

Demystifying Multimedia Conferencing Over the Internet Using the H.323 Set of Standards

James Toga, Emerging Products Division, Intel Architecture Labs, Intel Corporation
Hani ElGebaly, Emerging Products Division, Intel Architecture Labs, Intel Corporation

Index words: H.323, conferencing, Internet, multimedia, gatekeeper

Abstract

The Telecommunication Sector of the International Telecommunication Union (ITU-T) has developed a set of standards for multimedia conferencing over packet-based networks. These standards are aggregated under a standard umbrella termed Recommendation H.323. Recommendation H.323 describes terminals, equipment, and services for multimedia communication over networks such as the Internet.

H.323 terminals and equipment carry real-time voice, video, and data, in any combination thereof. Terminals signal calls using Q.931-derived procedures defined in Recommendation H.225.0. After the call signaling phase, terminals proceed to the call control phase where they exchange capabilities and logical channel information using the H.245 protocol defined in Recommendation H.245. Once the call has been established, audio and video (if supported) are initiated. Both media types use the Real Time Protocol (RTP) defined by the Internet Engineering Task Force (IETF) as their Transport Protocol. Procedures for audio and video packet format and transport are described in Recommendation H.225.0. Recommendation H.323 allows for the use of a variety of video codecs (e.g., H.261, H.263, H.263+) and audio codecs (e.g., G.711, G.723.1). Data collaboration is also allowed using Protocol T.120. We provide an overview of these protocols and explain the H.323 call scenario.

Other entities such as gatekeepers, Multipoint Control Units (MCUs), and gateways are also addressed in Recommendation H.323. These entities allow network management, centralized multipoint, and interoperability with other conferencing standards. We explain each of these entities briefly and provide some scenarios of their interaction with the H.323 terminals.

IP telephony has become an important driver for packet-based communications. We address the role of H.323 procedures in deploying IP telephony, and describe how

new H.323 features such as supplementary services and security facilitate this purpose.

Introduction

Recommendation H.323 [8] describes the procedures for point-to-point and multipoint audio and video conferencing over packet-switched networks. In addition to video conferencing terminals, Specification H.323 describes other H.323 entities including gateways, gatekeepers, and MCUs. Gateways allow interoperation of H.323 systems with other audio/video conferencing systems on integrated services digital networks (ISDN), plain old telephone systems (POTS), asynchronous transfer mode (ATM), and other transports. Gatekeepers provide admission control and address translation to H.323 endpoints. MCUs can engage more than two H.323 endpoints in a centralized multipoint conference.

Recommendation H.323 comprises a number of related documents that describe terminals, equipment, services, and interactions. Examples of other core standards that are referred to in Recommendation H.323 are H.225.0 [5] (procedures for call signaling, media packet format, and synchronization); H.245 [7] (procedures for capability exchange, channel negotiation, and flow control); H.450.x [11] (procedures for supplementary services); H.246 [8] (procedures for terminals' interoperability through gateways); and H.235 (security and encryption procedures). Other referenced standards include media codecs for audio (such as, G.711, G.723.1, G.729, and G.722) and video (such as, H.261 and H.263).

The purpose of this paper is to present an overview of the H.323 core components and functionality and the current industry trends with respect to H.323, such as IP telephony and corporate conferencing. Towards this goal, we briefly describe the H.323 architecture, the responsibilities of H.323 entities, and basic call scenarios. The main obstacles to the adoption and deployment of the H.323 standard in industry are also explained. Finally,

H.323 zone management, multipoint operation, telephony services features, and security are highlighted.

H.323 Architecture

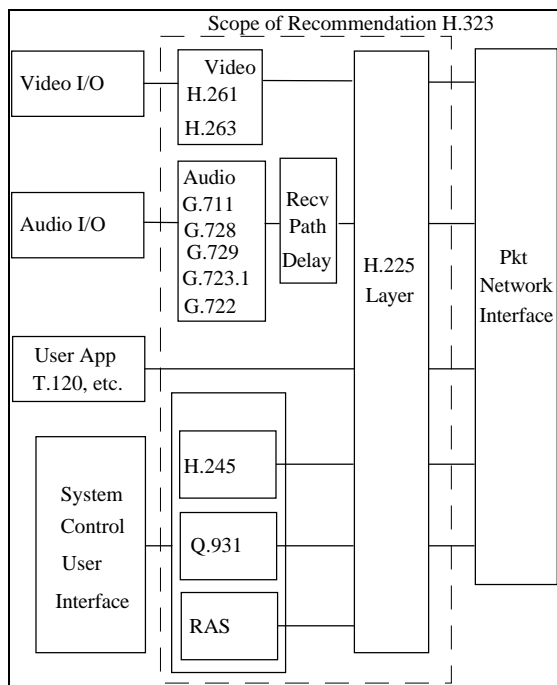


Figure 1: H.323 architecture

The architecture of an H.323 terminal is shown in **Figure 1**. The scope of Recommendation H.323 is limited to the definition of the media compression standard, packet format, signaling, and flow control. Media capturing such as video capturing schemes, audio recording, or user data applications are outside the scope of Recommendation H.323. Initial H.323 implementations targeted IP networks; however, Recommendation H.323 supports alternative transports such as IPX. The Transport layer carries control and media packets over the network; there are no specific requirements for the underlying transport except for support for reliable and unreliable packet modes. An example of a well-known transport is User Datagram Protocol (UDP) over Internet Protocol (IP).

The video codec (e.g., H.261, etc.) encodes the video from the video source (i.e., camera) for transmission and decodes the received video code that is output to a video display. The mandatory video codec for an H.323 terminal is H.261 [3] with quarter common intermediate format (QCIF) resolution. (QCIF is a video picture size.) Other codecs such as H.263 may be supported. H.263 [4] has better picture quality and more options. A terminal can also support other picture sizes such as CIF and SQCIF. During the terminal capability exchange phase, the video

codecs, resolution, bitrate, and algorithm options are exchanged between terminals, using the H.245 protocol. Terminals can open channels only with parameters and options chosen from the intersection of the capability sets. In general, the receiver always specifies what the transmitter may send.

The audio codec (G.711, etc.) encodes the audio signal from the microphone for transmission and decodes the received audio code that is output to the loudspeaker. G.711 0 is the mandatory codec for an H.323 terminal. A terminal may be capable of optionally encoding and decoding speech using Recommendations G.722, G.728, G.729, MPEG1 audio, and G.723.1. Since G.711 is a high-bitrate codec (64Kb/s or 56Kb/s), it cannot be carried over low-bitrate (< 56 kbps) links. G.723.1 [2] is the preferred codec in this situation because of its reasonably low rate (5.3Kb/s and 6.4 Kb/s). H.323 terminals open logical channels using a common capability that is supported by all entities and exchanged during the H.245 capability exchange phase.

The Data Channel supports telematic applications such as electronic whiteboards, still image transfer, file exchange, database access, audiographics conferencing, etc. The standardized data application for real-time audiographics conferencing is T.120. Other applications and protocols may also be used via H.245 negotiation such as chatting and fax.

The System Control Unit (H.245, H.225.0) provides signaling and flow control for proper operation of the H.323 terminal. H.245 [7] is the media control protocol that allows capability exchange, channel negotiation, switching of media modes, and other miscellaneous commands and indications. The H.225 standard describes the Call-Signaling Protocol used for admission control and for establishing connection between two or more terminals. The Connection Establishment Protocol is derived from the Q.931 specification [12]. The H.225.0 [5] Layer also formats the transmitted video, audio, data, and control streams into messages for output to the network interface, and it retrieves the received video, audio, data, and control streams from messages that have been input from the network interface, using the Real Time Transport Protocol (RTP) and its companion, the Real Time Control Protocol (RTCP). The RTP performs logical framing, sequence numbering, timestamping, payload distinction, source identification, and occasionally, error detection and correction as appropriate to each media type. The RTCP provides reporting and status that may be used by both senders and receivers to correlate performance on the media streams.

H.323 Entities and Responsibilities

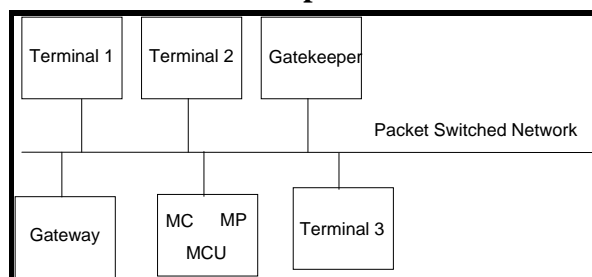


Figure 2: H.323 entities

Figure 2 shows the different entities that can be involved in an H.323 conference. A gatekeeper is an optional element in an H.323 conference that provides call control services to the H.323 endpoints such as address translation, admission control, bandwidth control, and zone management. The gatekeeper may also provide other optional functions such as call authorization and call accounting information. A gatekeeper cannot be specified as a destination in a call. Address translation is the method by which an alias address (e.g., e-mail address) is translated to a transport address. Admission control is a way of limiting H.323 access to a network, or to a particular conference using Request, Admission, and Status (RAS) messages defined in Recommendation H.225.0. A gatekeeper also manages bandwidth allocation to H.323 endpoints using RAS messages. Zone management defines the scope of entities over which a gatekeeper has control. These include endpoints, gateways, and MCUs. In addition, a gatekeeper can allow secure access to the conference using various authentication mechanisms. Q.931 and H.245 messages can be routed through the gatekeeper, and statistical information about the calls in progress can be collected. Gatekeepers may also perform telephony service operations such as call forwarding and call transfer.

H.323 endpoints can interact with each other directly in a point-to-point or multipoint conference if no gatekeeper is present. When a gatekeeper is present, all endpoints have to register with it.

A gateway operates as an endpoint on the network that provides real-time, two-way communication between H.323 terminals on the packet-based network and other ITU terminals on a switched-circuit network, or to another H.323 gateway. Other ITU terminals include those complying with Recommendations H.310 (B-ISDN), H.320 (ISDN), H.321 (ATM), H.322 (GQoS-LAN), H.324 (GSTN), H.324M (Mobile), and POTS. The gateway provides the appropriate translation between transmission formats (for example, H.225.0 of an H.323 endpoint to/from H.221 of an H.320 endpoint) and between communication procedures (for example, H.245

of an H.323 endpoint to/from H.242 of an H.320 endpoint). This translation is specified in Recommendation H.246. The gateway also performs call setup and clearing on both the network side and the Switched-Circuit Network (SCN) side. Translation between video, audio, and data formats may also be performed in the gateway. In general, the purpose of the gateway is to complete the call in both directions between the network endpoint and the SCN endpoint in a transparent fashion.

The MCU is an endpoint on the network, which provides the capability for three or more terminals or gateways to participate in a multipoint conference. It may also connect two terminals in a point-to-point conference, which may later develop into a multipoint conference. The MCU consists of two parts: a mandatory Multipoint Controller (MC) and optional Multipoint Processors (MP). The MC provides the capability for call control to negotiate with all terminals to achieve common levels of communication. It is this element that is required for all multipoint conferences. The MP allows mixing, switching, or other processing of media streams under the control of the MC. The MP may process a single media stream or multiple media streams depending on the type of conference supported. In the simplest case, an MCU may consist only of an MC with no MPs.

The following section provides an overview of the basic operation of an H.323 endpoint in a point-to-point conference without a gatekeeper and then with a gatekeeper.

H.323 General Operation

Figure 3 shows call establishment and tear down steps between two H.323 endpoints without a gatekeeper. All of the mandatory Q.931 and H.245 messages exchanged are listed. Note that some of these messages may be overlapped for increased performance. Each message has an assigned sequence number at the originating endpoint. Endpoint A starts by sending a *Setup* message (1) to endpoint B containing the destination address. Endpoint B responds by sending a Q.931 *Alerting* message (2) followed by a *Connect* message (3) if the call is accepted. At this point, the call establishment signaling is complete, and the H.245 negotiation phase is initiated. Both terminals will send their terminal capabilities (4) in the *terminalCapabilitySet* message. The terminal capabilities include media types, codec choices, and multiplex information. Each terminal will respond with a *terminalCapabilitySetAck* (5) message. The terminals' capabilities may be resent at any time during the call. The Master/Slave determination procedure (6-8) is then started. The H.245 Master/Slave determination procedure is used to resolve conflicts between two endpoints that can

both be the MC for a conference, or between two endpoints that are attempting to open bi-directional channels at the same time. In this procedure, two endpoints exchange random numbers in the H.245 *masterSlaveDetermination* message to determine the master and slave endpoints. H.323 endpoints are capable of operating in both master and slave modes. Following master/slave determination procedures, both terminals proceed to open logical channels (9-10). Both video and audio channels are unidirectional while data is bi-directional. Terminals may open as many channels as they want. Only one logical channel is shown in **Figure 3**, yet the same procedure applies if more channels are opened.

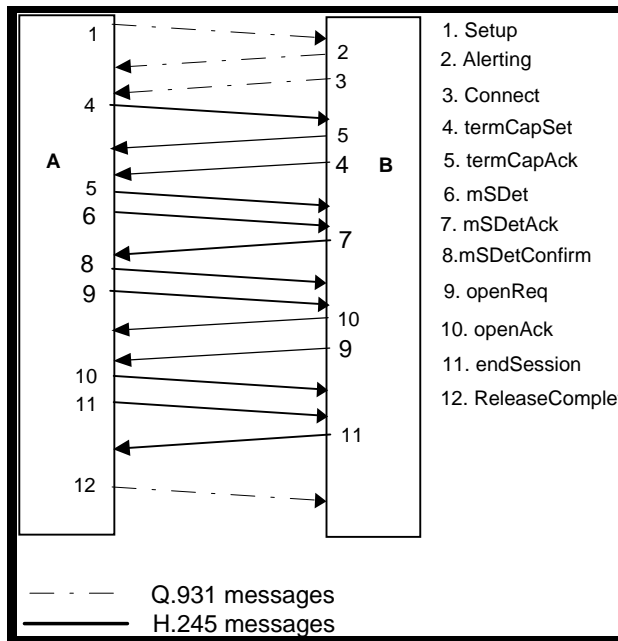


Figure 3: Q.931 and H.245 messages exchanged between two H.323 endpoints

Other H.245 control messages may be exchanged between the endpoints to change media format, request video key frames, change the bitrate, etc.

Termination of a call is initiated by one endpoint sending an *endSession* message (11). Endpoint B, on receiving the *endSession* command, will respond with another *endSession* message (11) to Endpoint A. Endpoint A will finally send a Q.931 *ReleaseComplete* message (12), and the call is terminated.

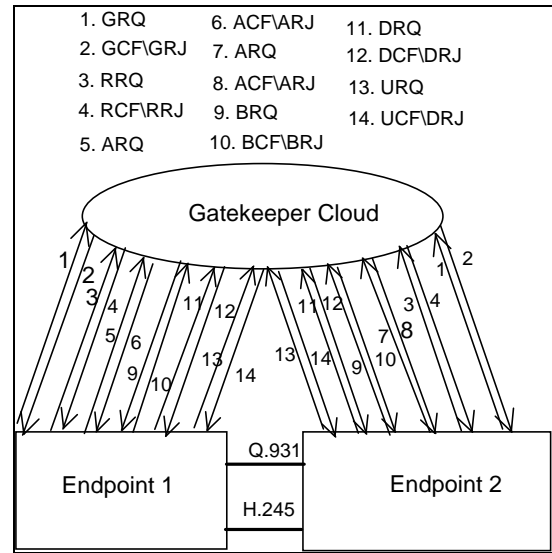


Figure 4: Gatekeeper interaction in an H.323 call

Figure 4 shows messages exchanged between a gatekeeper and an H.323 endpoint. Before the conference starts, both endpoints look for a gatekeeper by multicasting a *GatekeeperDiscovery* (GRQ). Request. The gatekeeper will reply either with a *GatekeeperConfirm* (GCF) message or with a *GatekeeperReject* (GRJ) message. Both endpoints will then register their alias names with the gatekeeper using the *RegistrationRequest* (RRQ) message. The gatekeeper will acknowledge by sending a *RegistrationConfirm* (RCF) message or will deny it using a *RegistrationReject* (RRJ) message. Registering alias names with the Gatekeeper allows endpoints to call each other using user-friendly addresses such as e-mail, etc., rather than the transport address. The discovery and registration procedure is valid until the gatekeeper indicates otherwise.

An endpoint or gatekeeper can request the location of another endpoint using its alias name by using a *LocationRequest* (LRQ) message, and the gatekeeper replies with a *LocationConfirm* (LCF) message containing the resolved address for the alias name.

When a user places a call from an endpoint, the endpoint starts by requesting admission from the gatekeeper using an *AdmissionRequest* (ARQ) message. The gatekeeper can accept (ACF) or deny the request (ARJ). If the call is accepted, the endpoint sends a Q.931 *Setup* message to the remote destination. The recipient of the *Setup* message in turn requests admission from its gatekeeper by sending an ARQ. When the call is accepted, the Q.931 call signaling sequence is completed followed by the H.245 message negotiation. The *AdmissionRequest* (ARQ) message carries the initial bandwidth the endpoint requires for the duration of the conference. If during H.245 logical channel negotiation, an endpoint requires more

bandwidth, it issues a *BandwidthRequest* (BRQ) message to the gatekeeper. If the request is accepted, the gatekeeper replies with a *BandwidthConfirm* (BCF) message; otherwise, it replies with a *BandwidthReject* (BRJ) message.

When the call is terminated, both endpoints send a *DisengageRequest* (DRQ) message to inform the gatekeeper that a call is being terminated. The gatekeeper replies with a *confirm* (DCF) or *reject* (DRJ) message. Alternatively, endpoints may unregister from the gatekeeper by sending an *UnregisterRequest* (URQ) message. The gatekeeper replies with an *UnregisterConfirm* (UCF) message or an *UnregisterReject* (URJ) message.

H.323 Deployment Obstacles

In order to achieve H.323 deployment in real networks, limitations at both the network level and the client platform level must be resolved at the H.323 client. The client should scale performance based on the available bandwidth. Given the inconsistencies of networks with best effort traffic, (i.e., no guaranteed Quality of Service (QoS), the Internet), it is extremely important to provide mechanisms for fault tolerance and error resiliency at the client platform, if it is to be used on an unmanaged network such as the Internet. At the network level, broad connectivity, policy management, and security are considered the major issues in the deployment of a new technology such as H.323. Switched Circuit Network connectivity is achieved in the H.323 context by using gateways for H.320, H.324, H.323, POTS, and other endpoints on other networks. Policy management is achieved using gatekeepers to provide call admission, authentication, and zone management. Deployment of QoS protocols such as RSVP can also help with policy management. The security of media streams is another important factor in the success of H.323 deployment, especially in unsecured environments such as the Internet.

The platform consists of two main components: the operating system and the CPU. Many media compression algorithms are currently limited by the machine speed. (We expect this issue to improve as more powerful processors hit the market.) Moreover, popular desktop operating systems do not supply consistent real-time services. Multitasking can adversely affect the quality of audio and video in an H.323 conference.

H.323 Applications

A number of applications can take advantage of H.323 technology both in the corporate environment and in the home-user environment. One obvious application is video

conferencing between two or more users on the network. In this application, the user is expecting the same services as those provided by a telephone. The quality of service should be equal to or better than POTS. The Internet does not appear to be readily addressing these issues; however, there is work in progress at the corporate Intranet level to improve the quality of multimedia communication.

Multimedia call centers are another application for H.323. An H.323 call center provides a well-integrated environment for Web access and other data/voice business situations. The call centers are used by banks for customer service, shops for extra retail outlets, etc. The call center can just be an H.323 terminal or an MCU, or it can be a full featured endpoint with a gatekeeper, a gateway, and MCU capability. The front end of the legacy call center may be a gateway, which allows installed systems to operate with minimal disruption.

Another compelling application for H.323 is telecommuting. Telecommuters can attend meetings at their companies, check their mail, or talk with someone at the company while on the road or at home.

Finally, IP telephony is another area in which H.323 has found a significant role; this will be covered in a later section.

H.323 Zone Management

Gatekeepers fulfill a required set of operational responsibilities and may offer a number of optional functions to entities within their *zone*. Before we describe how zones are managed, we will review some of these responsibilities and functions.

A gatekeeper acts as a monitor of all H.323 calls within its zone on the network. It has two main responsibilities: call approval and address resolution. An H.323 client that wants to place a call can do so with the assistance of the gatekeeper. The gatekeeper provides the address resolution to the destination client. (This division of work is due to alias name registration procedures.) During this address resolution phase, the gatekeeper may also make permissioning decisions based upon available bandwidth. The gatekeeper can act as an administrative point on the network for IT/IS managers to control H.323 traffic on and off the network.

Strictly speaking, a gatekeeper zone is defined by what it contains: it is defined by all of the endpoints, gateways, and MC(U)s that are or will be registered with a gatekeeper. Another way to describe a gatekeeper zone is to call it an "administrative domain," although the formal ITU-T recommendation text uses zone. An example of gatekeeper zones is given in **Figure 5**.

Some important aspects of zone coverage are summarized as follows:

- Zones are defined by all H.323 devices registered to a single gatekeeper.
- Zone design may be independent of physical topology.
- Each zone has only one gatekeeper.
- Zone definition is implementation-specific.
- Gatekeeper zones are logical in nature.
- Network topology and administrative scope are both factors in zone design.
- Resources such as gateways and proxies may affect the partitioning of zones.

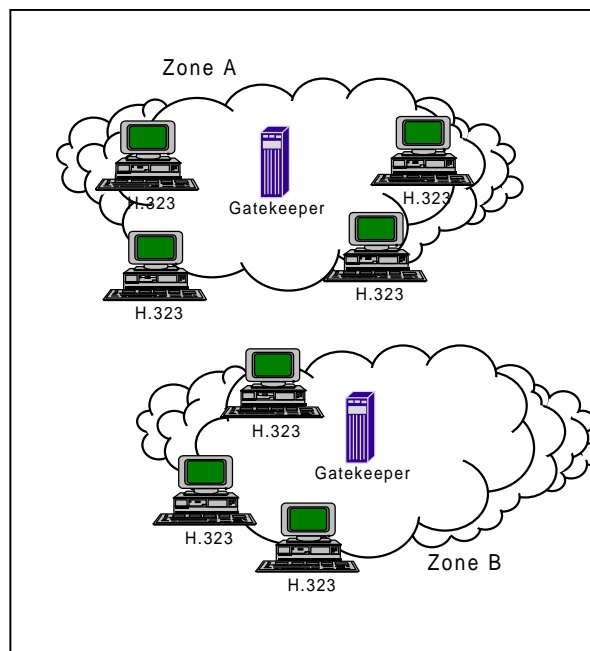


Figure 5: Gatekeeper zones

Multipoint Conferences: Centralized Versus Decentralized

Multipoint conferences are defined as calls between three or more parties. During these conferences, call control and media operations can become considerably more complex than during a simple point-to-point conference. The coordination and notification of participants entering and leaving a conference along with the marshalling of the media streams require the presence of at least a Multipoint

Control (MC). In other situations, an MCU is required; we will explain why in the next section.

The two broad models of multipoint, centralized and decentralized, differ in their handling of real-time media streams (audio and video). The centralized model operates in the same fashion as other circuit-based conferencing (e.g., H.320 or telephony “bridges”). In this model, all of the audio and video is transmitted to a central MCU that mixes the multiple audio streams, selects the corresponding video stream, and re-transmits the result to all of the participants. **Figure 6** illustrates this procedure.

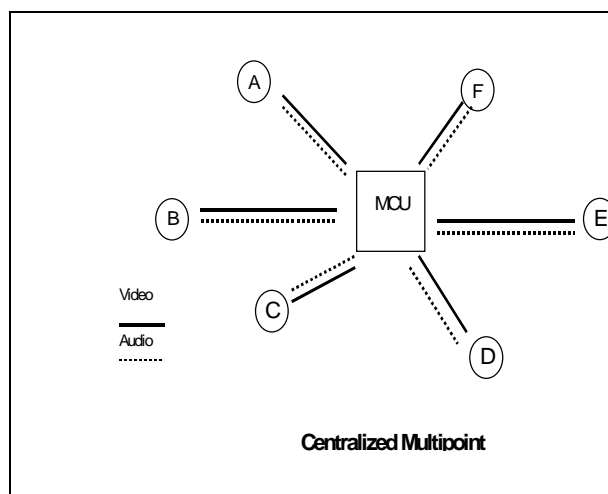


Figure 6: Centralized model

Note that the MCU is only ‘logically’ at the center of the conference configuration; the physical topology may be very different. The operational model of the conference is such that each endpoint is exchanging media control signaling (H.245), audio, and video directly with the MCU. In order to prevent echo, the MCU must provide the current speaker with a custom audio mix that does not contain the speaker’s own audio. The remaining participants may all receive the same audio and video media. In some implementations, the MCU may actually decode the video streams and combine them in a mixing operation to provide continuous presence of all participants on the endpoint displays. The amount of processor speed required for this operation increases significantly.

The decentralized model shares common control characteristics with the centralized model, but the media streams are handled differently. One of the participating entities must be an MC, but it will typically be co-located with one of the endpoints. All of the H.245 connections will have one end terminating at the MC just as with the

MCU in the centralized model. Whereas the MCU does the media processing in the centralized model, the media streams are sent and received by all participating entities on a peer-to-peer basis in the decentralized model. As **Figure 7** demonstrates, the logical configuration for this is a bus. There is no MCU to process the multiple streams; each entity is responsible for its own audio mixing and video selection. The media may be sent between all entities utilizing either multicast, or a multiple uni-cast if the underlying network does not support multicast.

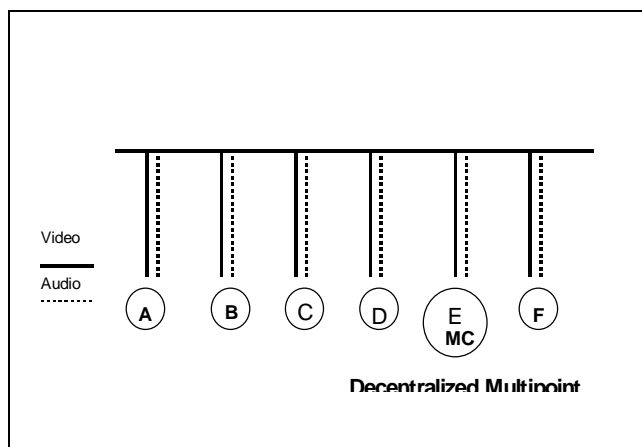


Figure 7: Decentralized model

There are a number of advantages and disadvantages to both the centralized and the decentralized models. In the centralized model, the endpoints that participate in this type of conference are not required to be as powerful as in the decentralized model. Each endpoint has only to encode its locally produced media streams and decode the set sent by the MCU. The MCU may provide specialized, or higher performance conferencing. The MCU's ability to provide custom-mixed media streams can allow otherwise constrained endpoints to participate in conferences. For example, if an endpoint is connected to the conference by a low-bitrate connection, it may only have the capacity to transmit and receive one media stream. If this same endpoint were connected to a decentralized conference, it would have no ability to detect which speaker was the focus when the focus changed. In most situations, the MCU will choose the largest common set of attributes within which to operate the conference. This may lead to a conference that is operated in a least common denominator mode, such as QCIF, rather than CIF or lower video resolution.

By definition, the decentralized multipoint model does not require the presence of an MCU, a potentially expensive and limited resource. It allows for disproportionate processing at the endpoints with each running at its own

level. The decentralized multipoint model does require that one of the participating entities contain an MC. If the endpoint that has the MC leaves the conference, the MC must stay active or the conference is terminated. In order for this model to be bandwidth efficient, the underlying network should support multicast. If it doesn't, each endpoint can send multiple uni-cast streams to all others, but this becomes increasingly inefficient with more than four entities participating in a conference.

H.323 Features for IP Telephony

The initial environment envisioned for H.323 was the corporate network environment, primarily local area networks. Wide Area Network (WAN) access was to be gained by using gateways to H.320/ISDN. During the implementation of Revision 1 of H.323, it became clear that IP telephony was gaining popularity and relevance as various infrastructure elements were improved upon. A number of proprietary IP-based telephones were creating many small islands that could not communicate with one another. Recommendation H.323 provided a good basis for establishing a universal IP voice and multimedia communication in larger, connected networks. With Revision 2 of the Recommendation, new additions and further extensions were added specifically to make it more suitable for IP telephony. These changes will be described in a later section.

By using Q.931 as its basis for establishing a connection, H.323 allows for relatively easy bridging to the public switched telephone networks (PSTN) and circuit-based phones. The required voice codec of G.711 also allows for easy connections to the legacy networks of telephones. The uncompressed 64kb/sec stream can easily be translated between digital and analog media. One of the addressing formats provided in Recommendation H.323 is the E.164 address. This is another ITU Recommendation that specifies standard telephone numbers (e.g., the digits 0-9, * and #). These addresses, which ultimately map onto the IP addresses for the H.323 endpoints, allow regular telephones to 'dial' them. Gatekeepers provide the final important element for IP telephony. Gatekeepers supply the ability to have integrated directory and routing functions within the course of the call setup. These operations are important for real-time voice or video when resources must be balanced, and points of connectivity are highly dynamic. The gatekeeper functions, which provide call permissioning and bandwidth control, enable load monitoring, provisioning, and ultimately, commercial-grade IP telephony service.

Interoperability

Recommendation H.323 comprises a number of interrelated documents and sub-recommendations.

Therefore, customers are faced with a number of options, which they may or may not implement. These design choices, in conjunction with differing interpretations of the required elements, can lead to implementations that cannot interoperate. Stated more simply, compliance with the recommendations does not always imply interoperability. The complexity and flexibility of H.323 essentially requires that implementations be tested together to ensure interoperability.

To this end, the International Multimedia Technical Consortium (IMTC) has focussed largely on this area of interoperability. The testing events sponsored by this consortium have provided a venue for more than thirty companies to participate in "bake-offs" to test the interoperability of their product. The methodology that is followed provides for a gradual testing of component protocol stacks that then leads to higher level end-end scenario testing. **Figure 8** illustrates the grouping of the components. Each horizontal level provides increasing H.323 functionality, and each vertical group indicates a specific stack function that is required.

Much of the mass interoperability testing has occurred in face-to-face events. In theory, the H.323 protocol should allow for testing from remote sites across the Internet; in practice, however, remote testing has not turned out to be useful. The lack of a controlled environment makes distinguishing interoperability issues from simple network problems extremely difficult.

Examples for initial interoperability problems encountered during some of the testing events include misaligned bit field and byte-ordering issues. These problems are common in the early development stage of protocols. For example, there are a number of adaptations that are specified in the Q.931 protocol used by H.323, which make the Protocol Data Units (PDUs) slightly different from the protocol messages used in the circuit world (such as ISDN).

Porting of the existing Q.931 code to the packet environment provided mis-matches. During the initial course of development, there were a small number of defects in the Recommendation that were discovered by the implementers. These defects were recorded in a document called the *Implementer's Guide* and eventually corrected in the revision of the base Recommendation. In addition to the low-level control protocol interoperability, the media stream is another potentially problematic area. Recommendation H.245 allows the receiver to specify the maximum bitrate that may be sent, but currently there is no way in which to specify the maximum Real Time Protocol (RTP) packet size. This can lead to interoperability problems if a receiver implementation makes assumptions about the buffering required and then cannot decode the packet. Although H.323 specifies G.711 and H.261 as its baseline codecs (audio and video respectively), a large number of initial H.323 implementations were targeted for low-bitrate connectivity where these codecs could not operate. The result of this was that the audio codec, G.723.1, was chosen due to its low (5.3-6.4kb/s) data rate. This low bitrate allowed H.263 video to operate in an acceptable fashion across a 33.6kb connection.

Implementation choices that are made as a result of operating environments elevate the interoperability issues to the next level. At an operational level, endpoint implementations may select to support an asymmetrical media model for optimal performance; other implementations may not. The ability to enter extended information at the user interface may determine whether an endpoint can operate with certain gateways or H.323 proxies. Gatekeepers may or may not provide intelligent zone control, which allows them to operate in a network connected to other activated gatekeepers. With the addition of a range of security options to H.323, strict policy constraints may prevent H.323 entities from interoperating while still complying with all the recommendations.

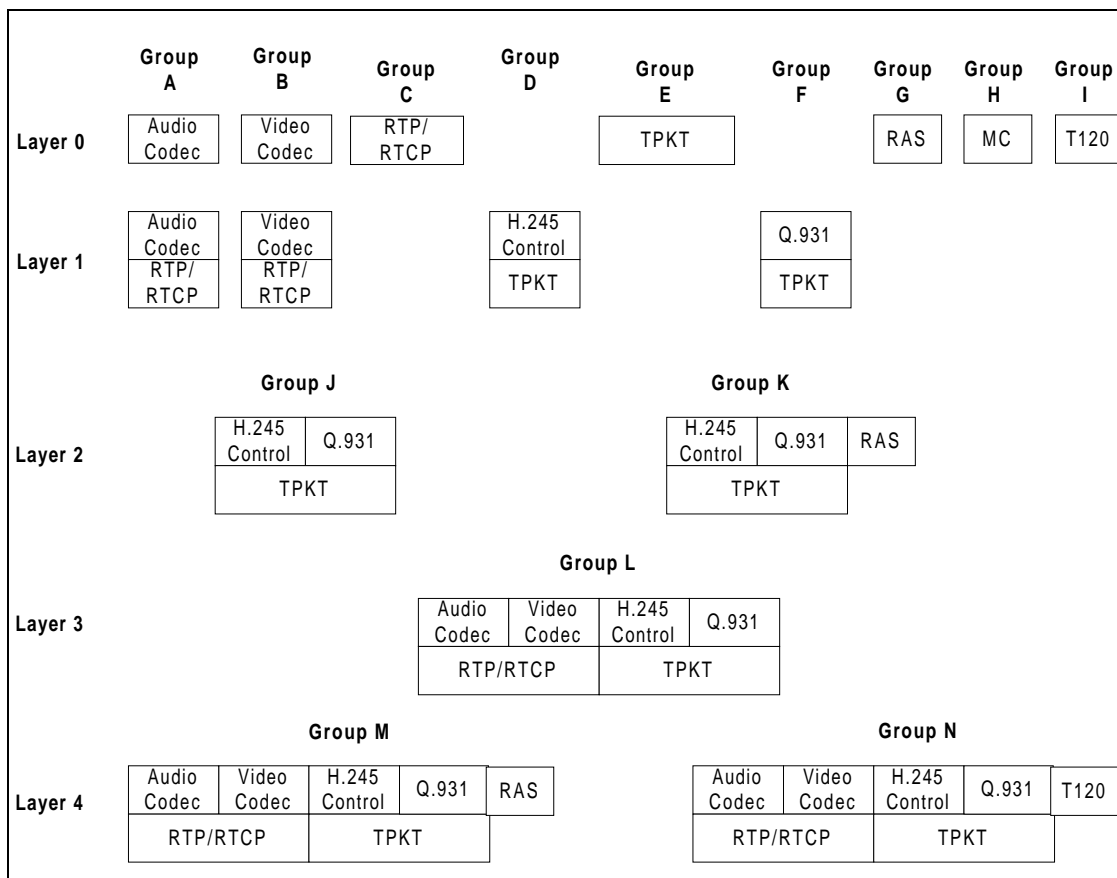


Figure 8: Testing matrix¹

New Features

Revision 2 (1998) of Recommendation H.323 contains a number of improvements for IP Telephony, among other areas. A new Recommendation (H.235) was developed to provide a full security framework for H.323 and other multimedia systems. It may provide the services of Authentication (which can be used for authorization), Privacy, and Integrity. The system can utilize underlying security protocols such as Internet Protocol Security (IPSEC) or Transport Layer Security (TLS) as established in the IETF. A number of new features and options were introduced regarding interactions with gatekeepers. During the process of discovering gatekeepers, endpoints may actually receive a number of gatekeeper addresses to utilize. This redundancy in the protocol will eliminate the single point of failure if a gatekeeper becomes inoperative. When endpoints register their current address with a gatekeeper, they may specify multiple addresses;

this allows a 'line hunting' mode of operation. Keep alive types of messages called Request In Progress (RIP) can stop premature retries on lengthy operations.

The call setup has a new method called FastStart that establishes bi-directional media in one round trip time of messages (discounting the establishment of the actual TCP connection). **Figure 9** shows FastStart messages exchanged. The *OpenLogicalChannels* messages that occur after H.245 is established in the regular case are piggy-backed on the Setup-Connect exchange. This facility allows instant audio connection that resembles the regular phone call model as opposed to the standard lengthy H.323 startup procedures.

¹ In this figure TPKT refers to the layer that segments a TCP stream into separate message 'packets' to be delivered to the H.323 application.

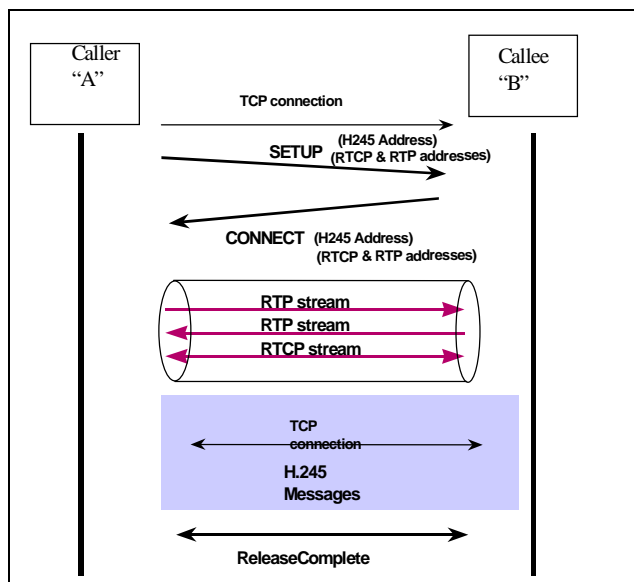


Figure 9: FastStart

Many new explicit address types have been added, including RFC822 (e-mail) formatted names and URLs, which network-based users find more familiar and easier to remember. Each endpoint may register under a number of these aliases. New unique call identifiers were added to better track calls as they traverse multiple network nodes and topologies. A number of new features were added to H.245 also, which allow for a much richer set of media possibilities. In addition to layered codecs, which divide up the video stream into additive layers of detail, Quality of Service parameters (e.g., RSVP) may now be signaled when opening up media streams. Finally, the ability to pass the media on other transports such as ATM (while keeping the rest of call and media control on IP) have been added.

In addition, a new family of recommendations has been developed to set up a framework for supplementary services. The H.450.1, H.450.2, and H.450.3 are loosely based on the Q-signals (QSIG) protocol, which is used by PBX and switch vendors. The initial services that have been defined are call transfer and call forward. Both of these features are a peer-peer option requiring no specific help from a centralized network entity (such as the PBX in the circuit world). A number of standardized supplementary services are expected to be developed in the future.

Along with the new features, the latest development of H.323 also includes an expanded topology model in the H.332 document. Recommendation H.332 allows an H.323 conference model to expand to literally thousands of participants. The tightly controlled H.323 'panel' is surrounded by a very large number of 'listeners.'

Members may participate in the conference by joining the panel. They may also leave as they wish. This H.323 panel is analogous to a panel on a stage in a large auditorium. Occasionally participants might get up from the audience and join the panel and others might leave the panel and join the audience. This movement of participants occurs using the standard Q.931 Invite and Join signaling as described in the H.225.0 document.

Conclusion

The H.323 system and its associated recommendations provide a useful and flexible system for multimedia communication. The factors that allow the protocols to easily bridge data and voice networks also make H.323 scalable. The ability of most, if not all of the system, to operate on a general-use platform such as a personal computer allows the H.323 system to scale with underlying processing power. As the processor's speed budget increases, the H.323 system can provide a better end-user experience. The dynamic exchange of capabilities allows the communications to change if needed during a call and adapt to any environmental or endpoint constraints.

Recommendation H.323 has become the single standards-based solution for a complete array of communication systems from simple point-to-point telephony to a rich multimedia conference with data sharing. Through continued effort by the ITU-T study group, Recommendation H.323 continues to evolve and adapt to new situations. Many of the real world difficulties in utilizing H.323 come about from infrastructure issues or other problems that are being resolved. Globally coordinated addressing and consistent QoS are two areas where we expect to see great improvements in the future. Some of these improvements will be facilitated by expected higher level communications between gatekeepers and gateways as their interaction is standardized.

References

- [1] ITU-T Recommendation G.711 (1988) "Pulse Code Modulation (PCM) of Voice Frequencies."
- [2] ITU-T Recommendation G.723.1 (1996) "Dual Rate Speech Coders for Multimedia Communication Transmitting at 5.3 & 6.3 kb/s."
- [3] ITU-T Recommendation H.261 (1993) "Video Codec for Audiovisual Services at p X 64 kb/s."

- [4] ITU-T Recommendation H.263 (1996) "Video Coding for Low Bit-rate Communication."
- [5] ITU-T Recommendation H.323 (1998) "Packet Based Multimedia Communications Systems."
- [6] ITU-T Recommendation H.225.0 (1998) "Call Signaling Protocols and Media Stream Packetization for Packet Based Multimedia Communications Systems."
- [7] ITU-T Recommendation H.245 (1998) "Control Protocol for Multimedia Communication."
- [8] ITU-T Recommendation H.246 (1998) "Interworking of H-Series Multimedia Terminals."
- [9] ITU-T Recommendation H.235 (1998) "Security and Encryption of H series (H.323 and other H.245 based) Multimedia Terminals."
- [10] ITU-T Recommendation H.332 (1998) "Loosely Coupled H.323 Conferencing."
- [11] ITU-T Recommendation H.450.1 (1998) "Generic Functional Protocol."
- [12] ITU-T Recommendation Q.931 (1993) "Digital Subscriber Signaling System No. 1 (DSS 1)-ISDN User-Network Interface Layer 3 Specification for Basic Call Control."

Outside of Intel, Mr. Toga develops standards and standards-based products within ITU-T, IETF, and IMTC. He is also involved in the following related activities:

Editor of ITU-T H.323 Implementors Guide
H.235 Security Standard

Chair of IMTC "Packet Networking Activity Group"

Chair of H.323 Interoperability Group

He has also written numerous articles for trade magazines. His e-mail is jim.toga@intel.com.

Hani ElGebaly is the project technical lead for the H.323/H.324 protocols development team within the Conferencing Products Division of Intel Architecture Labs. He received an M.Sc. in Computer Science from the University of Saskatchewan, Canada, and a B.Sc. in Electrical Engineering from Cairo University, Egypt. Mr. ElGebaly is currently pursuing a Ph.D. with the University of Victoria, Canada. His primary areas of interest include multimedia conferencing protocols, embedded programming, fault tolerant systems, and computer architecture. His e-mail is hani.el-gebaly@intel.com.

Authors' Biographies

Jim Toga holds a B.Sc in Chemistry from Tufts University and a M.Sc in Computer Science from Northeastern University. Before joining Intel, he was the principal engineer on StreetTalk* Directory with Banyan Systems where he designed and developed the Yellow Pages service. Presently, he is a senior staff software architect for the Standards and Architecture Group in the Intel Architecture Labs. He coordinates product groups giving guidance on architecture and standards. His primary tasks are H.323/Internet Telephony, Directory, and real-time security issues.

* All other trademarks are the property of their respective owners.