

# Voice over IP Signaling: H.323 and Beyond

Hong Liu and Petros Mouchtaris, *Telcordia Technologies*

## ABSTRACT

Signaling has been one of the key areas of Voice over IP (VoIP) technologies since inception. H.323 was the key protocol that allowed interoperability of VoIP products and moved the industry away from the initial proprietary solutions. Once the VoIP industry started maturing, some limitations of H.323 came to light. In this article we provide an overview of H.323, describe its capabilities, and discuss how its limitations are being addressed using the concept of gateway decomposition. We also discuss how H.323 can coexist with other protocols such as MGCP, H.248, and SIP which are attracting a lot of interest in the VoIP industry today.

## INTRODUCTION

Signaling is one of the most important functions in the telecommunications infrastructure because it enables various network components to communicate with each other to set up and tear down calls. Significant efforts were undertaken in past decades to develop the signaling protocols in use in today's telephone network, also known as the public switched telephone network (PSTN). These protocols, such as Signaling System No. 7 (SS7) and Q.931, are defined in large detailed specifications developed by various standardization organizations.

Similar efforts are now being undertaken to define voice over IP (VoIP) signaling. Since the very beginning of the VoIP industry, issues around signaling protocols for VoIP have been the focal point of industry debates. So far, the VoIP industry has gone through three stages in terms of signaling protocol evolution: *precommercial* (1980–1995), *PC-centric* (1995–1998), and *carrier grade* (1998 on).

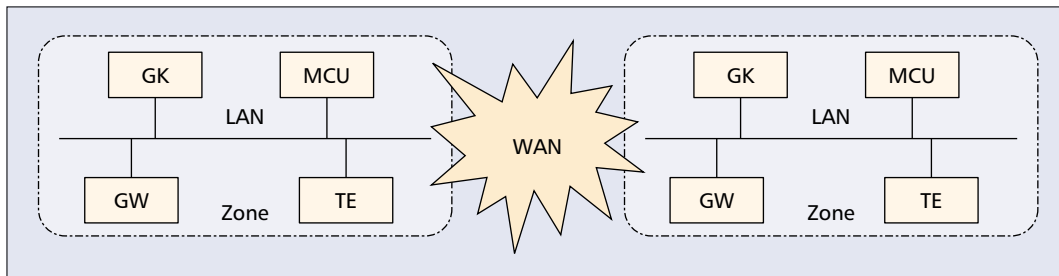
The precommercial stage was characterized by research activities in various universities and research organizations of the Internet community. Much of the work was coordinated by two working groups in the Internet Engineering Task Force (IETF), the Internet's standards organization. The Audio/Video Transport (AVT) working group produced the Real-Time Transport Protocol (RTP) [1]. The Multiparty Multimedia Ses-

sion Control (MMUSIC) working group designed a family of protocols for multimedia conferencing over the Internet, including the Session Initiation Protocol (SIP) [2] for session setup and teardown. The primary focus in this stage was on audio and video conferencing over the Internet. Interworking with the PSTN was only a small part of the overall effort. Until 1996, SIP was the only signaling protocol for multimedia conferencing over the Internet and was used by much Internet conferencing freeware/shareware such as VAT and CuSeeMe. The protocol underwent many revisions before it was approved by IETF as a proposed standard in March 1999.

The PC-centric stage started in early 1995 when commercial VoIP software first appeared on the market. Initially, these products allowed a user to place a call over the Internet from a multimedia PC to another multimedia PC. All the signaling and control functions resided on the PCs. Each product relied on a proprietary signaling protocol for call setup and teardown, which made it virtually impossible for two vendors' products to interoperate. To address this problem, the International Telecommunication Union (ITU) started work on standardizing VoIP signaling protocols in May 1995. In June 1996, Study Group 16 of ITU — Telecommunication Standardization Sector (ITU-T) decided on H.323 v. 1, referred to as a standard for real-time videoconferencing over nonguaranteed quality of service (QoS) LANs.

H.323 came out at the right time for the fledgling VoIP industry. The momentum of H.323 for VoIP was so great that by the end of 1996, most PC client software vendors were moving toward building H.323-compliant products. Unlike the previous stage, interworking with PSTN was the focus from the very beginning since bypass of PSTN telephone charges was then regarded as one of the main economic drivers for VoIP. Consequently, we also witnessed a proliferation of H.323 gateway products that enable phone calls to be made across the PSTN and the Internet.

The carrier-grade stage started around early 1998. As IP telephony service providers began to deploy networks of H.323 gateways to offer VoIP services, they soon realized that H.323 has



■ **Figure 1.** An H.323 network.

some limitations. H.323 assumed that a gateway handles signaling conversion, call control, and media transcoding in one box, which poses scalability problems for large-scale deployment. H.323 also had no provision for SS7 connectivity, which hinders its seamless integration with PSTN. In order to provide carrier-grade VoIP services, in May 1998 the concept of a decomposed gateway was introduced where call control resides in one box, called the *media gateway controller*, and media transformation resides in another box called the *media gateway*. The Media Gateway Control Protocol (MGCP) was introduced in 1998 [3]. After about two years of extensive work, ITU-T SG16 and IETF defined the media gateway control standard, called H.248 or *Megaco*, in June 2000.

In this article we provide a high-level overview of H.323, describing the various components of H.323 and the various signaling protocols defined as part of H.323. We discuss how H.323 interworks with the PSTN through a monolithic gateway and how the limitations of this approach can be addressed by decomposing the monolithic gateway. We discuss how H.323 can coexist and interwork with other VoIP signaling protocols such as MGCP and SIP. Then we provide a summary of this article.

## H.323 OVERVIEW

In this section we describe, at a high level, the H.323 architecture by defining the main components of the architecture: the terminal, the gatekeeper, the gateway, and the multipoint control unit. We then define the various protocols that are part of the H.323 family and are used by the components of the architecture for communicating with each other. We also define how services can be implemented within the H.323 architecture.

### THE H.323 ARCHITECTURE

The H.323 standard [4] was initially targeted to multimedia conferencing over LANs that do not provide guaranteed QoS. The functional architecture of an H.323 system is depicted in Fig. 1.

A typical H.323 network is composed of a number of *zones* interconnected via a WAN. Each zone consists of a single H.323 gatekeeper (GK), a number of H.323 terminal endpoints (TEs), a number of H.323 gateways (GWs), and a number of multipoint control units (MCUs), interconnected via a LAN. A zone can span a number of LANs in different locations, or just a single LAN. The only requirement is that each zone contain exactly one GK, which acts as the administrator

of the zone. The functionality of each component of the architecture is defined as follows:

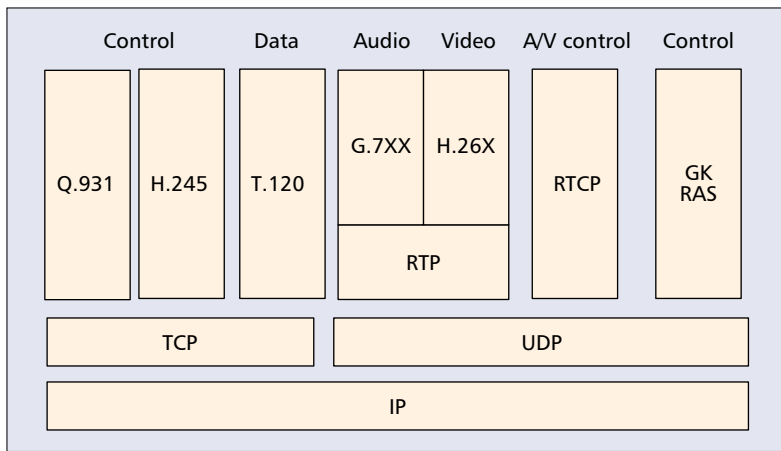
- **Terminal:** An H.323 TE is an endpoint in the network, which provides for real-time two-way communications with another H.323 terminal, GW, or MCU. This communication consists of control, indications, audio, moving color video pictures, and/or data between the two terminals. A terminal may set up a call to another terminal directly or with the help of a GK.
- **Gatekeeper:** The GK is an H.323 entity in the network that provides address translation and controls access to the network for H.323 terminals, GWs, and MCUs. The GK may also provide other services to the terminals, GWs, and MCUs such as bandwidth management and locating GWs. The GK function is optional in H.323 systems.
- **Gateway:** An H.323 GW is an endpoint in the network that provides real-time two-way communications between H.323 TEs on the packet-based network and terminals on the PSTN.
- **Multipoint control unit:** The MCU is an endpoint in the network that provides the capability for three or more terminals and GWs to participate in a multipoint conference.

### H.323 SIGNALING AND CONTROL

H.323 is an umbrella of the following four protocols:

- **Registration Admission and Status (RAS):** RAS is a transaction-oriented protocol between an H.323 endpoint (usually a TE or GW) and a GK. An endpoint can use RAS to discover a GK, register/unregister with a GK, requesting call admission and bandwidth allocation, and clearing a call. A GK can use RAS for inquiring on the status of an endpoint. There is also a mechanism for GKs to communicate with each other for address resolution across multiple zones. RAS is used only when a GK is present.
- **Q.931:** Q.931 is the signaling protocol for call setup and teardown between two H.323 TEs and is a variation of the Q.931 protocol defined for PSTN. H.323 adopted Q.931 so that interworking with PSTN/ISDN and related circuit-based multimedia conferencing standards such as H.320 and H.324 can be simplified. H.323 only uses a subset of the Q.931 messages in ISDN and a subset of the information elements (IEs). All the H.323-related parameters are encapsulated in the user-user IE (UUIE) of a Q.931 message.

A zone can span a number of LANs in different locations, or just a single LAN. The only requirement is that each zone contains exactly one gatekeeper, which acts as the administrator of the zone.



■ **Figure 2.** Protocol relationships in H.323.

- **H.245:** H.245 is used for connection control, allowing two endpoints to negotiate media processing capabilities such as audio/video codecs for each media channel between them. It is a common protocol for all H-series multimedia conferencing standards, including H.310, H.320, and H.324, and contains detailed descriptions of many media types. In the context of H.323, H.245 is used to exchange terminal capability, determine master-slave relationships of endpoints, and open and close logical channels between two endpoints.
- **Real-Time Transmission Protocol:** RTP is used as the transport protocol for packetized VoIP in H.323. It is adopted directly from IETF and is usually associated with Real-Time Control Protocol (RTCP).

Figure 2 summarizes the relationship of various protocols involved in H.323.

When GKs are used within the network, an H.323 call generally goes through seven phase, shown in Table 1. The first three phases correspond to call setup. The last three phases correspond to call teardown. When no GK is involved, phases 1 and 7 are omitted. For simple VoIP

calls, H.323 defines *fast connect* which reduces the seven phases of a call by combining the Q.931 and H.245 phases.

Two call control models are supported in H.323: direct call and GK-routed call, as shown in Fig. 3.

In the direct call model, all Q.931 and H.245 signaling messages are exchanged directly between the two endpoints; so are the RTP media streams. As long as the calling endpoint knows the transport address of the called endpoint, it can set up a direct call with the other party. This corresponds to the early PC client model, using IP as transport for free Internet phone calls. The GK cloud and RAS channels are optional. When GKs are present, the calling TE (TE1) may request address resolution service from its GK, and the called TE (TE2) may ask for permission from its GK to accept the call. This model is unattractive for large-scale carrier deployments because carriers may be unaware of calls being set up, which may prevent them from providing sufficient resources for the call and charging for it.

In the GK-routed call model, all signaling messages are routed through the GK cloud. In this case, use of RAS is necessary. This model allows endpoints to access services provided by the GK cloud, such as address resolution and call routing. It also allows the GKs to enforce admission control and bandwidth allocation over their respective zones. This model is more suitable for IP telephony service providers since they can control the network and exercise accounting and billing functions.

### H.323 SERVICES

H.323 v. 1 only defined the basic call control and signaling for setting up multipoint multimedia conferences and did not address enhanced services. To enable enhanced services on top of H.323, ITU-T SG16 created the H.450 series which specify supplementary services similar to features available in the PSTN. We will first describe how H.450 can be used for implementing services, and then discuss how services can be implemented without the use of H.450.

Phase	Protocol	Intended functions
1	Call admission	RAS
2	Call setup	Q.931
3	Endpoint capability negotiation and logical channel setup	H.245
4	Stable call	RTP
5	Channel closing	H.245
6	Call teardown	Q.931
7	Call disengage	RAS

■ **Table 1.** The seven phases of an H.323 call.

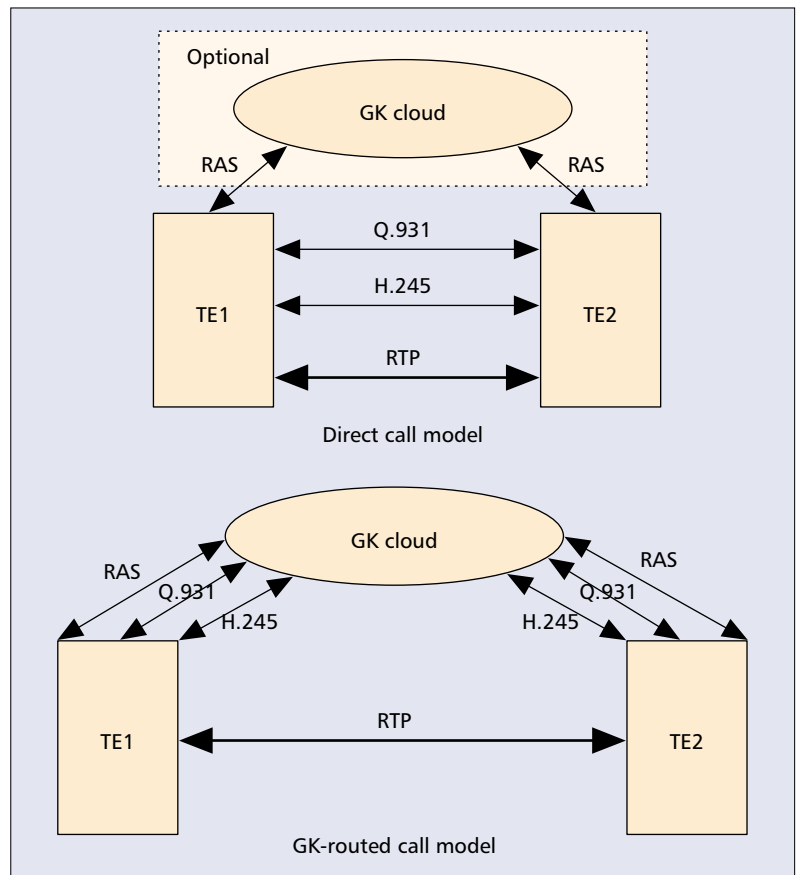
**H.450-Based Services** — In H.323 v. 2, three H.450 Recommendations were ratified: H.450.1 for generic functional protocol and procedures; H.450.2 for call transfer; and H.450.3 for call diversion, including various flavors of call forwarding and deflection. In H.323 v. 3, which was approved in September 1999, five more supplementary services are defined: H.450.4 for call hold; H.450.5 for call park and pickup; H.450.6 for message waiting indication; and H.450.7 for call waiting. Currently, ITU-T SG 16 is working on H.323 v. 4, which will include five more supplementary services: H.450.8 for name identification; H.450.9 for call completion; H.450.10 for call offer; H.450.11 for call intrusion; and H.450.12 for additional common information network services.

H.450.1 defines a generic functional protocol on top of Q.931 for all supplementary services. It also defines the control procedures for the TEs involved in handling the protocol messages. The functional protocol defined in H.450.1 is an end-to-end signaling protocol, derived from the QSIG protocol for interconnecting private branch exchanges (PBXs). In this sense, supplementary services in H.323 can be viewed as the adaptation of PBX services to the IP domain. Since H.450 is an end-to-end protocol, it requires that both TEs understand the service logic in order to make a supplementary service work. This functional model assumes that the TEs execute most of the service logic. This is a departure from traditional PSTN enhanced services, where the service logic resides in the switches, not the endpoints (phones), and poses significant problems for large-scale deployments where TEs may support different releases of H.323.

Each H.450.X with X larger than 1 defines a supplementary service application protocol for a specific service. A supplementary service application protocol data unit (SS-APDU) is encapsulated in the UUIE of a Q.931 message as the `h4501SupplementaryService` parameter. For example, H.450.3 specifies the call diversion services, which includes call forwarding unconditional, call forwarding busy, call forwarding no reply, and call deflection. These services roughly correspond to various call forwarding features in the PSTN. For each service, a set of procedures and the corresponding message flows are defined, such as activation, deactivation, interrogation, registration, and invocation.

**Non-H.450-Based Services** — The disadvantage of H.450-based services is that new specifications need to be developed by the ITU, and TEs may need to be upgraded before a service is deployed. This slows down deployment of new services, an undesirable feature in the VoIP environment. There are alternatives to H.450-based services that carriers can use to deploy services. Services can be implemented inside GKs. When the GK is used for address resolution, there are many services, such as mobility services, that can be offered to customers. The GK-routed call model allows carriers to introduce more advanced services.

Most services implemented inside the GK are implemented in a proprietary manner. This

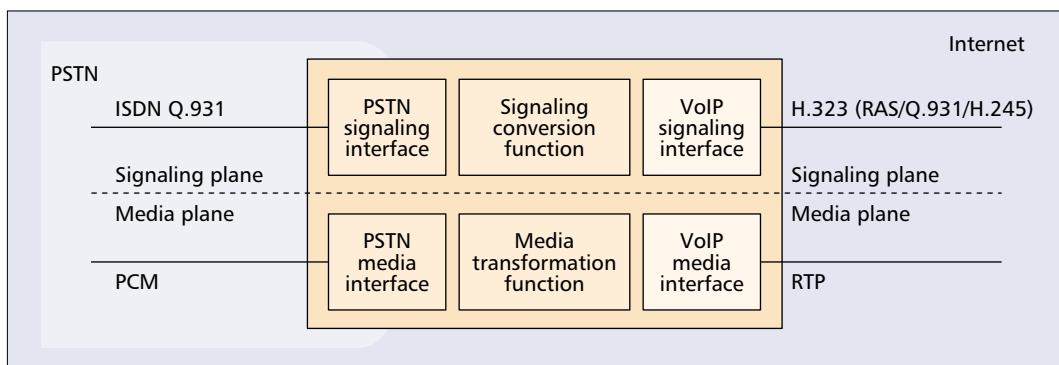


■ **Figure 3.** The direct and gatekeeper-routed call models of H.323.

is similar to what happened initially in the PSTN environment. Eventually, as the VoIP industry matures, a more standardized approach will become important. Intelligent network (IN) was introduced in the PSTN to standardize the development of services and locate service logic in a separate platform. There are already discussions of integrating IN with GKs, and there are some GK products that provide some support for IN-based services. ITU-T SG16 began standardizing this work in August 1999 as Annex D of H.246. However, due to lack of contributions, the work is progressing very slowly. It is unclear at this point whether an IN approach will be the final solution to standardizing development of VoIP services, or another alternative will emerge.

Realizing the limitation of H.450, ITU-T SG16 initiated two new work items in 1999 for version 4. One is to introduce an HTTP-based control channel for H.323 devices so that a service provider is able to display web pages to the user with H.323 call-related contents. This is addressed in Annex K of H.323, which provides a new way to create new services using a mechanism similar to third party call control. The other work is to provide a new “stimulus-based” control mechanism for H.323 systems so that a relatively simple H.323 endpoint can rely on the intelligence residing in the network elements such as feature servers. This is addressed in Annex L of H.323, which utilizes the “package” concept introduced in MGCP or H.248 for endpoint capability customization. In effect,

Even though H.323 was designed for multi-point multimedia conferencing over packet networks, its usage has been primarily driven by VoIP applications, and interworking with PSTN has been a focus from the very beginning.



■ **Figure 4.** Components of a gateway bridging the PSTN and the Internet.

Annex L creates a class of H.323 devices whose intelligence lies between a dumb residential GW as used in MGCP or H.248 and a full H.323 endpoint. It represents a departure from the end-to-end H.323 architectural principle. Both Annexes are scheduled for decision in November 2000.

## H.323 INTERWORKING WITH THE PSTN

Even though H.323 was designed for multipoint multimedia conferencing over packet networks, its usage has been primarily driven by VoIP applications, and interworking with the PSTN has been a focus from the very beginning. Unlike SIP, the GW to the PSTN has been an integral part of the H.323 architecture.

Interworking with PSTN usually concerns three call setup scenarios: H.323 TE to phone; phone to H.323 TE; and phone to phone via intermediate H.323 networks. In all cases, an H.323 GW is involved in connecting the PSTN with the Internet. Generally speaking, a GW needs to provide the following functionality, as depicted in Fig. 4.

- **PSTN interfaces:** This function includes the PSTN signaling interface that terminates signaling protocols such as ISDN Q.931, and the PSTN media interface that terminates media streams such as pulse code modulation (PCM) voice streams.
- **VoIP interfaces:** This function includes the VoIP signaling interface that terminates H.323 (including RAS, Q.931 and H.245), and the packet media interface that handles RTP.
- **Signaling conversion:** This function typically translates between ISDN Q.931 signaling and H.323 signaling for call control.
- **Media transformation:** This function typically translates between the 64 kb/s PCM streams and RTP streams of various speeds.
- **Connection management:** A major function implied by the above diagram is that a GW must internally coordinate between signaling flows and media transformations. This involves creating, modifying, and deleting the association between the PSTN and Internet flows during the lifetime of a call.

In 1998, as carriers gained experience with VoIP and got ready to move from small-scale to

large scale deployments, they realized that H.323 gateways have the following limitations:

- **Scalability:** The maximum number of lines an H.323 GW can support is a few thousand. This is small when compared to a regular telephone switches with tens of thousands of lines.
- **SS7 connectivity:** Until the end of 1998, all H.323 GWs on the market did not have this capability and connected to switches via ISDN trunks. Without SS7 connectivity, VoIP cannot provide the same rich set of services enabled by SS7.
- **Availability:** When a GW is down, all active calls through the GW disappear. There was no mechanism in H.323 for failover.
- **User friendliness:** Most VoIP services require that a subscriber dial the phone number to connect to the GW and then dial the number of the destination of the call. This procedure is called *two-stage dialing*. This is largely a result of the lack of SS7 connectivity.

The fundamental factor limiting the number of lines an H.323 GW can handle is the monolithic packaging of signaling and media transformation into one box. The number of lines in a GW is determined by the number of simultaneous calls it can handle, which is limited by its CPU processing power and memory capacity. The signaling and media transformation functions have very different processing requirements. Generally, signaling is less computationally intensive, mostly involved in call setup and teardown. Media transformation is much more computationally intensive because low-bandwidth codecs employ sophisticated algorithms, and GWs may have to apply echo cancellation and silence compression on the media for each active call. Media transformation also occurs through almost the entire duration of a call. From the above analysis, it was concluded that separating the signaling and media transformation functions would allow for more scalable GWs. This is the idea behind GW decomposition, discussed in detail next.

## FUNCTIONAL DECOMPOSITION OF H.323 GATEWAYS

We illustrate the idea of functional decomposition in Fig. 5, which was used as an earlier reference model by the European Telecommunications Standards Institute (ETSI) TIPPHON. ETSI



An H.323 GW (bounded by the dotted line in Fig. 5) is decomposed into three functional components:

- Let's see how functional decomposition of GWs overcomes the deficiency of monolithic H.323 GWs in carrier-grade VoIP deployments:

- 
- ```
graph LR
    GK1[GK] -- D --- GK2[GK]
    GK1 -- A --- H323[H.323 terminal]
    GK2 -- G --- BackEnd[Back-end]
    GK2 -- C --- MG1[Media GW]
    BackEnd -- F --- MG1
    MG1 -- J --- SGW[Signaling GW]
    MG1 -- N --- MG2[Media GW]
    H323 -- B --- MG2
    MG1 --- E1[E.a]
    SGW --- E2[E.b]
```

## H.323 AND OTHER VOIP SIGNALING PROTOCOLS

## H.323 AND MGCP

Media gateway control protocols may not just complement H.323 but also present an alternative to H.323 altogether in VoIP deployments. Some of the terminals being considered today for VoIP deployment include cable and DSL modems that may have limited computing resources. For those types of devices it may be more appropriate to use MGCP to provide VoIP signaling instead of H.323. The MGCP architecture assumes that most of the intelligence is inside the network and that customer premises equipment (CPE) has limited functionality which reduces the cost of those devices [3]. New services can be introduced without requiring any CPE upgrades and handled by simply upgrading the centralized software that contains the intelligence for implementing services. Services can be made available to all customers willing to pay without requiring that the customer download and install any new software. In addition, MGCP does not allow terminals to make calls directly to other terminals, which allows carriers to control QoS and charge appropriately.

147

*H.323 has the advantage that some vendors and carriers have made significant investments in H.323 and that there are significant deployments based on H.323 already. We expect that in the short term both protocols will co-exist. Thus, interworking of H.323 and SIP is being considered.*

has adopted MGCP as the interim network-based call signaling standard for cable modems supporting VoIP. Several cable and DSL modem VoIP products have also adopted MGCP. The use of H.323 vs. media gateway control protocols for terminals will depend on whether carriers and their customers prefer to deploy intelligence mostly at the terminal or mostly inside the network. It is conceivable that both models will coexist the same way answering machines and voicemail services coexist in today's telephone network.

### H.323 AND SIP

SIP has gained significant momentum recently as an alternative to H.323. Several companies have participated in SIP interoperability events in the past year or so, and carriers are considering basing their VoIP deployments on SIP. SIP has attracted a lot of attention because of its simplicity and ability to support rapid introduction of new services. Architecturally, SIP has some similarities with H.323 but is much more lightweight. When IETF defined SIP it did not adopt Q.931 or H.245, which made SIP much simpler than H.323. SIP did adopt the model where a lot of the intelligence may be at the terminal like H.323. SIP is such a simple protocol that inexpensive terminals based on SIP may be developed.

H.323 has the advantage that some vendors and carriers have made significant investments in H.323, and there are significant deployments based on H.323 already. We expect that in the short term both protocols will coexist. This is the reason why interworking of H.323 and SIP is being considered [5]. There are also products in the market today called *call agents* or *soft switches* that support H.323, SIP, and MGCP, and allow terminals supporting any of these protocols to place VoIP calls to other terminals regardless of the signaling protocol the terminal supports. Eventually market forces will determine whether all these VoIP signaling protocols will need to be supported.

### SUMMARY

H.323 was the first VoIP standard that helped move the VoIP industry away from proprietary solutions and toward interoperable products. The H.323 architecture is still evolving in several areas such as the gateway decomposition archi-

ture and integration of H.323 with IN. This evolution is addressing some of the original limitations of H.323. Other protocols such as SIP have also been introduced as alternatives to H.323. It is unclear how the VoIP signaling architecture will eventually evolve, but it is clear that these different signaling protocols will need to coexist for some time. The industry debate on the VoIP signaling architecture will continue to attract a lot of attention, and the evolution will be determined by the VoIP market forces.

### REFERENCES

- [1] H. Schulzrinne *et al.*, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 1889, Jan. 1996.
- [2] M. Handley *et al.*, "SIP: Session Initiation Protocol," IETF RFC, Mar. 1999.
- [3] C. Huitema *et al.*, "An Architecture for Internet Telephony Service for Residential Customers," *IEEE Network*, May/June 1999.
- [4] ITU-T Rec. H.323, "Packet Based Multimedia Communications Systems," Feb. 1998.
- [5] H. Agrawal, "SIP-H.323 Interworking Requirements," IETF draft, work-in-progress, July 2000.

### BIOGRAPHIES

HONG LIU (lhong@research.telcordia.com) is a senior research scientist in applied research, Telcordia Technologies. He got his Ph.D. degree in computer science in June 1996 from the University of Maryland at College Park. Since he joined Telcordia Technologies (then Bellcore), he has been involved in various research and development projects, including feature interaction detection for advanced intelligent networks, call signaling and control for voice over IP, VoIP gateway decomposition and control protocols, service interworking in next generation networks, and naming and addressing in converged networks. He has published more than 15 technical papers in various technical conferences and journals. Since 1998 he has been the principal representative for Telcordia Technologies in ITU-T SG16, IETF, and ETSI TIPHON for VoIP signaling protocol standardization. He is one of the major contributors to gateway control protocols SGCP and MGCP, and is an active contributor in ITU-T SG16 and IETF MEGACO WG in the standardization of MEGACO/H.248. He has also done extensive consulting to the telecommunication industry promoting VoIP technologies, and is a frequent speaker in VoIP industry fora.

PETROS MOUCHTARIS (pmouchta@telcordia.com) is executive director of Internet Services Research at Telcordia Technologies. His organization is focusing on services over packet-based networks such as Internet telephony and IP VPNs, and service management issues associated with those services. His interests include signaling, multimedia services, and network control. In the past he has managed organizations at Pacific Bell and Oracle. He received his M.S. and Ph.D. in electrical engineering from the California Institute of Technology and his Diploma in electrical engineering from the National Technical University of Athens, Greece.