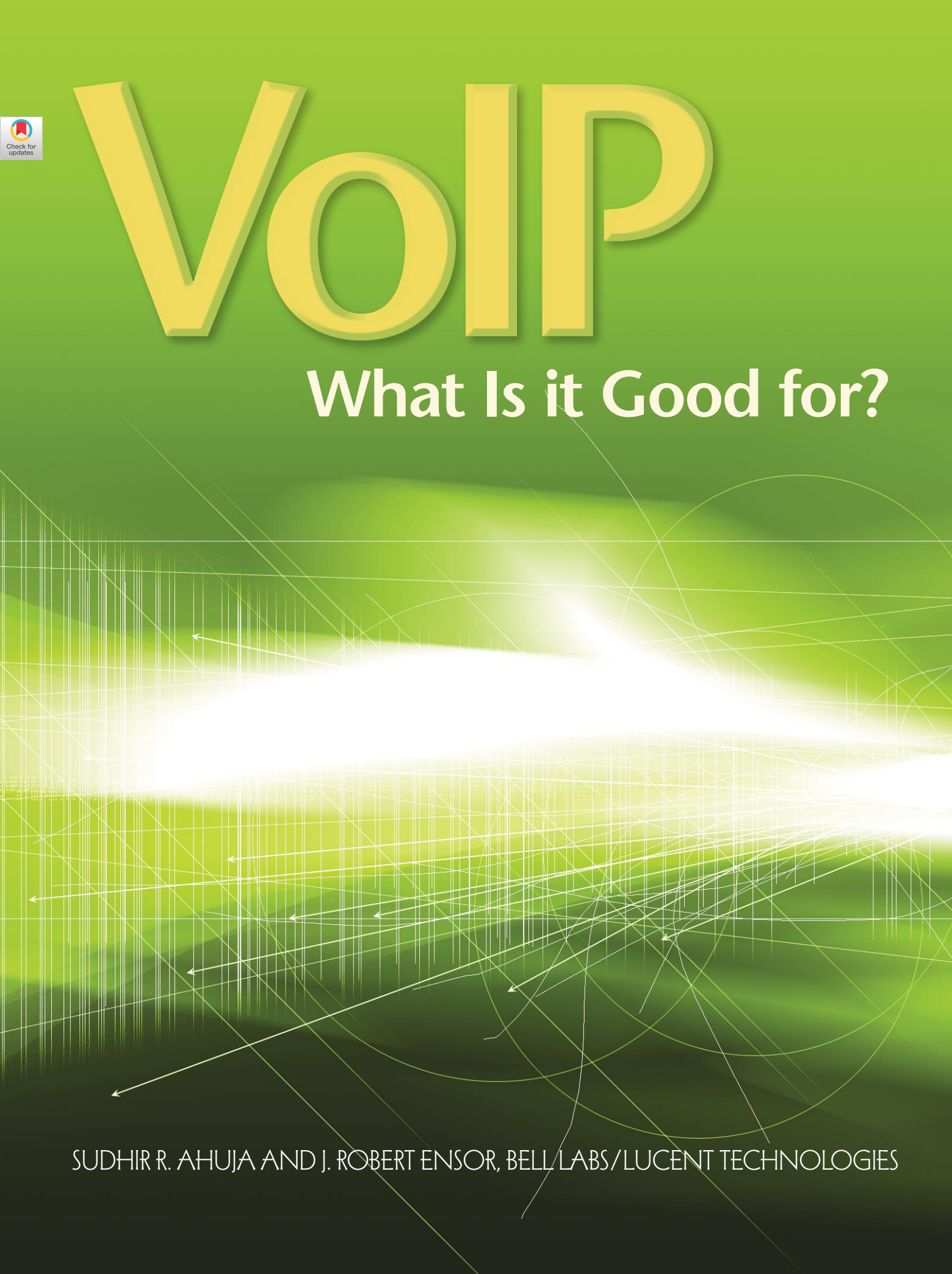
VoIP

What Is it Good for?



SUDHIR R. AHUJA AND J. ROBERT ENSOR, BELL LABS/LUCENT TECHNOLOGIES

If you think VoIP is just an IP version of telecom-as-usual, think again. A host of applications are changing the phone call as we know it.

VoIP (voice over IP) technology is a rapidly expanding field. More and more VoIP components are being developed, while existing VoIP technology is being deployed at a rapid—and still increasing—pace. This growth is fueled by two goals: decreasing costs and increasing revenues.

Network and service providers see VoIP technology as a means of reducing their cost of offering existing voice-based services and new multimedia services. Service providers also view VoIP infrastructure as an economical base on which to build new revenue-generating services. As deployment of VoIP technology becomes widespread and part of a shared competitive landscape, this second goal will become more important, with service providers working to increase their market bases.

Most current and envisioned VoIP services are so-called converged services, integrating features and functions from multiple existing services. Often, features from conventional voice-based telephony services are combined with those found in data network services. For example, click-to-dial services allow users to control telephone calls from Web browsers running on their personal computers. Converged services may also provide users with new media integration. For example, multimedia conference services allow users to interact with each other through calls in which they exchange both audio and video information (i.e., new versions of videophones).

VoIP What Is it Good for?

The growing opportunities for converged telephony-Web services are motivating convergence of telephony and data networks. VoIP services are also driving another network convergence: integration of wireless and wireline networks. More general network convergence seems likely. Because IP networks can be relatively inexpensive, network providers are encouraged to build common IP core networks surrounded by various access networks. These access networks (wireless, wireline, cable, etc.) can share the IP core resources, and thus reduce the costs of providing common services to customers with different access devices.

Many engaging VoIP services are already available, and service providers are planning even more exciting services. Continued deployment of IP networks and IP endpoint devices will enable further development of new services. Also, as the processing capacity of IP endpoints increases—allowing them to deal directly with network access controls, multiple data formats, and transformations—more innovative and convenient services will become possible. This article introduces some noteworthy services that are being deployed today and highlights a few of the interesting future services.

CREATING NEW SERVICES

Conventional telephony services—those available to customers through the public switched telephone network (PSTN)—are built upon a highly structured technology base. This base was created and optimized to support voice calls using analog telephones. The base provides application developers with integrated signaling/media transport (in-band signaling) and a limited set of signal handlers and media processors, which are isolated from other networks through their switched circuit connections. Since telephones support very limited signaling mechanisms, invocation and control of PSTN services have been awkward. Some services are invoked by dialing special phone numbers such as 800 or 900 numbers. PSTN services are often invoked and controlled through in-band signaling, which is typically activated through touch tones (DTMF, dual-tone multi-frequency) or voice (IVR, interactive voice response).

Fundamental control and media handling needed by PSTN service providers must be performed by special network elements (signaling control points, service nodes, etc.).

Figure 1 illustrates key components of a call-center service. In this figure, the 800 server is a service node; it is an application server that communicates with the class 5 and 4 switches via SS7 (Signal System 7) signaling protocols. This server deals only with control messages and not with voice itself. It helps establish the final route for the voice call based on the features it has implemented. For example, it can determine whether a call is routed to a company's call center or to one of its retail outlets.

Service providers may require control and media

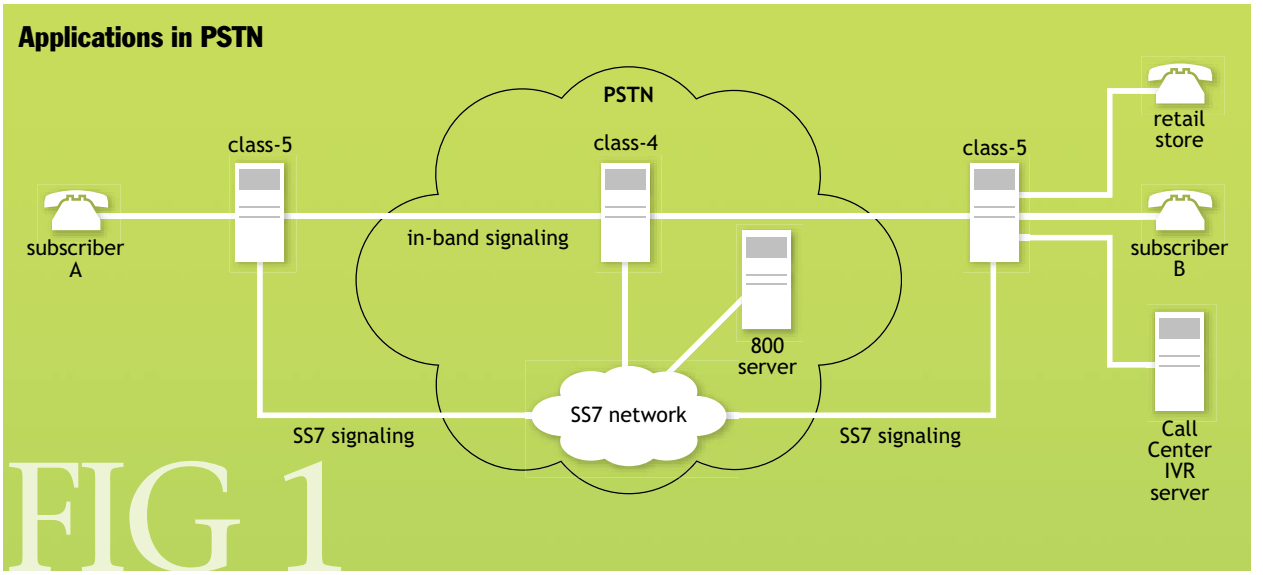


FIG 1

processing not supported by network elements. This additional processing must be handled at call endpoints. The flow of information into and from an endpoint is through the voice channel itself, and therefore specialized controls must be built on audio controls (e.g., conversations with human operators, DTMF, or IVR). In figure 1, these endpoint application servers are represented by the call-center IVR server, which terminates voice connections and communicates via in-band signaling using DTMF or voice recognition.

VoIP technology provides richer, more flexible foundations for building communication services. IP networks support independent connections for signaling and media traffic. This decoupling of signal and bearer traffic eliminates interference between the information flows; in-band signaling is not required. Thus, communication with application servers is simplified.

In addition, IP network topology allows any node to act as a server. Therefore, multiple application servers and user endpoints—located in one or several service provider domains—can communicate via IP to participate in service support.

Finally, IP transport is provided by various underlying networks, and different network technologies can support different sets of services. For example, DSL and cable networks provide broadband IP connections that support realtime voice, data, and video services. Hence, these network providers can offer "triple-play" services to their customers.

Figure 2 shows how to implement a call-center service using VoIP technology. In the figure, user endpoints are telephones (not IP-based devices) attached to wireline or wireless access networks. An IP backbone interfaces to these specific access networks through border elements (e.g., media gateways). These gateways terminate voice calls for the users; they handle all TDM (time division multiplexing) voice traffic to and from users. The gateways recognize DTMF signals from the users and convert them to SIP (session initiation protocol) messages for the IP-based application servers. In addition, they convert between the users' TDM voice payload and RTP (realtime transport protocol) media packets, which are used by the media processors. Several IP-based application servers work in concert, coordinating their activities through SIP signaling to provide the call-center service. The softswitch contains a SIP proxy to support this SIP coordination, and it contains media control functions to support coordination of media processing. The application servers may be geographically distributed and separated from endpoints and switches. For example, Web sites can use stored voice or music files to provide announcements. They can act as music-on-hold servers; a single announcement server is not required.

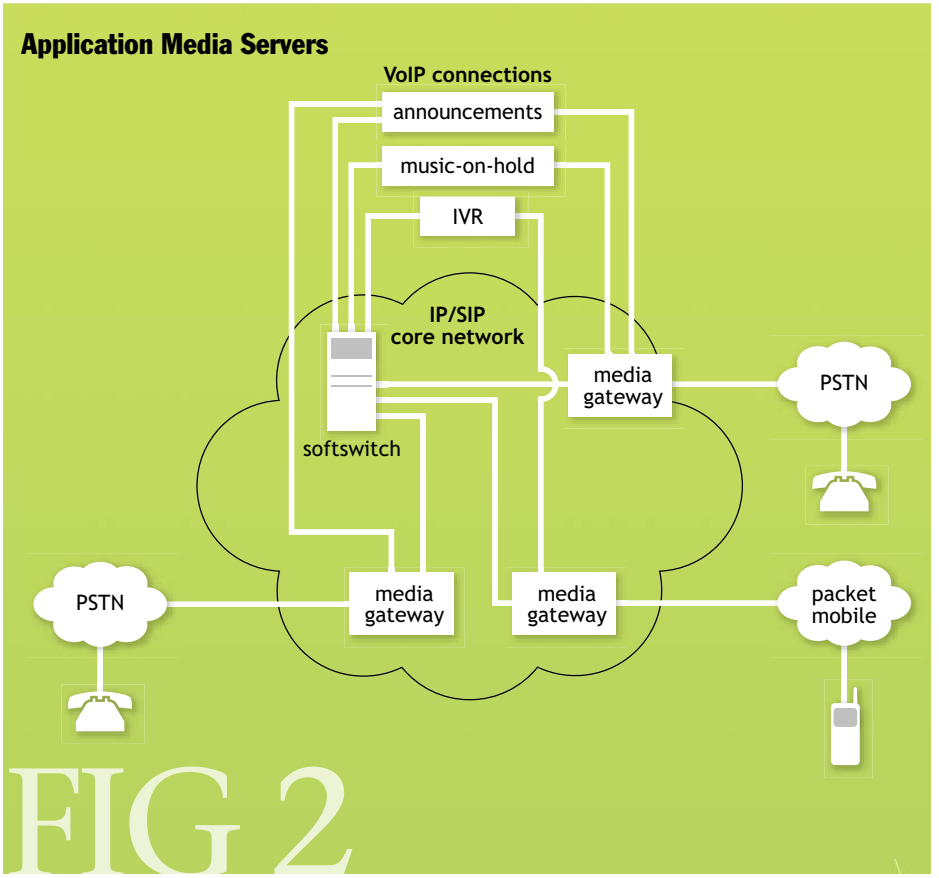


FIG 2

VoIP technology provides a foundation for creating many new converged services through different combinations of components. For example, IVR and Web components can combine—using SIP as a common signaling protocol—to create call-center services with access from Web browsers or IP phones, as well

VoIP What Is it Good for?

as voice-only telephones. Similarly, IVR servers and SMS (short message service) can combine to create call-center services that include SMS messages. Users will be able to access these call-center services via any of their access mechanisms or even simultaneously use multiple access technologies to provide better service. Alternatively, SMS systems can combine with Web-based information servers to create MMS (multimedia message services) in which messages may contain Web-based information and be retrieved by Web browsers.

NEW CONTROLS AND COORDINATION

Converged services can employ features from one set of services to control aspects of other sets of services. For example, click-to-dial services combine Web-based user interfaces with telephony servers to create Web-controllable phones. These services allow users to select (highlight) phone numbers embedded in Web pages, indicating that these numbers should be called. Such services are built by combining the PSTN, IP networks, and IP-based servers.

Figure 3 shows how a typical click-to-dial service works. When customers use their Web browsers to click on a telephone number within a Web page, their computer sends a message over a packet network to an IP-based click-to-dial server. This server, in turn, uses its connections to the PSTN to make telephone calls to the customer and to the number that customer is dialing. These calls are then bridged into a single call by a PSTN control element.

This example illustrates an important characteristic of VoIP services: they can be made as collections of multiple servers. These servers typically base their coordination on SIP signaling. SIP, however, provides a means only to locate and synchronize the initial interaction among the appropriate servers. Once the servers have rendezvoused through SIP, they must then exchange application-specific signaling through appropriate specialized protocols. In this example, the click-to-dial client and the Web

What Is SIP?

SIP (session initiation protocol) is a text-based protocol for initiating communication sessions between users. These sessions may include calls with conventional telephones, voice, video, and data calls, multimedia conferencing, streaming media services, games, etc. SIP is defined by a collection of Requests for Comment managed by the Internet Engineering Task Force (IETF).

SIP messages are exchanged among two or more peers (IP nodes) for rendezvous and synchronization, thus supporting initiation of interactive communication sessions.

Once communicating parties have started their session through SIP messages, they are able to conduct the session through session-specific message exchange. These parties may also use SIP for additional session events, such as adding and dropping session members, changing media, and ending sessions.

SIP is fundamentally a protocol for communication among peers. SIP sessions are conducted by two or more communicating parties. These parties may be network endpoints—IP nodes associated with end-user devices—as well as network servers. If one SIP node knows the address of another node, the first may invite the second to join a SIP session. Thus, SIP sessions do not require support from network servers, but network intermediates typically help endpoints find one another. Users register their network addresses with SIP registrars. Users usually send session invitations to one another through SIP proxies, which use registration information to locate invitees.

SIP sessions provide an extensible framework for a wide variety of interactions. They do not define—hence, do not constrain—specialized service behavior. Thus, they form the basis for many different communication services. SIP sessions support services typically accessed through packet data networks (e.g., streaming video-on-demand service). They also support conventional telephony services (e.g., conference voice calls).

Because SIP is a framework in which both telephony and nontelephony services have been developed, SIP has encouraged convergence of services. In particular, SIP is encouraging convergence of telephony and Web-based services. These converged services include Web phones, Web-based management of telephony services, and interactive games in which players can talk with one another in conference calls.

Additional information is available from the SIP working group of the IETF at http://www.ietf.org/html.charters/sip-charter.html.

server must exchange agreed-upon protocols (typically including HTTP) so that Web pages can be transferred to the user. Also, the click-to-dial client and the click-to-dial server must exchange an agreed-upon protocol to request and control the required telephony functions.

Service coordination and composition become important issues in the development and execution of VoIP services, as multiple application servers are often involved. The industry must develop techniques to coordinate distinct service elements within sessions. One fundamental problem is that service behavior is difficult to describe both formally and conveniently, which makes service coordination labor-intensive. A related problem is that the multiple servers used to create a service might not be in the same network. Therefore, one service provider might not be willing to publish details of its server for another provider. Another difficulty is that services can interfere with each another. For example, if a conference participant temporarily leaves, generating music on hold, this behavior can interfere with or even block continuation of the conference by the remaining participants.

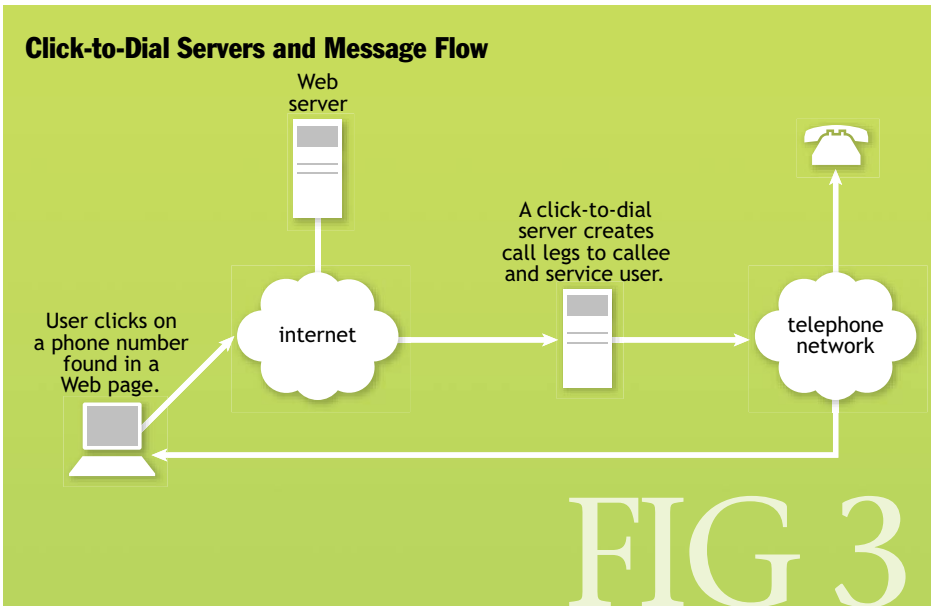
NEW MEDIA INTEGRATION

Many VoIP services are based on integration of multiple media. One such service is multimedia conferencing, which can be implemented by taking advantage of both SIP signaling and IP transport. SIP messages are available for server registration and rendezvous, as well as the controls that are needed to set up, conduct, and end sessions. Additional IP control messages are used to send media-specific commands. For example, service customers can use these commands to select video feeds, change codecs, change multicast groups, etc. IP transport is used to move the data representing the various media to and from servers and among users.

Figure 4 illustrates a conferencing service. Similar in overall structure to the IVR service depicted in figure 3, this system is based on a different set of servers: a multimedia conference server, an audio bridge, video server, and data-sharing server. The conferencing server coordinates the activities of the data-specific servers, which manipulate different sets of packet data corresponding to appropriate media. For example, the audio bridge receives encoded voice from all participants and distributes combined voice data back to the participants. As the figure illustrates, uniformity of endpoint devices is not required—each customer can participant in a conference through a different type of endpoint—e.g., cellphone, analog phone, or laptop. The media transmitted to/from each participant depends upon the capabilities of the participant's endpoint device.

QoS (quality of service) is an important issue for IP-based multimedia services. Many current IP services have been deployed without QoS guarantees from underlying network providers. These services are successful because transport quality is sufficient to meet customer demands. Providers of these services, however, do not have assurances that their services can grow to meet the needs of larger customer bases while also meeting time constraints for the services. For example, IP-based voice and video services are being deployed in enterprises without explicit QoS support. Since the enterprise LANs used for transport have enough bandwidth to allow over-provisioning for realtime voice, and video, these services are successful. Timely transport of time-sensitive data, however, to support realtime multimedia conversations across worldwide networks, is harder to ensure.

We must solve these problems by using adequate transport performance and servers within the signaling and media transport paths that can react to messages within realtime constraints. These servers must process both



VoIP What Is it Good for?

signaling and bearer traffic within time bounds to meet processing needs associated with transcoding, composition, distribution, etc. Currently, servers capable of this processing are economical only for certain functions.

NEW USES OF SESSIONS

SIP sessions can be long-lived, and persistent sessions provide the foundation for some interesting new VoIP services. One example is an enhanced chat-room service, called Telechat, illustrated in figure 5.

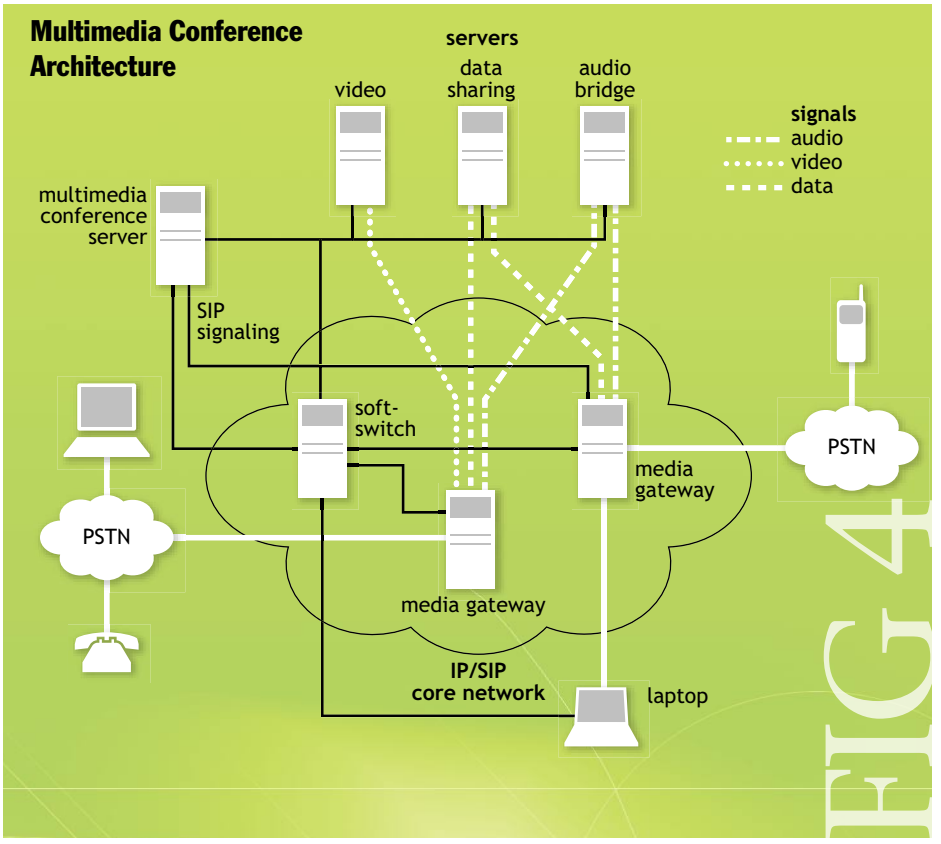
In this application users can interact through voice, video, and data during multimedia conferences. They can also exchange private and public (broadcast) messages. Users can create and access stored data in a shared repository. The data can be imported from other applications, generated during chat sessions, and accessed during or outside of multiparty conferences. Service sessions are not restricted to calls, so they can be long-lived, extending over multiple calls or over other, shorter sessions. These longer sessions can form the basis for persistent state and data storage.

Persistent sessions support long-term interactions—and can serve as the rendezvous point for multiple calls. In addition, a persistent session can provide storage for data used in these calls. Hence, a persistent session can act as a direct representation for a long-term group effort. Enhanced chat-room services can be built upon persistent sessions, which can maintain a room state that is stable over the span of several chat sessions. This persistent state creates a context or surrounding environment for a series of chat sessions.

Persistent sessions create new challenges for system designers. Developers must decide where to maintain session state, which can be distributed among network servers and endpoints or restricted to subsets of these elements. Designers must also decide where to store the data associated with the sessions. In Telechat, for example, session state is stored on multiple servers. In a related issue, service providers must decide who owns what data. Billing for the resources needed to store persistent state is also a source of several design decisions. For example, service providers must specify whether a person who joins a long-term session pays for the session or pays for the connection/interaction with the session.

ONLY THE BEGINNING

VoIP is a disruptive technology that is causing significant change in the way voice communication services are delivered. It is providing future roadmaps for telecom networks. This is only the beginning of a more significant move to convergence. As the world moves to a common IP-based data network as backbone, VoIP is only one of the realtime



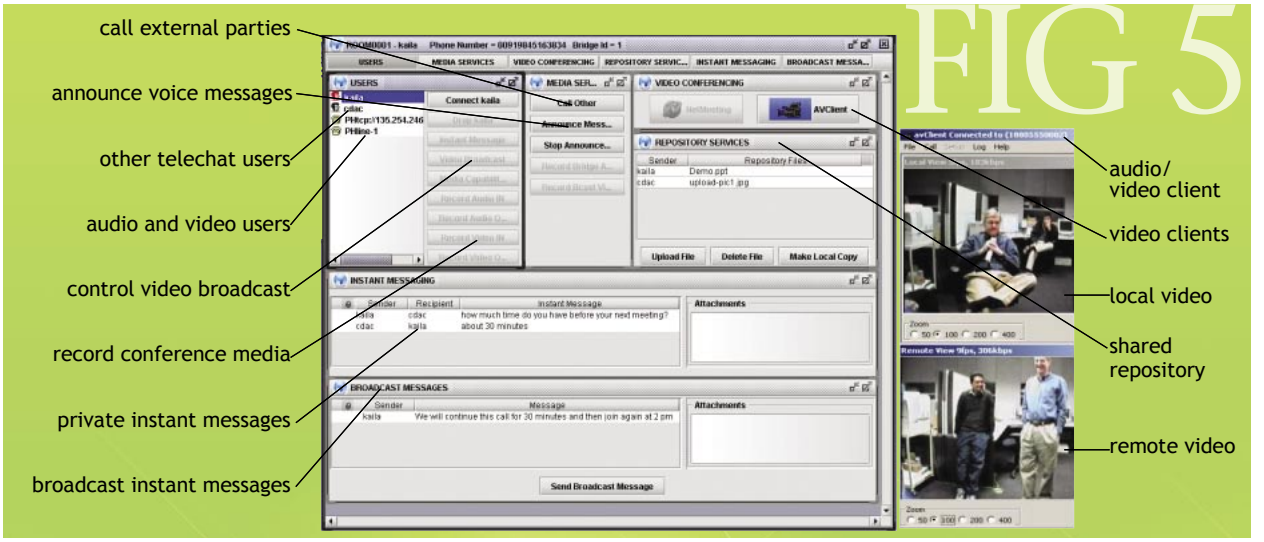


FIG 5

services offered on such networks, along with many data services. The same network will also support video services from videoconferencing to entertainment video.

More important, these services allow convergence at the control and user levels. A user can initiate a call or TV program from the Web and then send a video from a camera phone to the user's home Web site. Common Web-based services can be used for provisioning the user's personal choices. Clearly, this is only the beginning of exciting services offered by full multimedia on IP.

An important architectural change is that all application servers will move out of specific networks and become more access-independent. Networks will become multiservice platforms. To do this effectively, networks have to provide flexible QoS mechanisms and the ability to create virtual networks to match the services being deployed. This is where many of VoIP challenges remain to be solved. Specifically, we still need ways to specify network requirements of a particular application (e.g., multiparty audio-conferencing) and we need to be able to map that to the multiservice network. Finally, we need to be able to provision such services and monitor their execution to guarantee delivery.

Last, but not least, is the challenge of integrating the ever-smarter endpoint and endpoint-based applications with the network-centric view presented earlier. Besides new service interaction issues, this raises many new concerns about ownership of the user's data, authentication, billing for services, and responsibility for security.

VoIP is here and already leading the way not just to cheaper voice calls but also to a host of new applications. We need to focus on the challenges to enable a host of new multimedia applications. Q

LOVE IT, HATE IT? LET US KNOW

feedback@acmqueue.com or www.acmqueue.com/forums

SUDHIR AHUJA is vice president of the Converged Networks and Services Research Laboratory at Bell Labs/Lucent Technologies, where he is leading research in converged networks, services, speech recognition, text-to-speech coding techniques, video-based communication, and novel multimedia applications. He designed and developed the first large-scale multiprocessor at Bell Labs and championed the first Internet-based video conferencing system. His current interests are in the field of communication applications over the Internet.

Ahuja obtained his M.S. and Ph.D. degrees in electrical engineering from Rice University. His undergraduate education was at the Indian Institute of Technology, Bombay, where he received the President's Gold Medal for outstanding academic performance. He is a Fellow of Bell Labs and has served as chairman for the Multimedia Services and Terminals Committee of the IEEE Society, area editor for the IEEE Communications Committee, and editor for *Transactions on Networking*, a joint publication of IEEE and ACM.

BOB ENSOR is a technical manager in the Services Infrastructure Research Department at Bell Labs/Lucent Technologies. He leads research and development efforts in next-generation network architectures and components. Earlier, he served as principal researcher in several projects at Bell Labs, including broadband service data centers, multimedia messaging systems, shared virtual worlds for the Internet, and multimedia conferencing systems. Ensor holds several patents and has published numerous papers. He received his Ph.D. in computer science from SUNY at Stony Brook.