standards for the Internet. In many ways SIP is similar to H.323: it also can be used to setup audio and video calls, and it also refers to a long list of other standards (called "Request for Comments" or "RFCs" in the IETF lingo) that constitute "the SIP family of protocols." For example, SIP refers to the Session Description Protocol (SDP)[10] as format for describing media parameters.

IETF envisioned SIP to be generic protocol that can setup any kind of session, not just audio and video, for example, SIP can be used for instant messaging, data transfer, and so on. SIP was designed to be similar to the Hyper Text Transfer Protocol (HTTP)[11] which is used for Web browsing on the Internet. The idea was that HTTP developers should be able to easily learn the SIP protocol and develop Voice over IP and Video over IP applications, the same way they develop Web applications. While this did not exactly happen, SIP became easier to read and understand than H.323, mainly because it uses readable clear-text messages (by comparison, H.323 messages are ASN.1/BER encoded).

Since IETF develops standards for Internet, it is very concerned about the scalability of networking protocols. Therefore, SIP was designed to be lightweight and scale well. In fact, many of the SIP enthusiasts in the early SIP days criticized H.323 for it complexity, and it was an easy argument to make because SIP had just few specs. Over the years however, IETF created about 140 SIP specifications (RFCs) which increased the complexity for SIP implementations tremendously. There is a movement in IETF towards creating a "SIP core" that combines the most important 5-10 RFCs into a single spec which should be sufficient for a functional SIP implementation. Discussion continues which specs will make the list.

SIP Elements and Call Flow

The equivalent of H.323 Terminal in SIP is the SIP User Agent (UA). The name "user agent" leans towards mobile communication and user mobility, that is, the ability of the user to log on at a communication device which then becomes the user's agent.

Different from H.323, SIP splits the server functions (concentrated in the H.323 Gatekeeper) into several entities: SIP Redirect Server, SIP Proxy Server, and SIP Registrar. This is also in line with the Internet philosophy that the server that registers and authenticates you (the Registrar) does not need be the server that gets your requests (the Proxy) and does not need be the server that knows the current location of the destination (the Redirect Server).

Figure 3 shows the basic SIP message exchange necessary to setup an audio/video call.

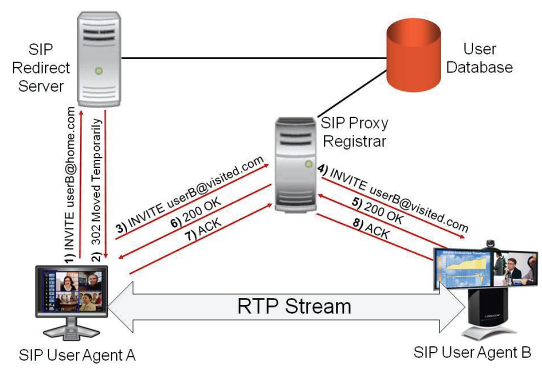

Figure 3: SIP Basic Call Flow

The UA's learn the SIP servers' addresses (Domain Name like www.sipregistrar1.com or IP address like 110.168.1.2) by configuration/provisioning or dynamically, that is, by sending a DNS SRV request asking the Internet "What SIP servers are there?" and receiving a list of servers.

Subsequently, UA's register with their home Registrars (registration procedure not shown here), and get authenticated, for example, the Registrar queries a user data base to verify user name, user password, and an additional authentication parameters called "SIP Realm."

While H.323 uses E.164 phone numbers (e.g. +14085551212) or aliases to identify the destination, SIP uses Unified Resource Identifier (URI) in the format user@<domain name>. In our example, user agent A is in the domain home.com and wants to reach "userB" which is currently in a different domain visited.com. A starts the session (call) by sending an INVITE message (the equivalent of a H.323 SETUP message) for userB@home.com to the local Redirect Server asking for the current location of "userB." The Redirect Server responds with error code 302 which means that the user has moved temporarily. SIP error codes are similar and often equivalent to the HTTP error codes that you see when there is an error in a Web application. The response includes the new domain of the user: visited.com.

A then sends a new INVITE to the local Proxy Server (for simplicity Proxy and Registrar are residing in the same server in Figure 3), and the Proxy server routes the INVITE through the network to the destination. A handshake procedure including the SIP messages 200OK and ACK makes sure both communicating partners and the proxy server know that the session is successfully setup.

Similar to H.323, the signaling procedure ends with the setup of media streams, e.g. for audio and video. As in H.323, audio streams and video streams are transmitted via the Real Time Protocol (RTP), and for each RTP stream there is an associated Real Time Control Protocol (RTCP) channel. The importance of the RTP/RTCP use in both H.323 and SIP will be highlighted later in the discussion around SIP-H.323 gateways.

SIP for Enterprise Video

As mentioned above, the H.323 community invested much effort adding new functionality to H.323 for the purposes of visual communication. SIP on the other hand was embraced by the Voice over IP community and extended in many ways to support voice communications. The challenge for the video industry today is to support some SIP-specific functions in video equipment and to map the video-specific functionality from H.323 to SIP.

Some of the functionality is easy to map. For example, H.323 systems deploy the Advanced Encryption Standard (AES)[12] for media encryption, that is, all RTP packets carrying audio and video are encrypted by the sender using AES. SIP refers to Secure Real Time Protocol (SRTP)[13] which lists AES as default cipher. Therefore, AES can be used in both H.323 and SIP networks.

Other functions, for example, in the area of firewall traversal, cannot be mapped. H.323 relies on H.460.17, H.460.18, and H.460.19 standards for firewall traversal. IETF originally developed the Simple Traversal of UDP through NATs (STUN)[14] mechanism, then added Traversal Using Relay NAT (TURN)[15] mechanism to increase the firewall traversal success rate, and finally created the Interactive Connectivity Establishment (ICE)[16] specification that combines STUN and TURN functions into one.

The Video Channel Control functions (Fast Update and Flow Control) are well-defined in H.323 and can be mapped in SIP using the RTCP Feedback function.[17]

Dual Video Streams is a standard video function and can be implemented in SIP by using the 'label' attribute18 and the "content" attribute[19] in SDP and by grouping the content stream with a live stream.[20]" The remaining issue is how to identify who is sending the content and who is receiving it. This is usually done by tokens (the party that has the token can send content), and token-management protocols ensure there is only one token in the session, and that anyone can request and receive the token. The RFC 458221 defines the token-management mechanism, and is used in the dual-video stream implementation in SIP."

Far End Camera Control information in H.224/H.281 format can be tunneled through the SIP network using the special RTP Payload Format for H.224.[21] (standard authored by Polycom).

How To Connect H.323 And SIP Networks?

Although we expect SIP deployments to grow rapidly in the future, the installed base of H.323 endpoints and infrastructure is here to stay in the healthcare, government, education, and general enterprise markets. Interworking between the two protocols becomes an important issue.

In general, there are three ways to bridge the SIP and H.323 networks: through dual-stack endpoints, through multi-protocol conference servers, and through signaling gateways.

Dual-Stack Endpoints

New video endpoints such as Polycom HDX have sufficient performance to run both protocol stacks—SIP and H.323—in parallel. If the same endpoint is registered via H.323 to a gatekeeper and via SIP to a SIP registrar, they can place and receive video calls over either of the networks. Figure 4 depicts the configuration.

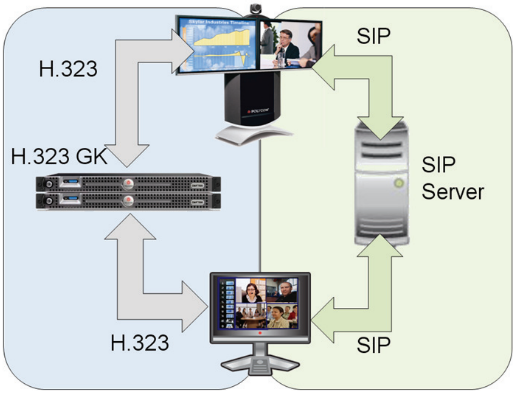

Figure 4: Dual-Stack Endpoints

The limitation of this approach is that only new endpoints have the performance to support two protocol stacks, and that solution cannot be used for the large installed base of older video endpoints that support only H.323.

Multiprotocol Conference Servers

Conference servers such as Polycom RMX 2000 and 4000 have much more resources than endpoints and support H.323 and SIP (as well as H.320) protocols. Therefore, endpoints can connect to the conference server via SIP or H.323 and be part of the same conference. Figure 5 describes the configuration.

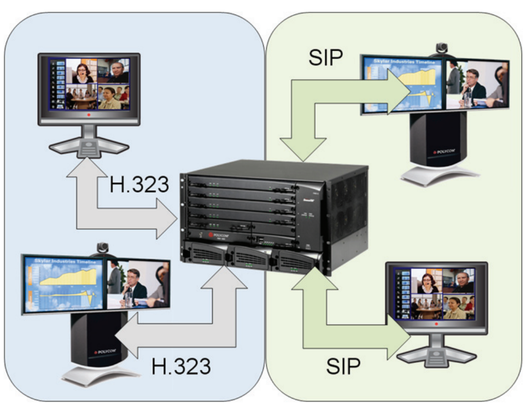

Figure 5: Multiprotocol Conference Servers

This approach works well for multipoint conferences but is not efficient for point-to-point calls between H.323 and SIP endpoints. Connecting every call through the conference server is a challenge not only from user interface perspective; it also uses scarce resources and reduces the quality unnecessarily.

However, the price-performance ratio in conference servers is going down rapidly, and new technologies such as Polycom Video Clarity™ technology in the Polycom RMX, actually improve the video quality, so the use of the conference server as a bridge between SIP and H.323 networks is becoming more attractive.

Signaling Gateway

Since both SIP and H.323 rely on the same protocols (RTP and RTCP) for transmitting media streams, it is possible to create a signaling gateway that maps the H.323 and SIP signaling without processing the media.

Figure 6 depicts the configuration.

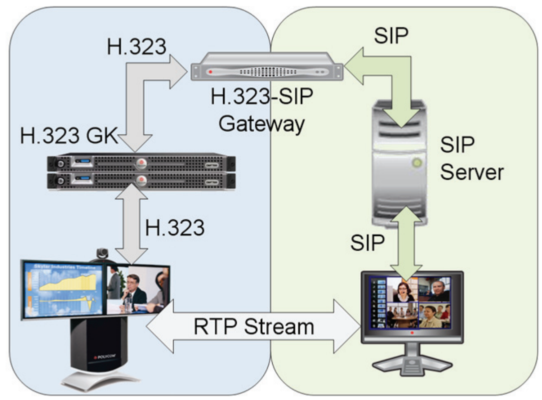

Figure 6: Signaling Gateway

Media processing, and especially video processing, is very resource-intensive. While signaling messages generate traffic in the magnitude of few kilobits per second, video media streams can be in the megabits per second. As a result, signaling gateways are much less expensive and much more scalable than their media counterparts. Video quality is not impacted by the gateway function either.

The signaling gateway function is implemented in Polycom DMA 7000 V4.0 which allows SIP and H.323 endpoints to connect and communicate.

In summary, each of the approaches to connecting SIP and H.323 networks has benefits and drawbacks. A combination of them is required to assure smooth migration of H.323 customers to SIP without losing functionality and video quality.

Conclusion

Visual communication is expanding beyond enterprise conference rooms to the user's desktop. The trend towards Unified Communications requires integrating video with variety of SIP-based systems in corporate environments.

SIP has already been deployed for visual communication in certain scenarios. The trend towards UC will increase the importance of the SIP protocol as glue that holds together the multi-vendor UC network.

Transition from H.323 to SIP will be gradual, and interoperability with the installed H.323 base throughout the process is a key requirement and main technical challenge.

Polycom is uniquely positioned to leverage its broad product portfolio, market leadership and extensive partner network to lead customers towards UC and transform traditional video conferencing into tomorrow's visual communications.

About the Author

Stefan Karapetkov is Emerging Technologies Director at Polycom, Inc. where he focuses on the visual communications market and technology. He has MBA from Santa Clara University (USA) and an MS degree in Engineering from the University of Chemnitz (Germany). He has spent more than 15 years in product management, new technology development, and product definition. His blog is http://videonetworker.blogspot.com/.

Acknowledgements

I would like to thank my colleagues Mike Tucker and Rick Flott for their contributions to this work.

References

[1] Definition of "Unified Communications": http://en.wikipedia.org/wiki/Unified_communications

[2] ITU-T Recommendation, 2006. H.323 v6 - Packet-Based Multimedia Communications Systems, ITU-T Document, May 2006. http://www.itu.int/itudoc/itu-t/aap/sg16aap/recaap/h323/h323v6.html

[3] Rosenberg, J. et al, 2002. RFC 3261 SIP: Session Initiation Protocol. IETF Request For Comments, June 2002. http://www.ietf.org/rfc/rfc3261.txt

[4] Campbell, B., 2002. RFC 3428 Session Initiation Protocol (SIP) Extension for Instant Messaging. IETF Request For Comments, December 2002. http://www.ietf.org/rfc/rfc3428.txt

[5] Saint-Andre, P., 2004. RFC 3921 Extensible Messaging and Presence Protocol (XMPP): Instant Messaging and Presence. IETF Request For Comments, October 2004. http://www.ietf.org/rfc/rfc3921.txt

[6] International Multimedia Telecommunications Consortium (IMTC) http://www.imtc.org/

[7] Andreasen, F. et al, 2003. RFC 3435 Media Gateway Control Protocol (MGCP), IETF Request for Comments, January 2003. http://www.ietf.org/rfc/rfc3435.txt

[8] Karapetkov, S. 2009, page 2. Scalable Infrastructure for Distributed Video. Polycom White Paper, November 2009. http://www.polycom.com/global/documents/whitepapers/wp_scalable_architecture_for_distributed_video.pdf

[9] Schulzrinne, H. et al, 2003. RFC 3550 RTP: A Transport Protocol for Real-Time Applications. IETF Request For Comments, July 2003. http://www.ietf.org/rfc/rfc3550.txt

[10] Handley, M. et al, 1998. RFC 2327 SDP: Session Description Protocol. IETF Request For Comments, April 1998. http://www.ietf.org/rfc/rfc2327.txt

[11] Fielding, R. et al, 1999. RFC 2616 Hypertext Transfer Protocol -- HTTP/1.1. IETF Request For Comments, June 1999. http://www.ietf.org/rfc/rfc2616.txt

[12] Advanced Encryption Standard definition http://en.wikipedia.org/wiki/Advanced_Encryption_Standard

[13] Baugher, M. et al, 2004. RFC 3711 The Secure Real-time Transport Protocol (SRTP). IETF Request For Comments, March 2004. http://www.ietf.org/rfc/rfc3711.txt

[14] Rosenberg, J. et al, 2003. RFC 3489 STUN - Simple Traversal of User Datagram Protocol. IETF Request For Comments, March 2003. http://www.ietf.org/rfc/rfc3489.txt

[15] Rosenberg, J. et al, 2009. Traversal Using Relays around NAT (TURN): Relay Extensions to Session Traversal Utilities for NAT (STUN). IETF Internet Draft, July 2009. http://tools.ietf.org/html/draft-ietf-behave-turn-16

[16] Rosenberg, J., 2010. RFC 5245 Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols. IETF Request for Comments, April 2010. http://datatracker.ietf.org/doc/rfc5245/

[17] Ott, J. et al, 2006. RFC 4585: Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF). IETF Request for Comments, July 2006. http://www.ietf.org/rfc/rfc4585.txt

[18] Levin, O., 2006 et al. RFC 4574: The Session Description Protocol (SDP) Label Attribute. IETF Request For Comments, August 2006. http://www.ietf.org/rfc/rfc4574.txt

[19] Hautakorpi, J. et al, 2007. RFC 4796: The Session Description Protocol (SDP) Content Attribute. IETF Request For Comments, February 2007. http://www.ietf.org/rfc/rfc4796.txt

[20] Camarillo, G. et al, 2002. RFC 3388: Grouping of Media Lines in the Session Description Protocol. IETF Request For Comments, December 2002. http://www.ietf.org/rfc/rfc3388.txt

[21] Even, R., 2006. RFC 4573: MIME Type Registration for RTP Payload Format for H.224. IETF Request for Comments, July 2006.http://www.ietf.org/rfc/rfc4573.txt