Internet Telephony

0.1 Introduction

The International Engineering Consortium (IEC) describes Internet Telephony as follows:

> Internet telephony refers to communications services – voice, facsimile, and/or voice-messaging applications – that are transported via the Internet, rather than the public switched telephone network (PSTN). The basic steps involved in originating an Internet telephone call are conversion of the analog voice signal to digital format and compression/translation of the signal into Internet protocol (IP) packets for transmission over the Internet; the process is reversed at the receiving end.

More technically, Internet telephony is the real-time delivery of of voice and possibly other multimedia data types between two or more parties, across networks using the Internet protocols, and the exchange of information required to control this delivery.

The terms Internet telephony, IP telephony and voice-over-IP (VoIP) are often used interchangeably, while some people consider Internet telephony that subset of voice-over-IP (or IP telephony) that travels over the public global Internet as opposed to private IP-based packet networks.

Neither term is fully satisfactory since it implies voice service only, while one of the strengths of Internet telephony is the ability to be *media-neutral*, that is, almost all of the infrastructure does not need to change if a conversation includes video, shared applications or text chat.

Voice services can also be carried over other packet networks, without an mediating IP layer, for example, voice-over-DSL (VoDSL) Ploumen and de Clercq [2000] for consumer and business DSL subscribers, and voice-over-ATM (VoATM) for carrying voice over ATM Wright [1996, 2002], typically as a replacement for inter-switch trunks. Many consider these as transition technologies until VoIP reaches maturity. They are usually designed for single-carrier deployments and aim to provide basic voice transport services, rather than competing on offering multimedia or other advanced capabilities. For brevity, we will not discuss these other voice-over-packet technologies (VoP) further in this chapter.

A related technology, multimedia streaming, shares the point-to-point or multipoint delivery of multimedia information with IP telephony. However, unlike IP telephony, the source is generally a server, not a human being and, more importantly, there is no bidirectional real-time media interaction between the parties. Rather, data flows in one direction, from media server to clients. Like IP telephony, streaming media requires synchronous data delivery where the short-term average delivery rate is equal to the native media rate, but streaming media can often by buffered for significant amounts of time, up to several seconds, without interfering with the service. Streaming and IP telephony share a number of protocols and codecs that will be discussed in this chapter, such as RTP and G.711. Media streaming can be used to deliver the equivalent of voice mail services. However, is beyond the scope of this chapter.

In the discussion below, we will occasionally use the term "legacy telephony" to distinguish plain old telephone service (POTS) provided by today's time-division multiplexing (TDM) and analog circuits from packet-based delivery of telephone-related services, the Next-Generation Network

0-8493-0052-5/00/$0.00+$.50

(NGN). Apologies are extended to the equipment and networks thus deprec(i)ated. The term public switched telephone network (PSTN) is commonly taken as a synonym for "the phone system", although pedants sometimes prefer the post-monopoly term GSTN (General Switched Telephone Network).

IP telephony is one of the core motivations for deploying quality-of-service into the Internet, since packet voice requires one-way network latencies well below 100 ms and modest packet drop rates of no more than about 10% to yield usable service quality Jiang and Schulzrinne [2003]; Jiang et al. [2003]. Most attempts at improving network-related QoS have focused on the very limited use of packet prioritization in access routers. Since QoS has been widely covered and is not VoIP specific, this chapter will not go into greater detail. Similarly, authentication, authorization and accounting (AAA) are core telephony services, but not specific to VoIP.

0.2 Motivation

The transition from circuit-switched the packet switched telephone services is motivated by cost savings, functionality and integration, with different emphasis on each depending on where the technology is being used.

0.2.1 Efficiency

Traditional telephone switches are not very cost effective as traffic routers; each 64 kb/s circuit in a traditional local office switch costs roughly between $150 and $500, primarily because of the line interface costs. Large-scale PBXs have similar per-port costs. A commodity Ethernet switch, on the other hand, costs only between $5 and $25 per 100 Mb/s port, so that switching packets has become significantly cheaper than switching narrowband circuits, even if one discounts the much larger capacity of the packet switch and only considers per-port costs Weiss and Hwang [1998].

"Free" long-distance phone calls were the traditional motivation for consumer IP telephony, even if they were only free incrementally, given that the modem or DSL connection had already been paid for. In the early 1990s, US long-distance carriers had to pay about 7c/minute to the local exchange carriers, an expense that gatewayed IP telephony systems could bypass. This allowed Internet telephony carriers to offer long-distance calls terminating at PSTN phones at significant savings. This charge has now been reduced to less than 1c/minute, decreasing the incentive McKnight [2000].

In many developing countries, carriers competing with the monopoly incumbent have found IP telephony a way to offer voice service without stringing wires to each phone, using DSL or satellite uplinks. Also, leased lines were often cheaper, on a per-bit basis, than paying international toll charges, opening another opportunity for arbitrage Vinall [1998].

In the long run, the cost differential in features such as caller ID, three-way calling and call waiting may well be more convincing than lower per-minute charges.

For enterprises, the current cost of a traditional circuit-switched PBX and a VoIP system are roughly similar, at about $500 a seat, due to the larger cost of IP phones. However, enterprises with branch offices can re-use their VPN or leased lines for intra-company voice communications and can avoid having to lease small numbers of phone circuits at each branch office*. Enterprises can realize operational savings since moves, adds and changes for IP phones are much simpler, only requiring that the phone be plugged in at its new location.

*It is well-known that a single large trunk for a large user population is more efficient than dividing the user population among smaller trunks, due to the higher statistical multiplexing gain.

As described in Section 0.2.3, having a single wiring plant rather than maintaining separate wiring and patch panels for Ethernet and twisted-pair phone wiring is attractive for new construction.

For certain cases, the higher voice compression and silence suppression found in IP telephony (see Section 0.6) may significantly reduce bandwidth costs. There is no inherent reason that VoIP has better compression, but end system intelligence makes it easier and more affordable to routinely compress all voice calls end-to-end. As noted, silence suppression is not well supported in circuit switched networks outside high-cost point-to-point links. (Indeed, in general, packetization overhead can eat up much of this advantage.)

0.2.2 Functionality

In the long run, increased functionality is likely to be a prime motivator for transitiong to IP telephony, even though current deployment largely limit themselves to replicating traditional PSTN features and functionality. PSTN functionality, beyond mobility, has effectively stagnated since the mid-1980 introduction of "CLASS" features Moulton and Moulton [1996] such as caller ID. Attempts at integrating multimedia, for example, have never succeeded beyond a few corporate teleconferencing centers.

Additional functionality is likely to arise from services tailored to user needs and vertical markets (Section 0.9), created by or close to their users, integration with presence and other Internet services, such as web and email. Since Internet telephony completes the evolution from in-band signaling found in analog telephony to complete separation of signaling and media flows, services can be offered equally well by businesses and specialized non-facility-based companies as they can by Internet service providers or telephone carriers.

Since telephone numbers and other identifiers are not bound to a physical telephone jack, it is fairly easy to set up virtual companies, where employee home phones are temporarily made part of the enterprise call center, for example.

It is much easier to secure VoIP services via signaling and media encryption, although legal constraints may never make this feature legally available.

0.2.3 Integration

Integration has been a leitmotif for packet-based communications from the beginning, with integration occurring at the physical layer (same fiber, different wavelengths), link layer (SONET), and, most recently, at the network layer (everything-over-IP). Besides the obvious savings in transmission facilities and the ability to allocate capacity more flexibly, managing a single network promises to be significantly simpler and reduce operational expenditures.

0.3 Standardization

While proprietary protocols are still commonly found in the applications for consumer VoIP services and indeed dominate today for enterprise IP telephony services (Cisco Call Manager protocol), there is a general tendency towards standardizing most components needed to implement VoIP services.

Note that standardization does not imply that there is only one way to approach a particular problem. Indeed, in IP telephony, there are multiple competing standards in areas such as signaling, while in others different architectural approaches are advocated by different communities. Unlike telephony standards, which exhibited significant technical differences across different countries, IP telephony standards so far diverge mostly for reasons of emphasis on different strengths of particular approaches, such as integration with legacy phone systems vs. new services or maturity vs. flexibility.

A number of organizations write standards and recommendations for telephone service, telecommunications and the Internet. Standards organizations used to be divided into “official” and “industry” standards organizations, where the former were established by international treaty or law, while the latter were voluntary organizations founded by companies or individuals. Examples of such treaty-based organizations include the International Telecommunications Union (ITU, www.itu.int), that in 1993 replaced the former International Telephone and Telegraph Consultative Committee (CCITT). The CCITT’s origins are over 100 years old. National organizations include the American National Standards Institute (www.ansi.org) for the United States and the European Telecommunications Standards Institute (ETSI) for Europe. Since telecommunications is becoming less regional, standards promulgated by these traditionally regional organizations are finding use outside those regions.

In the area of IP telephony, 3GPP, the 3rd Generation Partnership Project, has been driving the standardization for third generation wireless networks using “based on evolved GSM core networks and the radio access technologies that they support”. It consists of a number of organizational partners, including ETSI. A similar organization, 3GPP2, deals with radio access technologies derived from the North American CDMA (ANSI/TIA/EIA-41) system; it inherits most higher-layer technologies, such as those relevant for IP telephony, from 3GPP.

When telecommunications were largely a government monopoly, the ITU was roughly the “parliament of monopoly telecommunications carriers”, with a rough one-country, one-vote rule. Now, membership appears in the ITU to be open to just about any manufacturer or research organization willing to pay its dues. Thus, today there is no substantial practical difference between these different major standardization organizations. Standards are not laws or government regulations and obtain their force if customers require that vendors deliver products based on standards.

The Internet Engineering Task Force is an international volunteer standards development organization (SDO) that specifies standards for the Internet Protocol, its applications, such as SMTP, IMAP and HTTP, and related infrastructure services, such as DNS, DHCP and routing protocols. Many of the current IP telephony protocols described in this chapter were developed within the IETF.

In a rough sense, one can distinguish primary from secondary standardization functions. In the primary function, an organization develops core technology and protocols for new functionality, while the emphasis in secondary standardization is on adapting technology developed elsewhere to new uses or describing it more fully for particular scenarios. As an example, 3GPP has adopted and adapted SIP and RTP, developed within the IETF, for the Internet multimedia subsystem in 3G networks. 3GPP also develops radio access technology, which is then in turn used by other organizations.

In addition, some organizations, such as the IMTC and SIP Forum, provide interoperability testing, deployment scenarios, protocol interworking descriptions and educational services.

0.4 Architecture

IP telephony, unlike other Internet applications, is still dominated by concerns about interworking with older technology, here, the PSTN. Thus, we can define three classes Clark [1997] of IP telephony operation (Fig. 0.1), depending on the number of IP and traditional telephone end systems.

In the first architecture, sometimes called trunk replacement, both caller and callee use circuit-switched telephone services. The caller dials into a gateway, which then connects via either the public Internet or a private IP-based network or some combination to a gateway close to the callee. This model requires no changes in the end systems and dialing behavior and is often used, without the participants being aware of it, to offer cheap international prepaid calling card calls. However, it can also be used to connect two PBXs within a corporation with branch offices. Many PBX vendors now offer IP trunk interfaces that simply replace a T-1 trunk by a packet-switched connection.

Another hybrid architecture, sometimes called hop-on or hop-off depending on the direction, places calls from a PSTN phone to an IP-based phone or vice versa. In both cases, the phone is addressed by a regular telephone number, although the phone may not necessarily be located in the geographic area typically associated with that area code. A number of companies have started to offer IP phones for residential and small-business subscribers that follow this pattern. A closely related architecture is called an *IP PBX*, where phones within the enterprise connect to a gateway that provides PSTN dial tone.

If the IP PBX is shared among several organizations and operated by a service provider, it is referred to as *IP Centrex* or *hosted IP PBX*, as the economic model is somewhat similar to the centrex service offered by traditional local exchange carriers. Like classical centrex, IP centrex service decreases the initial capital investment for the enterprise and makes system maintenance the responsibility of the service provider. Unlike PSTN centrex, where each phone has its own access circuit, IP centrex only needs a fraction of the corporate Internet connectivity to the provider and is generally more cost-efficient. If the enterprise uses standards-compliant IP phones, it is relatively straightforward to migrate between IP centrex and IP PBX architectures, without changing the wiring plant or the end systems.

This architecture is also found in some cable systems, where phone service is provided by the cable TV operator (known as an MSO) Miller et al. [2001]; Wocjik [2000]. Note, however, that not all current cable-TV-phone arrangements use packet voice; some early experiments simply provide a circuit switched channel over coax and fiber.

The third architecture dispenses with gateways and uses direct IP-based communications end-to-end between caller and callee. This arrangement dominated early PC-based IP telephony, but only works well if all participants are permanently connected to the Internet.

The most likely medium-term architecture is a combination of the hybrid and end-to-end model, where calls to other IP phones travel direct, while others use gateways and the PSTN. If third-generation mobile networks succeed, the number of IP-reachable devices may quickly exceed those using the traditional legacy interface. If devices are identified by telephone numbers, there needs to be a way for the caller to determine if a telephone number is reachable directly. The ENUM directory mechanism described in Section 0.7.4 offers one such mapping.

0.5 Overview of Components

At the lower protocol layers, Internet components are easily divided into a small number of devices and functions that rarely cause confusion. For example, hosts, routers and DNS servers have clearly defined functionality and are usually in separate hardware, or at least servers. Usually, these servers are distinguished by the protocols they speak: a web server primarily deals with HTTP, for example. Things are not nearly as simple for IP telephony, where an evolving understanding, the interaction with the legacy telephony world and marketing have created an abundance of names that sometimes reflect function and sometimes common bundlings into a single piece of hardware.

In particular, the term “softswitch” is often used to describe a set of functions that roughly replicate the control functionality of a traditional telephone switch. However, this term is sufficiently vague that it should be avoided in technical discussions.

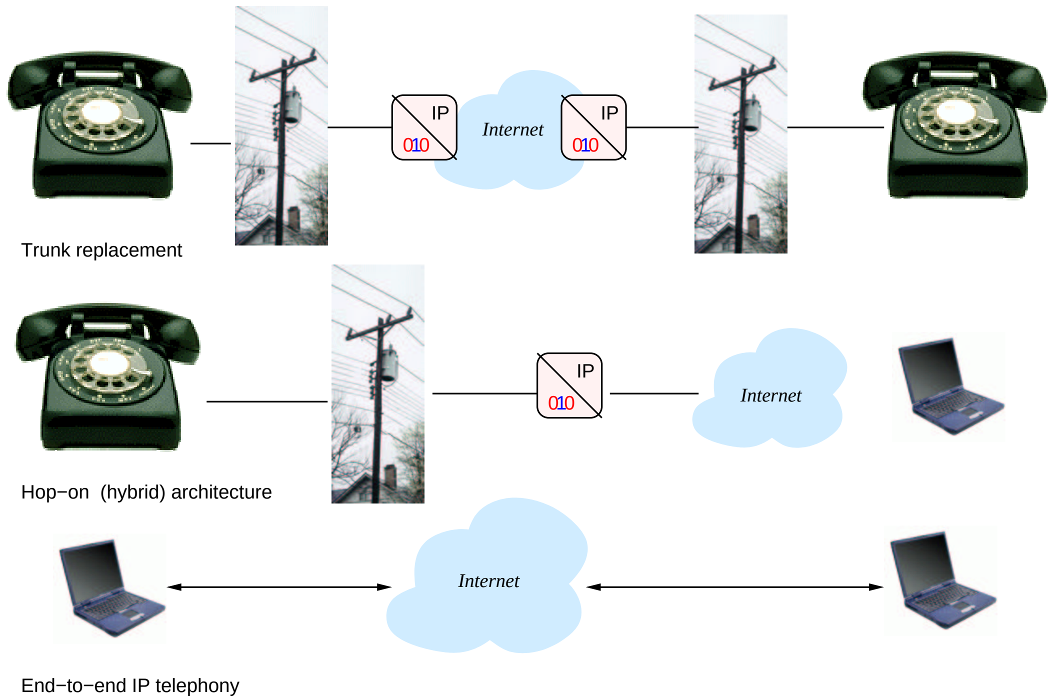

FIGURE 0.1
Internet telephony architectures

The International Packet Communications Consortium International Packet Communications Consortium has attempted to define these functional entities and common physical embodiments.

0.5.1 Common Hardware and Software Components

The most common hardware component in IP telephony are IP phones, access gateways and integrated access devices (IADs).

IP phones are end systems and endpoints for both call setup (signaling) and media, usually audio. There are both hardware phones that operate stand-alone, and softphones, software applications that run on common operating system platforms on personal computers. Hardware phones typically consist of a digital signal processor with analog-to-digital (A/D) and digital-to-analog (D/A) conversion, general-purpose CPU and network interface. The CPU often runs an embedded operating system and usually supports standard network protocols such as DNS for name resolution, DHCP for network autoconfiguration, NTP for time synchronization, tftp and HTTP for application configuration. Modern IP phones offer the same range of functionality as analog and digital business telephones, including speakerphones, caller ID displays and programmable keys. Some IP phones have limited display programmability or have a built-in Java environment for service creation.

Access gateways connect the packet and circuit-switched world, both in the control and media planes. They packetize bit streams or analog signals coming from the PSTN into IP packets and deliver them to their IP destination. In the opposite direction, they convert sequences of IP packets containing segments of audio into a stream of voice bits and “dial” the appropriate number in the legacy phone system. Small (residential or branch-office) gateways may support only one or two analog lines, while carrier-class gateways may have a capacity of a T1 (24 phone circuits) or even a T3 (720 circuits). Large-scale gateways may be divided into a media component that encodes and

FIGURE 0.2
Some examples of IP phones

decodes voice and a control component, often a general-purpose computer, that handles signaling.

An integrated access device (IAD) is a small gateway that typically has an Ethernet interface on one side and an ISDN or analog phone interface on the other. They allow commercial and residential users to re-use their large existing investment in analog and digital phones, answering machines and fax machines on an IP-based phone network. Sometimes the IAD is combined in the same enclosure with a DSL or cable modem and then, to ensure confusion, labeled a residential gateway (RG).

In addition to these specialized hardware components, there are a number of software functions that can be combined into servers. In some cases, all such functions reside in one server component (or a tightly coupled group of server processes), while in other cases they can be servers each running on its own hardware platform. The principal components are:

Signaling conversion: Signaling conversion servers transform and translate call setup requests. They may translate names and addresses, or translate between different signaling protocols. Later on, we will encounter them as gatekeepers in H.323 networks, proxy servers in SIP networks, and protocol translators in hybrid networks Liu and Mouchtaris [2000]; Singh and Schulzrinne [2000].

Application server: An application server implements service logic for various common or custom features, typically through an API such as JAIN, SIP servlets, CPL or proprietary versions, as discussed in Section 0.9. Often, they provide components of the operational support system (OSS), such as accounting, billing or provisioning. Examples include voice mail servers, conference servers, and calling card services.

Media server: A media server manipulates media streams, e.g., by recording, playback, codec translation or text-to-speech conversion. It may be treated like an end system, i.e., it terminates both media and signaling sessions.

0.6 Media Encoding

0.6.1 Audio

In both legacy and packet telephony, the most common way of representing voice signals is as a logarithmically companded† byte stream, with a rate of 8,000 samples of 8 bits each per second. This telephone-quality audio codec is known as G.711 International Telecommunication Union [1998b], with two regional variations known as μ-law or A-law audio, can reproduce the typical telephone frequency range of about 300 to 3,400 Hz. Typically, 20 to 50 ms worth of audio samples are transmitted in one audio packet. G.711 is the only sample-based codec in wide use.

As noted earlier, one of the benefits of IP telephony is the ability to compress telephone-quality voice below the customary rate of 64 kb/s found in TDM networks. All of commonly used codecs operate at a sampling rate of 8,000 Hz and encode audio into frames of between 10 and 30 ms duration. Each audio frame consists of a speech parameters, rather than audio samples. Only a few audio codecs are commonly used in IP telephony, in particular G.723.1 International Telecommunication Union [1996c] operating at 5.3 or 6.3 kb/s and modest speech quality, G.729 International Telecommunication Union [1996a] at 8 kb/s, and the GSM full-rate (FR) codec at 13 kb/s.

More recently, two royalty-free codecs have been published: iLBC Andersen et al. [2003] operating at 13.33 or 15.2 kb/s, with a speech quality equivalent to G.729, but higher loss tolerance, and Speex Herlein et al. [2003], operating at a variable bit rate ranging between 2.15 and 24.6 kb/s.

All codecs can operate in conjunction with silence suppression, also known as *voice activity detection* (VAD). VAD measures speech volume to detect when a speaker is pausing between sentences or letting the other party talk. Most modern codecs incorporate silence detection, while it is a separate speech processing function in codecs like G.711. Silence suppression can reduce the bit rate by 50-60%, depending on whether short silences between words and sentences are removed or not Jiang and Schulzrinne [2000a]. The savings can be much larger in multiparty conferences; there, silence suppression is required also to avoid that the summed background noise of the listeners does not interfere with audio perception.

During pauses, no packets are transmitted, but well-designed receivers will play *comfort noise* Gierlich and Kettler [2001] that avoids the impression to the listener that the line is “dead”. The sender occasionally updates Zopf [2002] the loudness and spectral characteristics, so that there is no unnatural transition when the speaker breaks her silence.

Silence suppression not only reduces the average bit rate, but also simplifies playout delay adaptation, which is used by the receiver to compensate for the variable queueing delays incurred in the network.

DTMF (“touchtone”) and other voiceband data signals such as fax tones pose special challenges to high-compression codecs and may not be rendered sufficiently well to be recognizable by the receiver. Also, it is rather wasteful to have an IP phone generate a waveform for DTMF signals, just to have the gateway spend DSP cycles recognizing it as a digit. Thus, many modern IP phones generate tones as a special encoding Schulzrinne and Petrack [2000].

While the bit rate and speech quality are generally the most important figures of merit for speech codecs, codec complexity, resilience to packet loss and algorithmic delay are other important considerations. The algorithmic delay is the delay imposed by the compression operation, as the compression operation needs to have access to a certain amount of audio data (block size) and may need to “look ahead” to estimate parameters.

Music codecs such as MPEG 2 Layer 3, commonly known as MP3, or MPEG-2 AAC can also compress voice, but since they are optimized for general audio signals rather than speech, they

† Smaller audio loudness values receive relatively more bits of resolution than larger ones.

typically produce much lower audio quality for the same bit rate. The typical MP3 encoding rates, for example, range from 32 kb/s for "better than AM radio" quality to 96 and 128 kb/s for "near CD quality". (Conversely, many low bit-rate speech codecs sound poor with music since their acoustic model is tuned towards producing speech sounds, not music.)

Generally, the algorithmic delay of these codecs is too long for interactive conversations, for example about 260 ms for AAC at 32 kb/s. However, the new AAC MPEG-4 low delay codec reduces algorithmic delays to 20 ms.

In the future, it is likely that "better-than-phone-quality" codecs will become more prevalent, as more calls are placed between IP telephones rather than from or into the PSTN. So-called conference-quality or wideband codecs typically have an analog frequency range of 7 kHz and a sampling rate of 16 kHz, with a quality somewhat better than static-free AM radio. Examples of such codecs include G.722.1 International Telecommunication Union [1999a]; Luthi [2001] at 24 or 32 kb/s, Speex Herlein et al. [2003] at 4 to 44.2 kb/s, AMR WB Sjoberg et al. [2002]; International Telecommunication Union [2002]; 3GPP [2001a,b] at 6.6-23.85 kb/s.

The quality of audio encoding with packet loss can be improved by using forward error correction (FEC) and packet loss concealment (PLC) Jiang et al. [2003]; Jiang and Schulzrinne [2002b]; Rosenberg and Schulzrinne [1999]; Jiang and Schulzrinne [2002c,a, 2000b]; Schuster et al. [1999]; Bolot et al. [1995]; Toutireddy and Padhye [1995]; Carle and Biersack [1997]; Stock and Adanez [1996]; Boutremans and Boudec [2001]; Jeffay et al. [1994].

0.6.2 Video

For video streams, the most commonly used codecs are H.261 International Telecommunication Union [1993b], which is being replaced by more modern codecs such as H.263 International Telecommunication Union [1998c] and H.263+. Like MPEG-1 and MPEG-2, H.261 and H.263 make use of interframe correlation and motion prediction to reduce the video bit rate. Sometimes, motion JPEG is used for high-quality video, which consists simply of sending a sequence of JPEG images. Compared to motion-compensated codecs, its quality is lower, but it also requires much less encoding effort and is more tolerant of packet loss.

0.7 Core Protocols

Internet telephony relies on five types of application-specific protocols to offer services: media transport (Section 0.7.1), device control (Section 0.7.2), call setup and signaling (Section 0.7.3), address mapping (Section 0.7.4) and call routing (Section 0.7.5). These protocols are not found in all Internet telephony implementations.

0.7.1 Media Transport

As described in Section 0.6, audio is transmitted in frames representing between 10 and 50 ms of speech content. Video, similarly, is divided into frames, at a rate of between 5 and 30 frames a second. However, these frames cannot simply be placed into UDP or TCP packets, as the receiver would not be able to tell what kind of encoding is being used, what time period the frame represents and whether a packet is the beginning of a talkspurt.

The Real-Time Transport Protocol (RTP Schulzrinne et al. [1996]) offers this common functionality. It adds a 12-byte header between the UDP packet header and the media content [‡] The packet header labels the media encoding so that a single stream can alternate between different codecs Schulzrinne [1996], e.g., for DTMF Schulzrinne and Petrack [2000] or different network conditions. It has a timestamp increasing at the sampling rate that makes it easy for the receiver to correctly place packets in a playout buffer, even if some packets are lost or packets are skipped due to silence suppression. A sequence number provides an indication of packet loss. A secure profile of RTP Baugher et al. [2003] can provide confidentiality, message authentication, and replay protection. Finally, a synchronization source identifier (SSRC) provides a unique 32-bit identifier for multiple streams that share the same network identity.

Just like IP has a companion control protocol, ICMP Postel [1981], RTP uses RTCP for control and diagnostics. RTCP is sent on an adjacent UDP port number to the main RTP stream and is paced to consume no more than a set fraction of the main media stream, typically 5%. RTCP has three main functions: (1) it identifies the source by a globally unique user@host-style identifier and adds labels such as the speaker's name; (2) it reports on sender characteristics such as the number of bytes and packets transmitted in an interval; (3) receivers report on the quality of the stream received, indicating packet loss and jitter. More extensive audio-specific metrics have been proposed recently Friedman et al. [2003].

While RTP streams are usually exchanged unmodified between end system, it is occasionally useful to introduce processing elements into these streams. RTP *mixers* take several RTP streams and combine them, e.g., by summing their audio content in a conference bridge. RTP *translators* take individual packets and manipulate the content, e.g., by converting one codec to another. For mixers, the RTP packet header is augmented by a list of contributing sources that identify the speakers that were mixed into the packet.

0.7.2 Device Control

In Section 0.4, we noted that some large-scale gateways are divided into two parts, a media-processing part that translates between circuit-switched and packet-switched audio and a media gateway controller (MGC) or call agent (CA) that directs its actions. The MGC is typically a general-purpose computer and terminates and originates signaling, such as SIP (see Section 0.7.3.2), but does not process media.

In an enterprise PBX or cable modem context (there, called network-based call signaling CableLabs [2003]), some have proposed that a central control agent provides low-level instructions to user end systems, such as IADs and IP phones, and receives back events such as numbers dialed or on/off hook status.

There are currently two major protocols that allow such device control, namely the older MGCP Arango et al. [1999] and the successor Megaco/H.248 Groves et al. [2003]. Currently, MGCP is probably the more widely used protocol. MGCP is text-based, while Megaco has a text and binary format, with the latter apparently rarely implemented. Fig. 0.3 gives a flavor of the protocol operation, drawn from CableLabs [2003]. First, the CA sends a NotificationRequest (RQNT) to the client, i.e., the user's phone. The 'N' parameter identifies the call agent, the 'X' parameter identifies the request, 'R' parameter enumerates the events, where 'hd' stands for off-hook. The 200 response by the client indicates that request was received. When the user picks up the phone, a Notify (NTFY) message is sent to the CA, including the 'O' parameter that describes the event that was observed. The CA then instructs the devices with a combined CreateConnection (CRCX) and NotificationRequest command to create a connection, labeled with a call ID 'C', provide dial tone ('dl' in the 'S' parameter) and collect digits according to digit map 'D'. The digitmap spells out the

[‡]TCP is rarely used since its retransmission-based loss recovery mechanism may not recover packets in the 100 ms or so required and congestion control may introduce long pauses into the media stream.

combinations of digits and time-outs (T) that indicate that the complete number has been dialed. The client responds with a ’200’ message indicating receipt of the CRCX request and includes a session description so that the CA knows where it should direct dialtone to. The session description uses the Session Description Protocol (SDP) Handley and Jacobson [1998]; we omitted some of the details for brevity. The ’c’ line indicates the network address, the ’m’ line the media type, port, the RTP profile (here, the standard audio/video profile) and the RTP payload identifier (0, which stands for G.711 audio). To allow later modifications, the connection gets its own label (I). The remainder of the call setup proceeds apace, with a notification when the digits have been collected. The CA then tells the calling client to stop collecting digits. It also creates a connection on the callee side and instructs that client to ring. Additional messages are exchanged when the callee picks up and when either side hangs up. For this typical scenario, the caller generates and receives a total of 20 messages, while the callee side sees an additional 15 messages.

As the example illustrated, MGCP and Megaco instruct the device in detailed operations and behavior and the device simply follows these instructions. The device exports low-level events such as hook switch actions and digits pressed, rather than, say, calls. This makes it easy to deploy new services without upgrades on the client side, but also keeps all service intelligence “in the network”, i.e., the CA. Since there is a central CA, device control systems are limited to single administrative domains. Between domains, CAs use a peer-to-peer signaling protocol, such as SIP or H.323, described below, to set up the call.

```
RQNT 1201 aaln/1@ec-1.whatever.net MGCP 1.0 NCS 1.0
N: ca@ca1.whatever.net:5678
X: 0123456789AB
R: hd
-------------------------------
200 1201 OK
-------------------------------
NTFY 2001 aaln/1@ec-1.whatever.net MGCP 1.0 NCS 1.0
N: ca@ca1.whatever.net:5678
X: 0123456789AB
O: hd

-------------------------------
CRCX 1202 aaln/1@ec-1.whatever.net MGCP 1.0 NCS 1.0
C: A3C47F21456789F0
L: p:10, a:PCMU
M: recvonly
N: ca@ca1.whatever.net:5678
X: 0123456789AC
R: hu, [0-9#*T](D)
D: (0T | 00T | [2-9]xxxxxx | 1[2-9]xxxxxxxxx | 011xx.T)
S: dl

--------------------------------
200 1202 OK
I: FDE234C8

c=IN P4 128.96.41.1
m=audio 3456 RTP/AVP 0
```

FIGURE 0.3
Sample call flow CableLabs [2003]

0.7.3 Call Setup and Control: Signaling

One of the core functions of Internet telephony that distinguishes it from, say, streaming media is the notion of call setup. Call setup allows a caller to notify the callee of a pending call, to negotiate call parameters such as media types and codecs that both sides can understand, to modify these parameter in mid-call and to terminate the call.

In addition, an important function of call signaling is "rendezvous", the ability to locate an end system by something other than just an IP address. Particularly with dynamically assigned network addresses, it would be rather inconvenient if callers had to know and provide the IP address or host name of the destination. Thus, the two most prevalent call signaling protocols both offer a binding (or registration) mechanism where clients register their current network address with a server for a domain. The caller then contacts the server and obtains the current whereabouts of the client.

The protocols providing these functions are referred to as signaling protocols; sometimes, they are also further described as peer-to-peer signaling protocols since both sides in the signaling transactions have equivalent functionality. This distinguishes them from the device control protocols like MGCP and Megaco, where the client reacts to commands and supplies event notifications.

Two signaling protocols are in common commercial use at this time, namely H.323 (Section 0.7.3.1 and SIP (Section 0.7.3.2). Their philosophies differ, although the evolution of H.323 has brought it closer to SIP.

0.7.3.1 H.323

The first widely used standardized signaling protocol was provided by the ITU in 1996, as the H.323 family of protocols. H.323 has its origins in extending ISDN multimedia conferencing, in Recommendation H.320 International Telecommunication Union [1999b], to LANs and inherits aspects of ISDN circuit-switched signaling. Also, H.323 has evolved considerably, through four versions, since its original design. This makes it somewhat difficult to describe its operation definitively in a modest amount of space. In addition, many common implementations, such as Microsoft NetMeeting, only support earlier versions, typically version 2, of the protocol. (Later versions are supposed to support all earlier versions and fall back to the less-functional version if necessary.)

H.323 is actually a whole suite of protocol specifications. The basic architecture is described in H.323 International Telecommunication Union [2000], registration and call setup signaling ("ringing the phone") is described in H.225.0 International Telecommunication Union [1996d], media negotiation and session setup in H.245 International Telecommunication Union [1998a], and the ISDN signaling underlying the protocol in Q.931 International Telecommunication Union [1993a]. The two sub-protocols for call and media setup, Q.931 and H.245, use different encodings. Q.931 is a simple binary protocol with mostly fixed-length fields, while H.245 is encoded as ASN.1. To make matters more confusing, H.245 and other parts of H.323 use an ASN.1 encoding that is rarely found elsewhere, the so-called packed encoding rules (PER) International Telecommunication Union [1997a]. Generally, H.323 applications developers rely on libraries or ASN.1 code generators.

The protocols listed so far are sufficient for basic call functionality and are those most commonly implemented in endpoints. Classical telephony services such as call forwarding, call completion or caller identification are described in the H.450.x series of recommendations. (They use yet another way to encode requests.) Security mechanisms are discussed in H.235. Functionality for application sharing and shared whiteboards, with its own call setup mechanism, is described in the T.120 series of recommendations International Telecommunication Union [1996b].

H.323 uses similar component labels as we have seen earlier, namely terminals (that is, end systems) and gateways. It also introduces gatekeepers, which route signaling messages between domains and register users, provide authorization and authentication of terminals and gateways, manage bandwidth, and provide accounting, billing and charging functions. Finally, from its origin in multimedia conferencing, H.323 describes multipoint control units (MCUs), the packet equivalent to a conference bridge.

Each gatekeeper is responsible for one *zone*, which can consist of any number of terminals, gateways and MCUs.

Fig. 0.4 shows a typical call setup between two terminals within the same zone. (Inter-gatekeeper communications is specified in H.323v3).

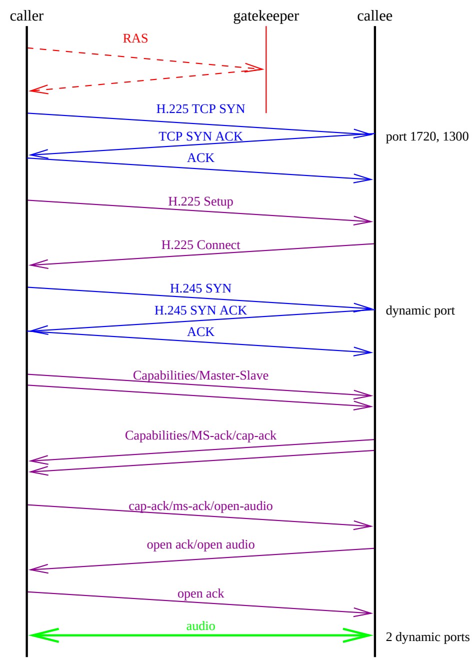

FIGURE 0.4
Example H.323 call flow

0.7.3.2 Session Initiation Protocol (SIP)

The Session Initiation Protocol (SIP) is a protocol framework originally designed for establishing, modifying and terminating multimedia sessions such as VoIP calls. Beyond the session setup functionality, it also provides event notification for telephony services such as supervised call transfer and message waiting indication and more “modern” services such as presence.

SIP does not describe the audio and media components of a session; instead, it relies on a separate session description carried in the body of INVITE and ACK messages. Currently, only the Session Description Protocol (SDP) Handley and Jacobson [1998] is being used, but an XML-based replacement Kutscher et al. [2003] is being discussed. The example in Fig. 0.5 Johnston [2003] shows a simple audio session originated by user “alice” to be received by IP address 192.0.2.101 and port 49172 using RTP and payload type 0 (μ-law audio).

Besides carrying session descriptions, the core function of SIP is to locate the called party, mapping a user name such as sip:alice@atlanta.example.com to the network addresses used by devices owned by Alice. Users can re-use their email address as a SIP URI or choose a different one. As for email addresses, users can have any number of SIP URIs with different providers that all

```
v=0
o=alice 2890844526 2890844526 IN IP4 client.atlanta.example.com
s=-
c=IN IP4 192.0.2.101
t=0 0
m=audio 49172 RTP/AVP 0
a=rtpmap:0 PCMU/8000
```

FIGURE 0.5
Example session description

reach the same device.

User devices such as IP phones and conferencing software run SIP *user agents*; unlike for most protocols, such user agents usually can act as both clients and servers, i.e., they both originate and terminate SIP requests.

Instead of SIP URIs, users can be identified also by telephone numbers, expressed as “tel” URIs Schulzrinne and Vaha-Sipila [2003] such as `tel:+1-212-555-1234`. Calls with these numbers are then either routed to an Internet telephony gateway or translated back into SIP URIs via the ENUM mechanism described in Section 0.7.4.

A user provides a fixed contact point, a so-called SIP proxy, that maps incoming requests to network devices registered by the user. The caller does not need to know the current IP addresses of these devices. This decoupling between the globally unique user-level identifier and device network addresses supports *personal mobility*, the ability of a single user to use multiple devices, and deals with the practical issue that many devices acquire their IP address temporarily via DHCP. The proxy typically also performs call routing functions, for example, directing unanswered calls to voice mail or an auto-attendant. The SIP proxy plays a role somewhat similar to an SMTP Mail Transfer Agent (MTA) rfc [2001], but naturally does not store messages. Proxies are not required for SIP; user agents can contact each other directly.

A request can traverse any number of proxies, but typically at least two, namely one “outbound proxy” in the caller’s domain and the proxy in the callee’s domain. For reliability and load balancing, a domain can use any number of proxies. A client identifies a proxy by looking up the DNS SRV Gulbrandsen et al. [2000] record enumerating primary and fall-back proxies for the domain in the SIP URI.

Session setup messages and media generally traverse independent paths, that is, they only join at the originating and terminating client. Media then flows directly on the shortest network path between the two terminals. In particular, SIP proxies do not process media packets. This makes it possible to route call setup requests through any number of proxies without worrying about audio latency or network efficiency. This “path-decoupled” signaling completes the evolution of telephony signaling from in-band audio signaling to out-of-band, disassociated channel signaling introduced by Signaling System No. 7. Since telephony signaling needs to configure switch paths, it generally meets up with the media stream in telephone switches; there is no such need in IP telephony.

Just like a single phone line can ring multiple phones within the same household, a single SIP address can contact any number of SIP devices with one call, albeit potentially distributed across the network. This capability is called “forking” and is performed by proxies. These forking proxies gather responses from the entities registered under the SIP URI and return the best response, typically the first one to pick up. This feature makes it easy to develop distributed voicemail services and simple automatic call distribution (ACD) systems.

Fig. 0.6 shows a simple SIP message and its components. SIP is a textual protocol, similar to SMTP rfc [2001] and HTTP Fielding et al. [1999]. A SIP request consists of a request line containing the request method and the SIP URI identifying the destination, followed by a number of header fields that help proxies and user agents to route and identify the message content.

There are a large number of SIP request methods, summarized in Table 0.1.

ACK	Rosenberg et al. [2002b]	acknowledges final INVITE response
BYE	Rosenberg et al. [2002b]	terminates session
CANCEL	Rosenberg et al. [2002b]	cancels INVITE
INFO	Donovan [2000]	mid-session information transfer
INVITE	Rosenberg et al. [2002b]	establishes session
NOTIFY	Roach [2002]	event notification
OPTIONS	Rosenberg et al. [2002b]	determine capabilities
PRACK	Rosenberg and Schulzrinne [2002]	acknowledge provisional response
REGISTER	Rosenberg et al. [2002b]	register name-address mapping
SUBSCRIBE	Roach [2002]	subscribe to event
UPDATE	Rosenberg [2002]	update session description
MESSAGE	rfc [2002]	user-to-user messaging
REFER	Sparks [2003]	transfer call

SIP request methods

SIP messages can be requests or responses, which only differ syntactically in their first lines. Almost all SIP requests generate a final response indicating whether the request succeeded or why it failed, with some requests producing a number of responses that update the requestor on the progress of the request via provisional responses.

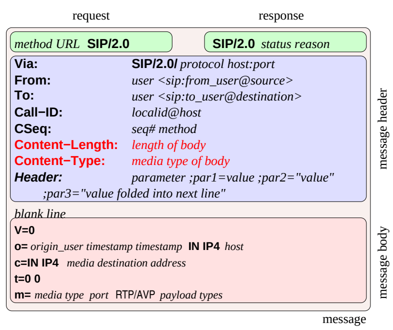

FIGURE 0.6
Example SIP INVITE message

Unlike other application-layer protocols, SIP is designed to run over both reliable and unreliable transport protocols. Currently, UDP is the most common transport mechanism, but TCP and SCTP, as well as secure transport using TLS Dierks and Allen [1999] are also supported. To achieve reliability, a request is retransmitted until it is acknowledged by a provisional or final response. The INVITE transaction, used to set up sessions, behaves a bit differently since considerable time may elapse between the call arrival and the time that the called party picks up the phone. An INVITE transaction is shown in Fig. 0.7.

Once a request has reached the right destination, the two parties negotiate the media streams using an offer-answer model, where the caller typically offers a capability and the callee makes a counter-proposal. Sessions can be changed in the middle of a session, e.g., to add or remove a media

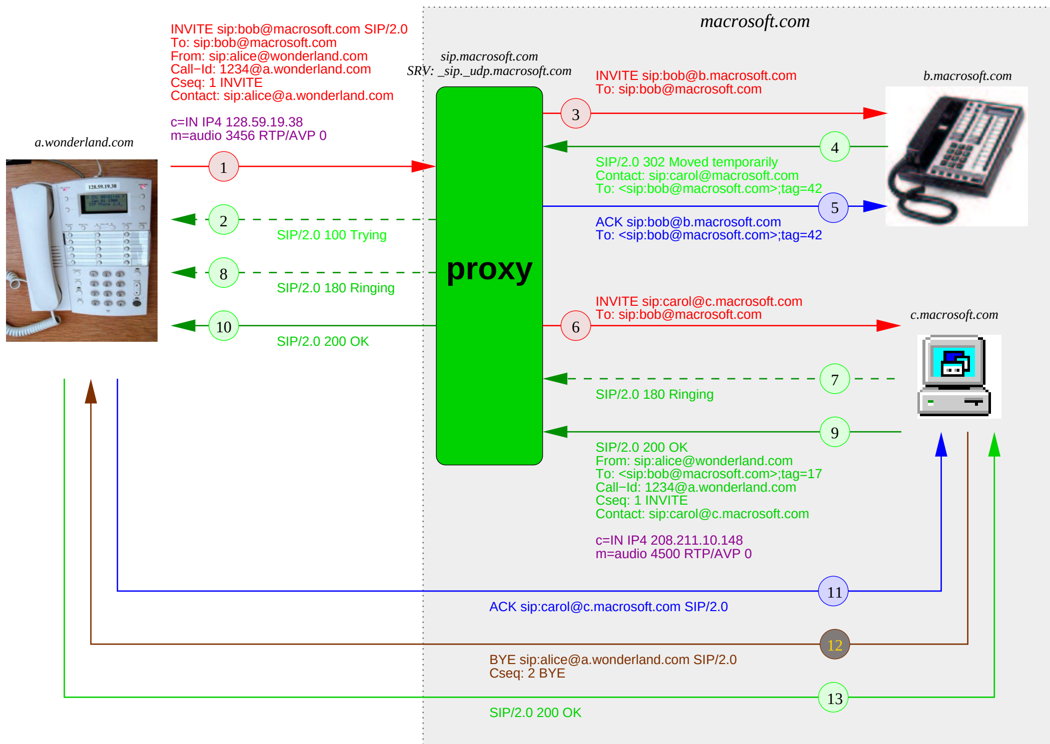

FIGURE 0.7
Example SIP call flow

stream.

SIP can be extended by adding new methods, message body types or header fields. Generally, receivers and proxies are free to ignore header fields that they do not understand, but a requestor can require that the receiver understand a particular feature by including a Require header field. If the receiver does not implement that feature, it must reject the request.

SIP user agents can initiate sessions between two other entities, acting as third-party call controllers or back-to-back user agents (B2BUAs) Rosenberg et al. [2003b].

While the basic protocol mechanisms are stable, components of the SIP infrastructure are currently still under active development within the IETF and, for third-generation mobile networks, in 3GPP. Such features include support for legacy telephone features such as overlap dialing as well as advanced call routing features such as caller preferences Rosenberg et al. [2003a]; Rosenberg and Kyzivat [2003].

0.7.4 Telephone Number Mapping

In the long run, VoIP destinations may well be identified by textual SIP URIs, but familiarity, deployed infrastructure and end system user interface limitations dictate the support of telephone numbers International Telecommunication Union [1997b] for the foreseeable future. To facilitate the transition to an all-IP infrastructure, it is helpful if telephone numbers can be mapped to SIP and other URIs. This avoids, for example, that a VoIP terminal needs to go through a gateway to reach a terminal identified by a telephone number, even though that terminal also has VoIP capability.

The ENUM service Faltstrom [2000]; Faltstrom and Mealling [2003] offers a standardized mapping service from global telephone numbers to one or more URIs. It uses the Dynamic Delegation Discovery System (DDDS) system Mealling [2002] and a relatively new DNS record type, NAPTR. NAPTR records allow for mapping of the name via a regular expression, as shown in Fig. 0.8 for the telephone number +46-89761234. Since the most significant digit for telephone numbers is on the left, while the most significant component of DNS names is on the right, the telephone number is reversed and converted into the DNS name “4.3.2.1.6.7.9.8.6.4.e164.arpa” in this example.

```
$ORIGIN 4.3.2.1.6.7.9.8.6.4.e164.arpa.
  IN NAPTR 10 100 "u" "E2U+sip"  "!^.*$!sip:info@example.com!"     .
  IN NAPTR 10 101 "u" "E2U+h323" "!^.*$!h323:info@example.com!"    .
  IN NAPTR 10 102 "u" "E2U+msg:mailto" "!^.*$!mailto:info@example.com!"  .
```

FIGURE 0.8
ENUM example Faltstrom and Mealling [2003]

0.7.5 Call Routing

Any IP telephony gateway can reach just about any telephone number and any VoIP device can reach any gateway. Since saving on international transit is a prime motivation for deploying IP telephony, gateways are likely to be installed all over the world, with gateways in each country handling calls for that country or maybe a region. Such gateways may be operated by one large corporation or a set of independent operators that exchange billing information via a clearinghouse Hoffman and Yergeau [2000].

Each operator divides their gateways into one or more Internet Telephony administrative domains (ITADs), represented by a Location Server (LS). The location servers learn about the status of gateways in their domain through a local protocol, such as TGREP Bangalore et al. [2003] or SLP

Zhao and Schulzrinne [2002]. Through the Telephony Routing over IP protocol (TRIP) Rosenberg et al. [2002a], location servers peer with each other and exchange information about other ITADs and their gateways.

0.8 Brief History

The first attempt to treat speech as segments rather than a stream of samples was probably Time-Assigned Speech Interpolation (TASI). TASI uses silence gaps to multiplex more audio streams than the nominal circuit capacity of a TDM system, by re-assigning time slots to active speech channels. It has been used in transoceanic cables since the 1960s Easton et al. [1982]; Fraser et al. [1962]; Miedema and Schachtman [1962]; Weinstein and Hofstetter [1979]; Campanella [1978]; Rieser et al. [1981]. While TASI is not packet switching, many of the analysis techniques to estimate the statistical multiplexing gains apply to packet voice as well.

Attempts to transmit voice across IP-based packet networks date back to the earliest days of ARPAnet, with the first publication in 1973, only two years after the first email. Magill [1973]; Cohen [1976a,b, 1977b, 1978]; Anonymous [1983]. In August 1974, real-time packet voice was demonstrated between USC/ISI and MIT Lincoln Laboratories, using CVSD (Continuous Variable Slope Delta Modulation) and NVP Cohen [1977a]. In 1976, live packet voice conferencing was demonstrated between USC/ISI, MIT Lincon Laboratories, Chicago, and SRI, using linear predictive audio coding (LPC) and the Network Voice Control Protocol (NVCP). These initial experiments, run on 56 kb/s links, demonstrated the feasibility of voice transmission, but required dedicated signal processing hardware and thus did not lend themselves to large-scale deployments. Development appears to have been largely dormant since those early experiments.

In 1989, the Sun SPARCstation 1 introduced a small form-factor Unix workstation with a low-latency audio interface. This also happened to be the workstation of choice for DARTnet, an experimental T-1 packet network funded by DARPA (Defense Advanced Research Projects Agency). In the early 1990s, a number of audio tools such as vt, vat Jacobson [1994]; Jacobson and McCanne [1992] and nevot Schulzrinne [1992], were developed that explored many of the core issues of packet transmission, such as playout delay compensation Montgomery [1983]; Ramjee et al. [1994]; Rosenberg et al. [2000]; Moon et al. [1998], packet encapsulation, QoS and audio interfaces, were explored. However, outside of the multicast backbone overlay network (Mbone) Eriksson [1993]; Chuang et al. [1993] that reached primarily research institutions and was used for transmitting IETF meetings Casner and Deering [1992] and NASA space launches, the general public was largely unaware of these tools. More popular was Cu-SeeMe, developed in 1992/1993 Cogger [1992].

The ITU standardized the first audio protocol for general packet networks in 1990 International Telecommunication Union [1990], but this was used only for niche applications, as there was no signaling protocol to set up calls.

In about 1996, VocalTec Communications Ltd. commercialized the first PC-based packet voice applications, primarily used initially to place free long distance calls between PCs. Since then, standardization of signaling protocols like RTP and H.323 in 1996 Thom [1996], have started the transition from experimental research to production services.

0.9 Service Creation

Beyond basic call setup and teardown, the legacy telephone has developed a number of services or *features*, including such common ones as call forwarding on busy or three-way calling and more specialized ones such as distributed call center functionalities. Almost all such services were designed to be developed on PSTN or PBX switches and deployed as a general service, with modest user parameterization.

Both SIP and H.323 can support most Signaling System #7 features Lennox et al. [1999] through protocol machinery, although the philosphy and functionality differs between protocols Glasmann et al. [2001]. Unlike legacy telephones, both end systems and network servers can provide services Wu and Schulzrinne [2003, 2000], often in combination. End system services scale better and can provide a more customized user interface, but may be less reliable and harder to upgrade.

However, basic services are only a small part of the service universe. One of the promises of IP telephony is the ability for users or programmers working closely with small user groups to create new services or customize existing ones. Similar to how dynamic, data-driven web pages are created, a number of approaches have emerged for creating IP telephony services. Java APIs such as JAIN and SIP servlets are meant for programmers and expose almost all signaling functionality to the service creator. They are, however, ill-suited for casual service creation and require significant programming expertise.

Just like common gateway interface (cgi) services on web servers, SIP-cgi Lennox et al. [2001] allows programmers to create user-oriented scripts in languages such as Perl and Python. A higher-level representation of call routing services is exposed through the Call Processing Language (CPL) CPL Lennox and Schulzrinne [2000a]; Lennox et al. [2003].

With distributed features, the problem of feature interaction Cameron et al. [1994] arises. IP telephony removes some of the common causes of feature interaction such as ambiguity in user input, but adds others Lennox and Schulzrinne [2000b] that are just beginning to be explored.

0.10 Glossary

The following glossary lists common abbreviations found in IP telephony. It is partially extracted from International Packet Communications Consortium.

3G	Third Generation (wireless)
3GPP	3G Partnership Project (UMTS)
3GPP2	3G Partnership Project 2 (UMTS)
AAA	Authentication, Authorization and Accounting (IETF)
AG	Access Gateway
AIN	Advanced Intelligent Network
AS	Application Server
BICC	Bearer Independent Call Control (ITU Q.1901)
CPL	Call Processing Language
CSCF	Call State Control Function (3GPP)
DTMF	Dual Tone/Multiple Frequency
ENUM	E.164 Numbering (IETF RFC 2916)
GK	Gatekeeper
GPRS	General Packet Radio Service
GSM	Global System for Mobility
IAD	Integrated Access Device
IETF	Internet Engineering Task Force
IN	Intelligent Network
INAP	Intelligent Network Application Protocol
ISDN	Integrated Services Digital Network
ISUP	Integrated Services Digital Network User Part (SS7)
ITU	International Telecommunications Union
IUA	ISDN User Adaptation
IVR	Interactive Voice Response
JAIN	Java Application Interface Network
LDAP	Lightweight Directory Access Protocol (IETF)
M3UA	MTP3 User Adaptation (IETF SIGTRAN)
MEGACO	MEdia GAteway COntrol (IETF RFC 3015 or ITU H.248)
MG	Media Gateway
MGC	Media Gateway Controller
MGC-F	Media Gateway Controller Function (IPCC)
MGCP	Media Gateway Control Protocol (IETF, ITU-T J.162)
MPLS	Multi-Protocol Label Switching
MS	Media Server
MSC	Mobile Services Switching Center (GSM, 3GPP)
MTA	Multimedia Terminal Adaptor (PacketCable)
NCS	Network Call/Control Signaling (PacketCable MGCP)
NGN	Next Generation Network
OSS	Operational Support System
PBX	Private Branch eXchange
POTS	Plain Old Telephone Service
PSE	Personal Service Environment (3GPP)
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RAN	Radio Access Network
RFC	Request For Comment (IETF)
RG	Residential Gateway
RSVP	Resource ReSerVation Protocol (IETF)
RTCP	Real Time Transport Control Protocol (IETF)
RTP	Real Time Transport Protocol (IETF RFC 1889)
SCP	Service Control Point
SCTP	Stream Control Transmission Protocol
SDP	Session Description Protocol (IETF RFC 2327)
SG	Signaling Gateway
SIGTRAN	SIGnaling TRANsport (IETF)
SIP	Session Initiation Protocol (IETF)
SIP-T	SIP For Telephony (IETF)
SS7	Signaling System 7 (ITU)
TDM	Time Division Multiplexing
TRIP	Telephony Routing over IP (IETF RFC 2871)
UMTS	Universal Mobile Telecommunications System
VAD	Voice Activity Detection
VLR	Visitor Location Register (GSM, 3GPP)
VoDSL	Voice over DSL
VoIP	Voice over IP
VoP	Voice over Packet

References

Internet message format. RFC 2822, Internet Engineering Task Force, April 2001. URL http://www.rfc-editor.org/rfc/rfc2822.txt.

Session initiation protocol (SIP) extension for instant messaging. RFC 3428, Internet Engineering Task Force, December 2002. URL http://www.rfc-editor.org/rfc/rfc3428.txt.

3GPP. Amr wideband speech codec: Frame structure. Technical Specification TR 26.201, 3rd Generation Partnership Project (3GPP), March 2001a.

3GPP. Amr wideband speech codec: Transcoding functions. Technical Specification TR 26.190, 3rd Generation Partnership Project (3GPP), March 2001b.

S. C. Andersen et al. Internet low bit rate codec. Internet draft, Internet Engineering Task Force, July 2003. URL http://www.ietf.org/internet-drafts/draft-ietf-avt-ilbc-codec-02.txt. Work in progress.

Anonymous. Special issue on packet switched voice and data communication. *IEEE Journal on Selected Areas in Communications*, SAC-1(6), December 1983.

M. Arango, A. Dugan, I. Elliott, C. Huitema, and S. Pickett. Media gateway control protocol (MGCP) version 1.0. RFC 2705, Internet Engineering Task Force, October 1999. URL http://www.rfc-editor.org/rfc/rfc2705.txt.

M. Bangalore et al. A telephony gateway REgistration protocol (TGREP). Internet draft, Internet Engineering Task Force, July 2003. URL http://www.ietf.org/internet-drafts/draft-ietf-iptel-tgrep-02.txt. Work in progress.

Mark Baugher et al. The secure real-time transport protocol. Internet draft, Internet Engineering Task Force, July 2003. URL http://www.ietf.org/internet-drafts/draft-ietf-avt-srtp-09.txt. Work in progress.

J. C. Bolot, H. Crepin, and Anilton Garcia. Analysis of audio packet loss in the Internet. In *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Lecture Notes in Computer Science, pages 163–174, Durham, New Hampshire, April 1995. Springer. URL http://www.nossdav.org/1995/papers/bolot.ps.

Catherine Boutremans and Jean-Yves Le Boudec. Adaptive delay aware error control for Internet telephony. In *Internet Telephony Workshop*, New York, April 2001. URL http://www.cs.columbia.edu/ hgs/papers/iptel2001/34.ps.

CableLabs. Packetcable network-based call signaling protocol specification. Specification PKT-SP-EC-MGCP-I07-0, Cable Television Laboratories, April 2003. URL http://www.packetcable.com/downloads/specs/PKT-SP-MGCP-I07-030415.pdf.

E. J. Cameron, N. Griffeth, Y. Lin, Margaret E. Nilson, William K. Schure, and Hugo Velthuijsen. A feature interaction benchmark for IN and beyond. In *Feature Interactions in Telecommunications Systems*, pages 1–23, Amsterdam, Netherlands, 1994. ISBN 90-5199-165-7.

S. J. Campanella. Digital speech interpolation techniques. In *Conference record of the IEEE National Telecommunications Conference*, volume 1, pages 14.1.1–14.1.5, Birmingham, Alabama, December 1978. IEEE.

G. Carle and Ernst Biersack. Survey of error recovery techniques for IP-Based audio-visual multicast applications. *IEEE Network*, 11(6):24–36, November 1997. URL `http://207.127.135.8/ni/private/1997/nov/Carle.html`, `http://www.eurecom.fr/ btroup`.

Stephen Casner and S. E. Deering. First IETF Internet audiocast. *ACM Computer Communication Review*, 22(3):92–97, July 1992. URL `http://www.acm.org/sigcomm/ccr/archive/1992/jul92/casner.ps`.

S. Chuang, Jon Crowcroft, S. Hailes, Mark Handley, N. Ismail, D. Lewis, and Ian Wakeman. Multimedia application requirements for multicast communications services. In *International Networking Conference (INET)*, pages BFB–1–BFB–9, San Francisco, California, August 1993. Internet Society.

David D. Clark. A taxonomy of Internet telephony applications. In *25th Telecommunications Policy Research Conference*, Washington, DC, September 1997. URL `http://itc.mit.edu/itel/pubs/ddc.tprc97.pdf`.

R. Cogger. CU-SeeMe cornell desktop video, December 1992.

D. Cohen. Specifications for the network voice protocol (NVP). RFC 741, Internet Engineering Task Force, November 1977a. URL `http://www.rfc-editor.org/rfc/rfc741.txt`.

Danny Cohen. The network voice conference protocol (NVCP). NSC Note 113, February 1976a.

Danny Cohen. Specifications for the network voice protocol. Technical Report ISI/RR-75-39 (AD A02, USC/Information Sciences Institute, Marina del Rey, California, March 1976b. Available from DTIC).

Danny Cohen. Issues in transnet packetized voice communications. In *5th Data Communications Symposium*, pages 6–10–6–13, Snowbird, Utah, September 1977b. ACM, IEEE.

Danny Cohen. A protocol for packet-switching voice communication. *Computer Networks*, 2(4/5): 320–331, September/October 1978.

T. Dierks and C. Allen. The TLS protocol version 1.0. RFC 2246, Internet Engineering Task Force, January 1999. URL `http://www.rfc-editor.org/rfc/rfc2246.txt`.

S. Donovan. The SIP INFO method. RFC 2976, Internet Engineering Task Force, October 2000. URL `http://www.rfc-editor.org/rfc/rfc2976.txt`.

Robert E. Easton, P. T. Hutchison, Richard W. Kolor, Richard C. Mondello, and Richard W. Muise. TASI-E communications system. *IEEE Transactions on Communications*, COM-30(4):803–807, April 1982.

Hans Eriksson. MBone – the multicast backbone. In *International Networking Conference (INET)*, pages CCC–1–CCC–5, San Francisco, California, August 1993. Internet Society.

P. Faltstrom. E.164 number and DNS. RFC 2916, Internet Engineering Task Force, September 2000. URL `http://www.rfc-editor.org/rfc/rfc2916.txt`.

P. Faltstrom and M. Mealling. The E.164 to URI DDDS application

(ENUM). Internet draft, Internet Engineering Task Force, May 2003. URL http://www.ietf.org/internet-drafts/draft-ietf-enum-rfc2916bis-06.txt. Work in progress.

R. Fielding, J. Gettys, J. C. Mogul, H. Frystyk, L. Masinter, P. J. Leach, and T. Berners-Lee. Hypertext transfer protocol – HTTP/1.1. RFC 2616, Internet Engineering Task Force, June 1999. URL http://www.rfc-editor.org/rfc/rfc2616.txt.

Keir Fraser et al. Over-all characteristics of a TASI-system. *Bell System Technical Journal*, 41: 1439–1473, 1962.

Timur Friedman et al. RTP control protocol extended reports (RTCP XR). Internet draft, Internet Engineering Task Force, May 2003. URL draft-ietf-avt-rtcp-report-extns-06.txt,.pdf. Work in progress.

Hans Gierlich and Frank Kettler. Conversational speech quality - the dominating parameters in VoIP systems. In *Internet Telephony Workshop*, New York, April 2001. URL http://www.cs.columbia.edu/ hgs/papers/iptel2001/9.ps.

Josef Glasmann, Wolfgang Kellerer, and Harald Mller. Service development and deployment in H.323 and SIP. In *IEEE Symposium on Computers and Communications*, pages 378–385, Hammamet, Tunisia, July 2001. IEEE.

C. Groves, M. Pantaleo, Thomas Anderson, Tracy M. Taylor, and Editors. Gateway control protocol version 1. RFC 3525, Internet Engineering Task Force, June 2003. URL http://www.rfc-editor.org/rfc/rfc3525.txt.

A. Gulbrandsen, P. Vixie, and L. Esibov. A DNS RR for specifying the location of services (DNS SRV). RFC 2782, Internet Engineering Task Force, February 2000. URL http://www.rfc-editor.org/rfc/rfc2782.txt.

M. Handley and V. Jacobson. SDP: session description protocol. RFC 2327, Internet Engineering Task Force, April 1998. URL http://www.rfc-editor.org/rfc/rfc2327.txt.

G. Herlein et al. RTP payload format for the speex codec. Internet draft, Internet Engineering Task Force, July 2003. URL http://www.ietf.org/internet-drafts/draft-herlein-speex-rtp-profile-01.txt. Work in progress.

P. Hoffman and F. Yergeau. UTF-16, an encoding of ISO 10646. RFC 2781, Internet Engineering Task Force, February 2000. URL http://www.rfc-editor.org/rfc/rfc2781.txt.

International Packet Communications Consortium. http://www.softswitch.org/.

International Telecommunication Union. Voice packetization - packetized voice protocols. Recommendation G.764, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, 1990. URL http://www.itu.int/itudoc/itu-t/rec/g/g700-799/g764.html.

International Telecommunication Union. Digital subscriber signalling system no. 1 (DSS 1) - ISDN user-network interface layer 3 specification for basic call control. Recommendation Q.931, International Telecommunication Union, Geneva, Switzerland, March 1993a. URL http://www.itu.int/itudocs/itu-t/rec/q/q500-999/q931_24961.html.

International Telecommunication Union. Video codec for audiovisual services at px64 kbit/s. Recommendation H.261, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, March 1993b.

International Telecommunication Union. Coding of speech at 8 kbit/s using conjugatestructure algebraic-code-excited linear-prediction. Recommendation G.729, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, March 1996a. URL http://www.itu.int/itudoc/itu-t/rec/g/g700-799/g729_32350.html.

International Telecommunication Union. Data protocols for multimedia conferencing. Recommendation T.120, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, July 1996b. URL http://www.itu.int.

International Telecommunication Union. Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s. Recommendation G.723.1, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, March 1996c. URL http://www.itu.int/itudoc/itu-t/rec/g/g700-799/g723-1.html.

International Telecommunication Union. Media stream packetization and synchronization on non-guaranteed quality of service LANs. Recommendation H.225.0, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, November 1996d. URL http://www.itu.int.

International Telecommunication Union. ASN.1 encoding rules - specification of packed encoding rules (PER). Recommendation X.691, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, December 1997a. URL http://www.itu.int.

International Telecommunication Union. The international public telecommunication numbering plan. Recommendation E.164, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, May 1997b. URL http://www.itu.int.

International Telecommunication Union. Control protocol for multimedia communication. Recommendation H.245, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, February 1998a. URL http://www.itu.int.

International Telecommunication Union. Pulse code modulation (PCM) of voice frequencies. Recommendation G.711, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, November 1998b.

International Telecommunication Union. Video coding for low bit rate communication. Recommendation H.263, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, February 1998c.

International Telecommunication Union. Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss. Recommendation G.722.1, International Telecommunication Union, September 1999a. URL http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-REC-G.722.1.

International Telecommunication Union. Narrow-band visual telephone systems and terminal equipment. Recommendation H.320, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, May 1999b. URL http://www.itu.int/itu-t/rec/h/h320.html.

International Telecommunication Union. Packet based multimedia communication systems. Recommendation H.323, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, November 2000. URL http://www.itu.int/.

International Telecommunication Union. Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB). Recommendation, International Telecommunication Union, January 2002. URL http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-REC-G.722.2.

V. Jacobson. Multimedia conferencing on the Internet. In *SIGCOMM Symposium on Communications Architectures and Protocols*, London, England, August 1994. URL ftp://cs.ucl.ac.uk/darpa/vjtut.ps.Z. Tutorial slides.

V. Jacobson and Steve McCanne. vat - LBNL audio conferencing tool, July 1992. URL http://www-nrg.ee.lbl.gov/vat/. Available at http://www-nrg.ee.lbl.gov/vat/.

K. Jeffay, D. Stone, and F. Smith. Transport and display mechanisms for multimedia conferencing across packet-switched networks. *Computer Networks and ISDN Systems*, 26(10):1281–1304, July 1994. URL http://www.elsevier.com/locate/comnet.

Wenyu Jiang, Kazuumi Koguchi, and Henning Schulzrinne. QoS evaluation of VoIP end-points. In *Conference Record of the International Conference on Communications (ICC)*, May 2003.

Wenyu Jiang and Henning Schulzrinne. Analysis of on-off patterns in VoIP and their effect on voice traffic aggregation. In *International Conference on Computer Communication and Network*, Las Vegas, Nevada, October 2000a. URL http://www.cs.columbia.edu/IRT/papers/Jian0010_Analysis.pdf.

Wenyu Jiang and Henning Schulzrinne. Modeling of packet loss and delay and their effect on real-time multimedia service quality. In *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, June 2000b. URL http://www.nossdav.org/2000/papers/27.pdf.

Wenyu Jiang and Henning Schulzrinne. Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss. In *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Miami Beach, Florida, May 2002a. URL http://www.cs.columbia.edu/IRT/papers/Jian0205_Comparison.pdf.

Wenyu Jiang and Henning Schulzrinne. Comparisons of FEC and codec robustness on VoIP quality and bandwidth efficiency. In *ICN*, Atlanta, Georgia, August 2002b. URL http://www.cs.columbia.edu/IRT/papers/Jian0208_Comparisons.pdf.

Wenyu Jiang and Henning Schulzrinne. Speech recognition performance as an effective perceived quality predictor. In *IWQoS*, Miami Beach, May 2002c. URL http://www.cs.columbia.edu/ hgs/papers/Jian0205:Speech.pdf.

Wenyu Jiang and Henning Schulzrinne. Assessment of VoIP service availability in the current Internet. In *Passive & Active Measurement Workshop*, San Diego, CA, April 2003. URL http://www.cs.columbia.edu/IRT/papers/Jian0304_Assessment.pdf.

A. R. Johnston. Session initiation protocol basic call flow examples. Internet draft, Internet Engineering Task Force, April 2003. URL http://www.ietf.org/internet-drafts/draft-ietf-sipping-basic-call-flows-02.txt. Work in progress.

D. Kutscher, Juerg Ott, and Carsten Bormann. Session description and capability negotiation. Internet draft, Internet Engineering Task Force, March 2003. URL http://www.ietf.org/internet-drafts/draft-ietf-mmusic-sdpng-06.txt. Work in progress.

J. Lennox and Henning Schulzrinne. Call processing language framework and requirements. RFC 2824, Internet Engineering Task Force, May 2000a. URL http://www.rfc-editor.org/rfc/rfc2824.txt.

J. Lennox, Henning Schulzrinne, and Thomas La Porta. Implementing intelligent network services with the session initiation protocol. Technical Report CUCS-002-99, Columbia University, New York, New York, January 1999. URL ftp://ftp.cs.columbia.edu/reports/reports-1999/cucs-002-99.ps.gz.

J. Lennox, Henning Schulzrinne, and J. Rosenberg. Common gateway interface for SIP. RFC 3050, Internet Engineering Task Force, January 2001. URL http://www.rfc-editor.org/rfc/rfc3050.txt.

Jonathan Lennox and Henning Schulzrinne. Feature interaction in Internet telephony. In *Feature Interaction in Telecommunications and Software Systems VI*, Glasgow, United Kingdom, May 2000b. URL http://www.cs.columbia.edu/IRT/papers/Lenn0005_Feature.pdf.

Jonathan Lennox, Xinzhou Wu, and Henning Schulzrinne. CPL: a language for user control of Internet telephony services. Internet draft, Internet Engineering Task Force, August 2003. URL draft-ietf-iptel-cpl-07.txt,.ps. Work in progress.

Hong Liu and Petros N. Mouchtaris. Voice over IP signaling: H.323 and beyond. *IEEE Communications Magazine*, 38(10), October 2000. URL http://www.comsoc.org/livepubs/ci1/public/2000/oct/index.html.

P. Luthi. RTP payload format for ITU-T recommendation G.722.1. RFC 3047, Internet Engineering Task Force, January 2001. URL http://www.rfc-editor.org/rfc/rfc3047.txt.

D. T. Magill. Adaptive speech compression for packet communication systems. In *Conference record of the IEEE National Telecommunications Conference*, pages 29D–1–29D–5, 1973.

Lee McKnight. Internet telephony markets: 2000 – 3001. In *Carrier Class IP Telephony*, San Diego, California, January 2000. URL http://www.cs.columbia.edu/IRT/papers/others/McKn0001_Internet.ppt.gz.

M. Mealling. Dynamic delegation discovery system (DDDS) part one: The comprehensive DDDS. RFC 3401, Internet Engineering Task Force, October 2002. URL http://www.rfc-editor.org/rfc/rfc3401.txt.

H. Miedema and M. G. Schachtman. TASI quality — effect of speech detectors and interpolation. *Bell System Technical Journal*, 41(4):1455–1473, July 1962.

Ed Miller, Flemming Andreasen, and Glenn Russell. The packetcable architecture. *IEEE Communications Magazine*, 39(6), June 2001. URL http://www.comsoc.org/livepubs/ci1/public/2001/jun/index.html.

Warren A. Montgomery. Techniques for packet voice synchronization. *IEEE Journal on Selected Areas in Communications*, SAC-1(6):1022–1028, December 1983.

Sue Moon, James F. Kurose, and Donald F. Towsley. Packet audio playout delay adjustment: performance bounds and algorithms. *Multimedia Systems*, 5(1):17–28, January 1998. URL ftp://gaia.cs.umass.edu/pub/Moon95_Packet.ps.gz.

Pete Moulton and Jeremy Moulton. Telecommunications technical fundamentals. Technical handout, The Moulton Company, Columbia, Maryland, 1996. URL http://www.moultonco.com/semnotes/telecomm/teladd.htm. see http://www.moultonco.com/semnotes/telecomm/teladd.htm.

Frank M. Ploumen and Luc de Clercq. The all-digital loop: benefits of an integrated voice-data access network. In *Communication Technology (ICCT)*, Beijing, China, August 2000. IEEE.

John Postel. Internet control message protocol. RFC 792, Internet Engineering Task Force, September 1981. URL http://www.rfc-editor.org/rfc/rfc792.txt.

R. Ramjee, James F. Kurose, Donald F. Towsley, and Henning Schulzrinne. Adaptive playout mechanisms for packetized audio applications in wide-area networks. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, pages 680–688, Toronto, Canada, June 1994. IEEE Computer Society Press, Los Alamitos, California. URL `ftp://gaia.cs.umass.edu/pub/Ramj94:Adaptive.ps.Z`.

J. H. Rieser, H. G. Suyderhood, and Y. Yatsuzuka. Design considerations for digital speech interpolation. In *Conference Record of the International Conference on Communications (ICC)*, pages 49.4.1–49.4.7, Denver, Colorado, June 1981. IEEE.

A. B. Roach. Session initiation protocol (sip)-specific event notification. RFC 3265, Internet Engineering Task Force, June 2002. URL `http://www.rfc-editor.org/rfc/rfc3265.txt`.

J. Rosenberg. The session initiation protocol (SIP) UPDATE method. RFC 3311, Internet Engineering Task Force, October 2002. URL `http://www.rfc-editor.org/rfc/rfc3311.txt`.

J. Rosenberg, H. F. Salama, and M. Squire. Telephony routing over IP (TRIP). RFC 3219, Internet Engineering Task Force, January 2002a. URL `http://www.rfc-editor.org/rfc/rfc3219.txt`.

J. Rosenberg and Henning Schulzrinne. An RTP payload format for generic forward error correction. RFC 2733, Internet Engineering Task Force, December 1999. URL `http://www.rfc-editor.org/rfc/rfc2733.txt`.

J. Rosenberg and Henning Schulzrinne. Reliability of provisional responses in session initiation protocol (SIP). RFC 3262, Internet Engineering Task Force, June 2002. URL `http://www.rfc-editor.org/rfc/rfc3262.txt`.

J. Rosenberg, Henning Schulzrinne, G. Camarillo, A. R. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: session initiation protocol. RFC 3261, Internet Engineering Task Force, June 2002b. URL `http://www.rfc-editor.org/rfc/rfc3261.txt`.

Jonathan Rosenberg et al. Indicating user agent capabilities in the session initiation protocol (SIP). Internet draft, Internet Engineering Task Force, June 2003a. URL `http://www.ietf.org/internet-drafts/draft-ietf-sip-callee-caps-00.txt`. Work in progress.

Jonathan Rosenberg and P. Kyzivat. Guidelines for usage of the session initiation protocol (SIP) caller preferences extension. Internet draft, Internet Engineering Task Force, July 2003. URL `http://www.ietf.org/internet-drafts/draft-ietf-sipping-callerprefs-usecases-00.txt`. Work in progress.

Jonathan Rosenberg, James L. Peterson, Henning Schulzrinne, and Gonzalo Camarillo. Best current practices for third party call control in the session initiation protocol. Internet draft, Internet Engineering Task Force, July 2003b. URL `http://www.ietf.org/internet-drafts/draft-ietf-sipping-3pcc-04.txt`. Work in progress.

Jonathan Rosenberg, Lili Qiu, and Henning Schulzrinne. Integrating packet FEC into adaptive voice playout buffer algorithms on the Internet. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, Tel Aviv, Israel, March 2000. URL `http://www.cs.columbia.edu/ hgs/papers/Rose0003_Integrating.pdf`.

Henning Schulzrinne. Voice communication across the internet: A network voice terminal. Technical Report TR 92-50, Dept. of Computer Science, University of Massachusetts, Amherst, Massachusetts, July 1992. URL http://www.cs.columbia.edu/ hgs/papers/Schu9207_Voice.ps.gz.

Henning Schulzrinne. RTP profile for audio and video conferences with minimal control. RFC 1890, Internet Engineering Task Force, January 1996. URL http://www.rfc-editor.org/rfc/rfc1890.txt.

Henning Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: a transport protocol for real-time applications. RFC 1889, Internet Engineering Task Force, January 1996. URL http://www.rfc-editor.org/rfc/rfc1889.txt.

Henning Schulzrinne and S. Petrack. RTP payload for DTMF digits, telephony tones and telephony signals. RFC 2833, Internet Engineering Task Force, May 2000. URL http://www.rfc-editor.org/rfc/rfc2833.txt.

Henning Schulzrinne and A. Vaha-Sipila. The tel URI for telephone calls. Internet draft, Internet Engineering Task Force, July 2003. URL http://www.ietf.org/internet-drafts/draft-ietf-iptel-rfc2806bis-02.txt. Work in progress.

Guido Schuster, Jerry Mahler, Ikhlaq Sidhu, and Michael S. Borella. Forward error correction system for packet based real time media. U.S. Patent US5870412, 3Com, Chicago, Illinois, February 1999. URL http://www.patents.ibm.com/patlist?icnt=US&patent_number=5870412.

Kundan Singh and Henning Schulzrinne. Interworking between SIP/SDP and H.323. In *IP-Telephony Workshop (IPtel)*, Berlin, Germany, April 2000. URL http://www.cs.columbia.edu/ hgs/papers/Sing0004_Interworking.pdf.

J. Sjoberg, M. Westerlund, A. Lakaniemi, and Q. Xie. Real-time transport protocol (RTP) payload format and file storage format for the adaptive multi-rate (AMR) and adaptive multi-rate wideband (AMR-WB) audio codecs. RFC 3267, Internet Engineering Task Force, June 2002. URL http://www.rfc-editor.org/rfc/rfc3267.txt.

R. Sparks. The session initiation protocol (SIP) refer method. RFC 3515, Internet Engineering Task Force, April 2003. URL http://www.rfc-editor.org/rfc/rfc3515.txt.

T. Stock and Xavier Garcia Adanez. On the potentials of forward error correction mechanisms applied to real-time services carried over B-ISDN. In Bernhard Plattner, editor, *International Zurich seminar on Digital Communications, IZS (Broadband Communications Networks, Services, Applications, Future Directions)*, Lecture Notes in Computer Science, pages 107–118, Zurich, Switzerland, February 1996. Springer-Verlag. URL http://tcomwww.epfl.ch/ garcia/publications/publications.html/izs96.ps.

Gary A. Thom. H.323: the multimedia communications standard for local area networks. *IEEE Communications Magazine*, 34(12), December 1996. URL http://www.comsoc.org/pubs/ci/comsoc/private/1996/dec/Thom.html.

Kiran Toutireddy and J. Padhye. Design and simulation of a zero redundancy forward error correction technique for packetized audio transmission. Project report, Univ. of Massachusetts, Amherst, Massachusetts, December 1995. URL http://www-ccs.cs.umass.edu/ jitu/fec/.

George Vinall. Economics of Internet telephony. In *Voice on the Net*, San Jose, California, March/April 1998. URL http://www.pulver.com/oldslides.

Clifford J. Weinstein and Edward M. Hofstetter. The tradeoff between delay and TASI advantage in a packetized speech multiplexer. *IEEE Transactions on Communications*, COM-27(11):1716–1720, November 1979.

Mark Allen Weiss and Jenq-Neng Hwang. Internet telephony or circuit switched telephony: Which is cheaper? In *Telecommunications Policy Research Conference*, Washington, DC, October 1998. URL `http://www2.sis.pitt.edu/ mweiss/papers/itel.pdf`.

Ronald J. Wocjik. Packetcable network architecture. In *Carrier Class IP Telephony*, San Diego, California, January 2000.

D. Wright. Voice over ATM: an evaluation of network architecture alternatives. *IEEE Network*, 10(5):22–27, September 1996. URL `http://207.127.135.8/ni/private/1996/sep/Wright.html`.

David J. Wright. Voice over MPLS compared to voice over other packet transport technologies. *IEEE Communications Magazine*, 40(11):124–132, November 2002.

Xiaotao Wu and Henning Schulzrinne. Where should services reside in Internet telephony systems? In *IP Telecom Services Workshop*, Atlanta, Georgia, September 2000. URL `http://www.cs.columbia.edu/ hgs/papers/Wu0009_Where.pdf`.

Xiaotao Wu and Henning Schulzrinne. Programmable end system services using SIP. In *Conference Record of the International Conference on Communications (ICC)*, May 2003.

W. Zhao and Henning Schulzrinne. Locating IP-to-Public switched telephone network (PSTN) telephony gateways via SLP. Internet draft, Internet Engineering Task Force, August 2002. URL `http://www.ietf.org/internet-drafts/draft-zhao-iptel-gwloc-slp-05.txt`. Work in progress.

R. Zopf. Real-time transport protocol (RTP) payload for comfort noise (CN). RFC 3389, Internet Engineering Task Force, September 2002. URL `http://www.rfc-editor.org/rfc/rfc3389.txt`.