Documentation

| HOME | CONTENTS | PREVIOUS | NEXT | GLOSSARY | FEEDBACK | SEARCH | HELP |

Table of Contents

Chapter Goals

- Provide an overview of technologies and applications of integrated voice/data networking.

- Outline the differences between the various voice/data integration technologies, and tell when each should be used.

- Understand the specific protocols involved in voice/data networking.

- List specific network engineering challenges and solutions associated with the integration of voice and data.

Voice/Data Integration Technologies

Introduction

Voice/data integration is important to network designers of both service providers and enterprise. Service providers are attracted by the lower-cost model—the cost of packet voice is currently estimated to be only 20 to 50 percent of the cost of a traditional circuit-based voice network. Likewise, enterprise network designers are interested in direct cost savings associated with toll-bypass and tandem switching. Both are also interested in so-called "soft savings" associated with reduced maintenance costs and more efficient network control and management. Finally, packet-based voice systems offer access to newly enhanced services such as Unified Messaging and application control. These, in turn, promise to increase the productivity of users and differentiate services.

Integration of voice and data technologies has accelerated rapidly in recent years because of both supply- and demand-side interactions. On the demand side, customers are leveraging investment in network infrastructure to take advantage of integrated applications such as voice applications. On the supply side, vendors have been able to take advantage of breakthroughs in many areas, including standards, technology, and network performance.

Standards

Many standards for interoperability for voice signaling have finally been ratified and matured to the point of reasonable interoperability. This reduces the risk and costs faced by vendors offering components of a voice/data system, and it also reduces the risk to consumers. Standards such as H.323 (approved by the ITU in June 1996), are now evolving through their third and fourth iterations, while products based on initial standards still enjoy strong capabilities and interoperability. The general maturity of standards has in turn generated robust protocol stacks that can be purchased "off the shelf" by vendors, further ensuring interoperability.

Technology

Recent advances in technology have also enabled voice integration with data. For example, new Digital Signal Processor (DSP) technology has allowed analog signals to be processed in the digital domain, which was difficult or impossible only a few years earlier. These powerful new chips offer tremendous processing speeds, allowing voice to be sampled, digitized, and compressed in real time. Further breakthroughs in the technology allow as many as four voice conversations to be managed at the same time on a single chip, with even greater performance in development. These technologies greatly reduce the cost and complexity of developing products and deploying voice over data solutions.

In other areas, the industry has also enjoyed breakthroughs in voice codec (coder/decoder) technology. Previously, it was assumed that voice quality would suffer as bandwidth was decreased in a relatively linear fashion. However, new, sophisticated algorithms employed in new codecs have changed that view. It is now possible to obtain reasonably good-sounding voice at a fraction of the bandwidth once required. More importantly, these new algorithms have been incorporated into the standards to allow interoperability of highly compressed voice.

Network Performance

Finally, data-networking technology has improved to the point that voice can be carried reliably. Over the last few years, growth in voice traffic has been relatively small, while data traffic has grown exponentially. The result is that data traffic is now greater than voice traffic in many networks. In addition, the relative importance of data traffic has grown, as businesses and organizations come to base more business practices and policies on the ubiquity of data networks. This increase in importance of data networks has forced a fundamental change in the way data networks are engineered, built, and managed. Typical "best-effort" data modeling has given way to advanced policy-based networking with managed quality of service to support an even greater range of applications. Voice traffic, as an application on a data network, has benefited greatly from these technologies. For example, support of delay-sensitive SNA traffic over IP networks resulted in breakthroughs in latency management and queuing prioritization, which was then applied to voice traffic.

As stated previously, deployment of new technologies and applications must also be driven by greater demand from users. Breakthroughs in technology don't necessarily result in increased deployment unless they fill a real user need at a reasonable cost. For example, digital audio tape (DAT) technologies never enjoyed widespread use outside the audiophile community because of the high cost and only marginally better perceived performance than analog tapes. Voice/data integration, however, provides users with very real benefits, both now and in the future. Most users of voice/data integration technologies gain in two ways: Packet voice technologies are less expensive, and, in the future, they will offer much greater capabilities compared to today's circuit-based voice systems.

Economic Advantages

It has been estimated that packet voice networking costs only 20 to 30 percent of an equivalent circuit-based voice network. This is true for both carriers (service providers) and enterprise (private) users. Logically, this implies that enterprise users can operate long-distance voice services between facilities at less cost than purchasing long-distance voice services from a carrier, and it's often true. For example, many enterprise users have deployed integrated voice/data technologies to transport voice over data wide-area networks (WANs) between traditional PBXs across different geographical locations. The resulting savings in long-distance toll

charges often provide payback in as little as six months (especially if international calls are avoided). Using data systems to carry voice as "virtual tie lines" between switches is also useful to service providers. In fact, many new carriers have started to embrace packet-based voice technologies as their primary network infrastructure strategy going forward.

However, savings associated with packet voice technologies don't stop with simple transport. It is also possible to switch voice calls in the data domain more economically than traditional circuit-based voice switches. For large, multisite enterprises, the savings result from using the data network to act as a "tandem switch" to route voice calls between PBXs on a call-by-call basis. The resulting voice network structure is simpler to administer and uses a robust, nonblocking switching fabric made up of data systems at its core.

Advances in Applications

Real cost savings are sufficient for deployment of voice/data integration technologies. However, there are added benefits, which will become more evident in the future. As applications evolve, organizations will gain increased user productivity from the integration of voice and computer applications. Computer telephony integration (CTI) was begun by PBX vendors in the 1980s to integrate computers with PBXs to provide applications such as advanced call center features (for example, "screen pops" for agents).

However, as voice/data integration continues, the line between voice and data applications will continue to blur. For example, Unified Messaging systems are now available that combine voice mail, e-mail, and fax messaging into a single, convenient system. With these advanced systems, users can have e-mail read to them over the phone or can add document attachments to voice mail. At the enterprise level, new applications such as virtual call centers allow call center agents to be distributed anywhere within reach of the data network, while still receiving the full suite of call center functions and features. They can even receive calls over their computers rather than using a traditional telephone instrument, and they can provide "blended contact center" support to answer Web user questions with electronic chat capability and e-mail between voice calls. These capabilities go far beyond simple cost savings and will ultimately make organizations much more effective and profitable.

The strong pressures driving the integration of voice and data networks have resulted in various solutions to the problem, each with its own strengths and weaknesses. Three general approaches exist:

- Voice over ATM

- Voice over Frame Relay

- Voice over IP

There are also mixed solutions, including voice over IP, over Frame Relay, and so on. These are illustrated in Figure 19-1 The figure shows that voice over ATM and voice over Frame Relay are primarily transport mechanisms between PBXs, while voice over IP can connect all the way to the desktop. More details are available later in this chapter.

Figure 19-1: Mixed Solutions Including Voice over IP, Voice over Frame Relay, and so on.

Voice over IP
(over Frame Relay or ATM)

Voice over Frame
Relay or ATM

PBX

PBX

PBX

PBX

PBX

PBX
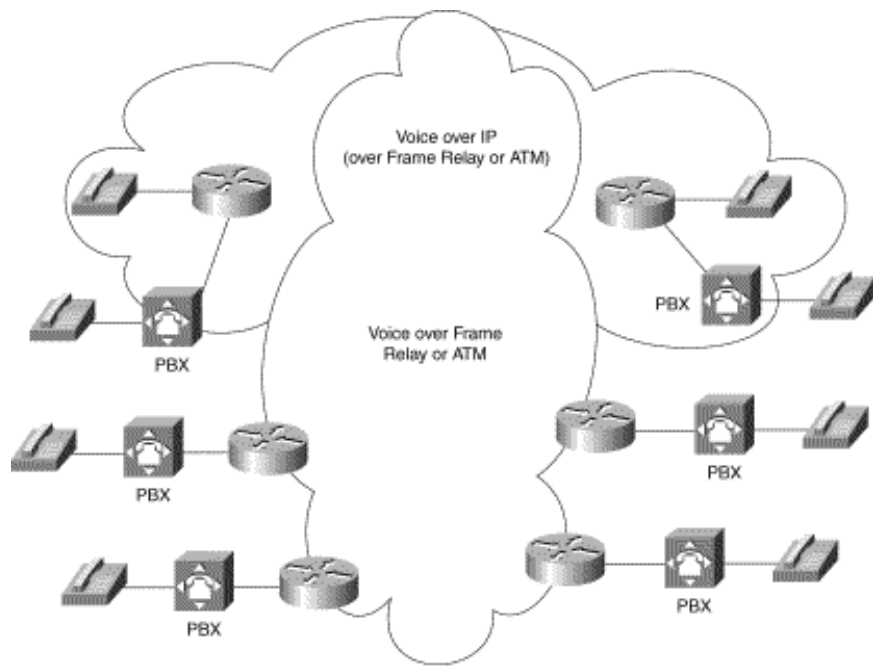
ct841901.eps
87050013
10/01/00
steve gifford

Voice over ATM

Voice over ATM (VoATM) can be supported as standard pulse code modulated (PCM) voice via circuit emulation (AAL1, described later) or as variable bit rate voice in ATM cells as AAL2 (also described later). ATM offers many advantages for transport and switching of voice. First, quality of service (QoS) guarantees can be specified by service provisioning or on a per-call basis. In addition, call setup signaling for ATM switched virtual circuits (SVCs), Q.2931, is based on call setup signaling for voice ISDN, Q.931. Administration is similar to circuit-based voice networks.

However, VoATM suffers from the burden of additional complexity and incomplete support and interoperability among vendors. It also tends to be more expensive because it is oriented toward all optical networks. Most importantly, ATM is typically deployed
as a WAN Layer 2 protocol and therefore does not extend all the way to the desktop. Nevertheless, ATM is quite effective for providing trunking and tandem switching

services between existing voice switches and PBXs.

Voice over Frame Relay (VoFR) has become widely deployed across many networks. Like VoATM, it is typically employed as a tie trunk or tandem-switching function between remote PBXs. It benefits from much simpler administration and relatively lower cost than VoATM, especially when deployed over a private WAN network. It also scales more economically than VoATM, supporting links from T1 down to 56 kbps. When deployed over a carefully engineered Frame Relay network, VoFR works very well and provides good quality. However, voice quality over Frame Relay can suffer depending on network latency and jitter. Although minimal bandwidth and burstiness are routinely contracted, latency and jitter are often not included in service level agreements (SLAs) with service providers. As a result, voice performance can vary. Even if quality is good at first, voice quality can degrade over time as a service provider's network becomes saturated with more traffic. For this reason, many large enterprise customers are beginning to specify latency and jitter, as well as overall packet throughput from carriers. In these situations, voice over Frame Relay can provide excellent service.

Voice over IP (VoIP) has begun to be deployed in recent years as well. Unlike voice over Frame Relay and Voice over ATM, Voice over IP is a Layer 3 solution, and it offers much more value and utility because IP goes all the way to the desktop. This means that in addition to providing basic tie trunk and tandem-switching functions to PBXs, VoIP can actually begin to replace those PBXs as an application. As a Layer 3 solution, VoIP is routable and can be carried transparently over any type of network infrastructure, including both Frame Relay and ATM. Of all the packet voice technologies, VoIP has perhaps the most difficult time supporting voice quality because QoS cannot be guaranteed. Normal applications such as TCP running on IP are insensitive to latency but must retransmit lost packets due to collisions or congestion. Voice is much more sensitive to packet delay than packet loss. In addition to normal traffic congestion, QoS for VoIP is often dependent on lower layers that are ignorant of the voice traffic mingled with the data traffic.

Voice Networking

Basic voice technology has been available for more than 100 years. During that time, the technology has matured to the point at which it has become ubiquitous and largely invisible to most users. This legacy of slow evolution continues to affect today's advanced voice networks in many ways, so it is important to understand the fundamentals of traditional voice technology before emulating it on data networks.

Traditional analog telephone instruments used for plain old telephone service (POTS) use a simple two-wire interface to the network. They rely on an internal two-wire/four-wire hybrid circuit to combine both transmit and receive signals. This economical approach has been effective but requires special engineering regarding echo.

Basic Telephony

Three types of signaling are required for traditional telephony: supervision, alerting, and addressing. Supervision monitors the state of the instrument—for example, allowing the central office or PBX to know when the receiver has been picked up to make a call, or when a call is terminated. Alerting concerns the notification of a user that a call is present (ringing) or simple call progress tones during a call (such as busy, ringback, and so on). Finally, addressing enables the user to dial a specific extension.

In addition to signaling, telephony services also provide secure media transport for the voice itself, analog-to-digital conversion, bonding and grounding for safety, power, and a variety of other functions when needed.

Analog voice interfaces have evolved over the years to provide for these basic functions while addressing specific applications. Because basic POTS two-wire analog interfaces operate in a master/slave model, two basic types of analog interfaces are necessary for data equipment to emulate: the user side and the network side. The user side (telephone) expects to receive power from the network as well as supervision.

A foreign exchange service (FXS) interface is used to connect an analog telephone, fax machine, modem, or any other device that would be connected to a phone line. It outputs 48 vdc power, ringing, and so on, and it accepts dialed digits. The opposite of an FXS interface is a foreign exchange office (FXO) interface. It is used to connect to a switching system providing services and supervision, and it expects the switch to provide supervision and other elements. (Why "foreign"? The terms *FXS* and *FXO* were originally used within telephone company networks to describe provision of telephone service from a central office other than normally assigned.)

Within FXS and FXO interfaces, it is also necessary to emulate variants in supervision. Typical telephones operate in a loop start mode. The telephone normally presents a high impedance between the two wires. When the receiver goes off-hook, a low-impedance closed circuit is created between the two wires. The switch, sensing current flow, then knows that the receiver is off-hook and applies a dial tone. The switch also checks to be sure that the receiver is on-hook before sending a ringing signal. This system works well for simple telephones, but it can cause problems on trunks between PBXs and COs with high activity. In that situation, the remote end and the CO switch can both try to seize the line at the same time. This situation, called *glare*, can freeze the trunk until one side releases it. The solution is to short tip or ring to ground as a signal for line seizure rather than looping it. This is called *ground start*.

After the line is seized, it is necessary to dial the number. Normal human fingers cannot outrun the dial receivers in a modern switch, but digits dialed by a PBX can. In that case, many analog trunks use a delay start or wink start method to notify the calling device when the switch is ready to accept digits.

Another analog interface often used for trunking is E&M. This is a four- or six-wire interface that includes separate wires for supervision in addition to the voice pair. *E&M* stands for "ear and mouth" or "Earth and magneto" and is derived from the early telephony days. The E&M leads are used to signal on-hook and off-hook states.

Analog voice works well for basic trunk connections between switches or PBXs, but it is uneconomical when the number of connections exceeds six to eight circuits. At that point, it is usually more efficient to use digital trunks. In North America, the T1 (1.544 Mbps) trunk speed is used, consisting of 24 digitized analog voice conversations. In other parts of the world, E1 (2.048 Mbps) is used to carry 30 voice channels. (Engineers refer to the adoption of E1 and T1 internationally as "the baseball rule"—there is a strong correlation of countries that play baseball to the use of T1. Therefore, the United States, Canada, and Japan have the largest T1 networks, while other countries use E1.)

The first step in conversion to digital is sampling. The Nyquist theorem states that the sampling frequency should be twice the rate of the highest desired frequency. Early telephony engineers decided that a range of 4000 hertz would be sufficient to capture human voices (which matches the performance of long analog loops). Therefore, voice channels are sampled at a rate of 8000 times per second, or once every 125 ms. Each of these samples consists of an 8-bit measurement, for a total of 64000 bits per second to be transmitted. As a final step, companding is used to provide greater accuracy of low-amplitude components. In North America, this is u-law (mu-law), while elsewhere it is typically A-law. For international interworking purposes, it is agreed that the North American side will make the conversion.

To construct a T1, 24 channels are assembled for a total of 1.536 Mbps, and an additional 8 bits are added every 125 ms for framing, resulting in a rate of 1.544 Mbps. Often, T1 frames are combined into larger structures called SuperFrames (12 frames) and Extended-SuperFrames (24 frames). Additional signaling can then be transmitted by "robbing bits" from the interior frames.

Basic T1 and E1 interfaces emulate a collection of analog voice trunks and use robbed bit signaling to transfer supervisory information similar to the E&M analog model. As such, each channel carries its own signaling, and the interface is called channel associated signaling (CAS). A more efficient method uses a common signaling channel for all the voice channels. Primary Rate Interface for ISDN is the most common example of this common channel signaling (CCS).

If voice/data integration is to be successful, all of these voice interfaces must be supported to provide the widest possible range of applications. Over the years, users have grown to expect a certain level of performance, reliability, and behavior of a telecommunications system, which must be supported going forward. All these issues have been solved by various packet voice systems today so that users can enjoy the same level of support to which they have become accustomed.

Voice over ATM

The ATM Forum and the ITU have specified different classes of services to represent different possible traffic types for VoATM.

Designed primarily for voice communications, constant bit rate (CBR) and variable bit rate (VBR) classes have provisions for passing real-time traffic and are suitable for guaranteeing a certain level of service. CBR, in particular, allows the amount of bandwidth, end-to-end delay, and delay variation to be specified during the call setup.

Designed principally for bursty traffic, unspecified bit rate (UBR) and available bit rate (ABR) are more suitable for data applications. UBR, in particular, makes no guarantees about the delivery of the data traffic.

The method of transporting voice channels through an ATM network depends on the nature of the traffic. Different ATM adaptation types have been developed for different traffic types, each with its benefits and detriments. ATM adaptation layer 1 (AAL1) is the most common adaptation layer used with CBR services.

Unstructured AAL1 takes a continuous bit stream and places it within ATM cells. This is a common method of supporting a full E1 byte stream from end to end. The problem with this approach is that a full E1 may be sent, regardless of the actual number of voice channels in use. (An EI is a wide-area digital transmission scheme used predominantly in Europe that carries data at a rate of 2.048 Mbps.)

Structured AAL1 contains a pointer in the payload that allows the digital signal level 0 (DS0) structure to be maintained in subsequent cells. This allows network efficiencies to be gained by not using bandwidth for unused DS0s. (A DS0 is a framing specification used in transmitting digital signals over a single channel at 64 kbps on a T1 facility.)

The remapping option allows the ATM network to terminate structured AAL1 cells and remap DS0s to the proper destinations. This eliminates the need for permanent virtual circuits (PVCs) between every possible source/destination combination. The major difference from the previous approach is that a PVC is not built across the network from edge to edge.

VoATM Signaling

Figure 19-2 describes the transport method, in which voice signaling is carried through the network transparently. PVCs are created for both signaling and voice transport. First, a signaling message is carried transparently over the signaling PVC from end station to end station. Second, coordination between the end systems allows the selection of a PVC to carry the voice communication between end stations.

Figure 19-2: The VoATM Signaling Transport Model Describes the Transport Method, in Which Voice Signaling Is Carried Through the Network Transparently
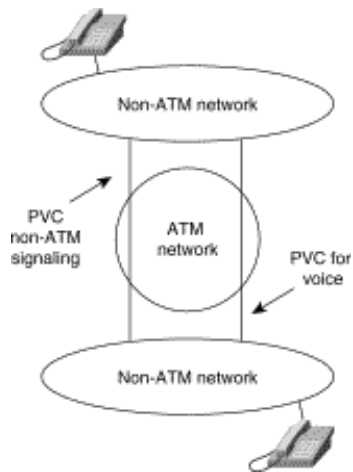
At no time is the ATM network participating in the interpretation of the signaling that takes place between end stations. However, as a value-added feature, some products are capable of understanding channel associated signaling (CAS) and can prevent the sending of empty voice cells when the end stations are on-hook.

Figure 19-3 shows the translate model. In this model, the ATM network interprets the signaling from both non-ATM and ATM network devices. PVCs are created between the end stations and the ATM network. This contrasts with the previous model, in which the PVCs are carried transparently across the network.

Figure 19-3: In the VoATM Signaling Translate Model, the ATM Network Interprets the Signaling from Both Non-ATM and ATM Network Devices
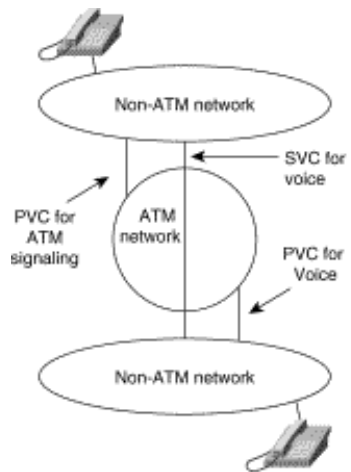
IT241804.EPS

A signaling request from an end station causes the ATM network to create an SVC with the appropriate QoS to the desired end station. The creation of an SVC versus the prior establishment of PVCs is clearly more advantageous for three reasons:

- SVCs are more efficient users of bandwidth than PVCs.

- QoS for connections do not need to be constant, as with PVCs.

- The capability to switch calls within the network can lead to the elimination of the tandem private branch exchange (PBX) and potentially the edge PBX. (A PBX

is a digital or analog telephone switchboard located on the subscriber premises and used to connect private and public telephone networks.)

VoATM Addressing

ATM standards support both private and public addressing schemes. Both schemes involve addresses that are 20 bytes in length (shown in Figure 19-4).

Figure 19-4: ATM Supports a 20-Byte Addressing Format
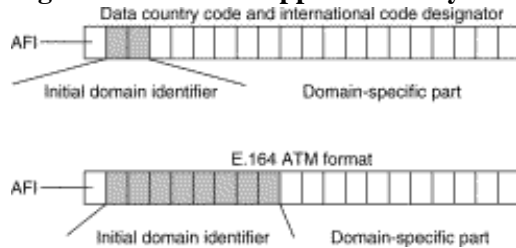
The *Authority and Format Identifier (AFI)* identifies the particular addressing format employed. Three identifiers are currently specified: data country code (DCC), international code designator (ICD), and E.164. Each is administered by a standards body. The second part of the address is the initial domain identifier (IDI). This address uniquely identifies the customer's network. The E.164 scheme has a longer IDI that corresponds to the 15-digit ISDN network number. The final portion, the domain-specific part (DSP), identifies logical groupings and ATM end stations.

In a transport model, you don't need to be aware of the underlying addressing used by the voice network. However, in the translate model, the capability to communicate from a non-ATM network device to an ATM network device implies a level of address mapping. Fortunately, ATM supports the E.164 addressing scheme, which is employed by telephone networks throughout the world.

VoATM Routing

ATM uses a *private network-to-network interface (PNNI), a hierarchical link-state routing protocol that is scalable for global usage. In addition to determining reachability and routing within an ATM network, it is also capable of call setup.*

A virtual circuit (VC) call request causes a connection with certain QoS requirements to be requested through the ATM network. The route through the network is determined by the source ATM switch based on what it determines is the best path through the network, based on the PNNI protocol and the QoS request. Each switch along the path is checked to determine whether it has the appropriate resources for the connection.

When the connection is established, voice traffic flows between end stations as if a leased line existed between the two. This specification spells out routing in private networks. Within carrier networks, the switch-to-switch protocol is B-ICI. Current research and development of integrated non-ATM and ATM routing will yield new capabilities to build translate-level voice and ATM networks.

VoATM and Delay

ATM has several mechanisms for controlling delay and delay variation. The QoS capabilities of ATM allow the specific request of constant bit rate traffic with bandwidth and delay variation guarantees. The use of VC queues allows each traffic stream to be treated uniquely. Priority can be given for the transmission of voice

traffic. The use of small, fixed-size cells reduces queuing delay and the delay variation associated with variable-sized packets.

Voice over Frame Relay

Voice over Frame Relay enables a network to carry live voice traffic (for example, telephone calls and faxes) over a Frame Relay network. Frame Relay is a common and inexpensive transport that is provided by most of the large telcos.

VoFR Signaling

Historically, Frame Relay call setup has been proprietary by vendor. This has meant that products from different vendors would not interoperate. Frame Relay Forum FRF.11 establishes a standard for call setup, coding types, and packet formats for VoFR, and it provides the basis for interoperability between vendors.

VoFR Addressing

Address mapping is handled through static tables, dialed digits mapped to specific PVCs. How voice is routed depends on which routing protocol is chosen to establish PVCs and the hardware used in the Frame Relay network. Routing can be based on bandwidth limits, hops, delay, or some combination, but most routing implementations are based on maximizing bandwidth utilization.

A full mesh of voice and data PVCs is used to minimize the number of network transit hops and to maximize the capability to establish different QoS. A network designed in this fashion minimizes delay and improves voice quality, but it represents the highest network cost.

Most Frame Relay providers charge based on the number of PVCs used. To reduce costs, both data and voice segments can be configured to use the same PVC, thereby reducing the number of PVCs required. In this design, the central site switch reroutes voice calls. This design has the potential of creating a transit hop when voice needs to go from one remote office to another remote office. However, it avoids the compression and decompression that occurs when using a tandem PBX.

A number of mechanisms can minimize delay and delay variation on a Frame Relay network. The presence of long data frames on a low-speed Frame Relay link can cause unacceptable delays for time-sensitive voice frames. To reduce this problem, some vendors implement smaller frame sizes to help reduce delay and delay variation. FRF.12 proposes an industry-standard approach to do this, so products from different vendors will be capable of interoperating and consumers will know what type of voice quality to expect.

Methods for prioritizing voice frames over data frames also help reduce delay and delay variation. This—and the use of smaller frame sizes—is vendor-specific implementations. To ensure voice quality, the committed information rate (CIR) on each PVC should be set to ensure that voice frames are not discarded. Future Frame Relay networks will provide SVC signaling for call setup and may also allow Frame Relay DTEs to request a QoS for a call. This will enhance VoFR quality in the future.

Voice over IP

As stated previously, voice over IP (VoIP) is an OSI Layer 3 solution rather than a Layer 2 solution. This feature allows VoIP to operate over Frame Relay and ATM networks autonomously. More importantly, VoIP operates over typical LANs to go all the way to the desktop. In this sense, VoIP is more of an application than a service, and VoIP protocols have evolved with this in mind.

VoIP protocols fall into two general categories: centralized and distributed. In general terms, centralized models follow a client/server architecture, while distributed models
are based on peer-to-peer interactions. All VoIP technologies use common media by transmitting voice information in RTP packets over IP. They also agree by supporting a wide variety of compression codecs. The difference lies in signaling and where call logic and call state are maintained, whether at the endpoints or at a central intelligent server. Both architectures have advantages and disadvantages. Distributed models tend to scale well and are more resilient (robust) because they lack a central point that could fail. Conversely, centralized call control models offer easier management and can support traditional supplementary services (such as conferencing) more easily, but they can have scaling limits based on the capacity of the central server. Hybrid and interworking models being developed also offer the best of both approaches.

Distributed VoIP call management schemes include the oldest architecture, H.323, and the newest, Session Initiation Protocol (SIP). Centralized call management methods include Media Gateway Control Protocol and proprietary protocols such as Skinny Station Protocol (from Cisco Systems). A brief overview of each of these protocols is provided next.

Voice Codec Overview

Voice coder/decoder (codec) technology has advanced rapidly over the last few years thanks to advancements in digital signal processor (DSP) architectures as well as research into human speech and recognition. New codecs do more than simply provide analog-to-digital conversion. They can apply sophisticated predictive patterns to analyze voice input and subsequently transmit voice using a minimum of bandwidth. Some examples of voice codecs and the bandwidth used are discussed in this section. In all cases, voice is carried in RTP packets over IP.

Simple pulse code modulated (PCM) voice is defined by ITU-T G.711. It allows two basic variations of 64-kbps PCM: Mu-law and A-law. The methods are similar in that they both use logarithmic compression to achieve 12 to 13 bits of linear PCM quality in 8 bits. However, they are different in relatively minor compression details (Mu-law has a slight advantage in low-level signal-to-noise ratio performance). Usage has historically been along country and regional boundaries, with North America using Mu-law and Europe using A-law modulation. Conversion from Mu-law to A-law is the responsibility of the Mu-law country. When troubleshooting PCM systems, a mismatch will result in terrible-sounding voice but will still be intelligible.

Another compression method often used is adaptive differential pulse code modulation (ADPCM). A commonly used instance of ADPCM, ITU-T G.726 encodes using 4-bit samples, giving a transmission rate of 32 kbps. Unlike PCM, the 4 bits do not directly encode the amplitude of speech, but encode the differences in amplitude as well as the rate of change of that amplitude, employing some very rudimentary linear prediction.

PCM and ADPCM are examples of *waveform codecs*, compression techniques that exploit redundant characteristics of the waveform itself. New compression techniques have been developed over the past 10 to 15 years that further exploit knowledge of the source characteristics of speech generation. These techniques employ signal-processing techniques that compress speech by sending only simplified parametric information about the original speech excitation and vocal tract shaping, requiring less bandwidth to transmit that information. These techniques can be grouped generally as "source" codecs and include variations such as linear predictive coding (LPC), code excited linear prediction (CELP), and multipulse, multilevel quantization (MP-MLQ).

There are also subcategories within these codec definitions. For example, code excited linear prediction (CELP) has been augmented by a low-delay version, predictably called *LD-CELP* (for low delay CELP). It has also been augmented by a more sophisticated vocal tract modeling technique using conjugate structure algebraic transformations. This results in a codec called CSA-CELP. The list goes on and on, but it is important for network designers to understand only the trade-offs of these approaches as they apply to network and application design.

Advanced predictive codecs rely on a mathematical model of the human vocal tract and, instead of sending compressed voice, send mathematical representations so that

voice can be generated at the receiving end. However, this required a great deal of research to get the bugs out. For example, some early predictive codecs did a good job of reproducing the developers' voices and were actively promoted—until it was discovered that they did
not reproduce female voices or Asian dialects very well. These codecs then had to be redesigned to include a broader range of human voice types and sounds.

The ITU has standardized the most popular voice coding standards for telephony and packet voice to include the following:

- G.711, which describes the 64-kbps PCM voice-coding technique outlined earlier. G.711-encoded voice is already in the correct format for digital voice delivery in the public phone network or through PBXs.

- G.726, which describes ADPCM coding at 40, 32, 24, and 16 kbps. ADPCM voice may also be interchanged between packet voice and public phone or PBX networks, provided that the latter has ADPCM capability.

- G.728, which describes a 16-kbps low-delay variation of CELP voice compression. CELP voice coding must be transcoded to a public telephony format for delivery to or through telephone networks.

- G.729, which describes CELP compression that enables voice to be coded into 8-kbps streams. Two variations of this standard (G.729 and G.729 Annex A) differ largely in computational complexity, and both generally provide speech quality as good as that of 32-kbps ADPCM.

- G.723.1, which describes a compression technique that can be used for compressing speech or other audio signal components of multimedia service at a very low bit rate. As part of the overall H.324 family of standards, this coder has two bit rates associated with it: 5.3 and 6.3 kbps. The higher bit rate is based on MP-MLQ technology and has greater quality; the lower bit rate is based on CELP, gives good quality, and provides system designers with additional flexibility.

As codecs rely increasingly on subjectively tuned compression techniques, standard objective quality measures such as total harmonic distortion and signal-to-noise ratios have less correlation with perceived codec quality. A common benchmark for quantifying the performance of the speech codec is the mean opinion score (MOS). Because voice quality and sound in general are subjective to the listener, it is important to get a wide range of listeners and sample material. MOS tests are given to a group of listeners who give each sample of speech material a rating of 1 (bad) to 5 (excellent). The scores are then averaged to get the mean opinion score. MOS testing is also used to compare how well a particular codec works under varying circumstances, including differing background noise levels, multiple encodes and decodes, and so on. This data can then be used to compare against other codecs.

MOS scoring for several ITU-T codecs is illustrated in Table 19-1. This table shows the relationship between several low bit rate codecs and standard PCM.

Table 19-1: Relative Processing Complexity and Mean Opinion Scores of Popular Voice Codecs

Compression Method	Bit Rate (kbps)	Processing [1] (MIPS)	Framing Size	MOS Score
G.711 PCM	64	0.34	0.125	4.1
G.726 ADPCM	32	14	0.125	3.85

G.728 LD-CELP	16	33	0.625	3.61
G.729 CS-ACELP	8	20	10	3.92
G.729 x2 Encodings	8	20	10	3.27
G.729 x3 Encodings	8	20	10	2.68
G.729a CS-ACELP	8	10.5	10	3.7
G.723.1 MPMLQ	6.3	16	30	3.9
G.723.1 ACELP	5.3	16	30	3.65

[1] MIP processing power given for Texas Instruments 54x DSPs

This table provides information useful in comparing various popular voice codec implementations. The relative bandwidth as well as processing complexity (in millions of instructions per second [MIPS]) is useful in understanding the trade-offs associated with various codecs. In general, higher mean opinion scores are associated with more complex codecs or more bandwidth.

VoIP Network Design Constraints

After voice has been compressed and converted to data, the next step is to put it into a Real Time Protocol (RTP) stream for transmission across an IP network. Network designers must consider both bandwidth and delay when implementing VoIP. Bandwidth requirements are critical and are determined not only by the codec selected, but also by the overhead added by IP headers and other factors. Bandwidth is especially critical across expensive WAN links. Delay is affected by propagation delay (speed of light constraints), serial delay (typically caused by buffering within devices in transit), and packetization delay.

Network Bandwidth Requirements

The bandwidth of a voice conversation over IP is affected by a variety of factors. First, as described previously, the codec employed for the conversation can vary widely from as little as 3 to 4 kbps to as much as 64 kbps. Layer 3 (IP) and Layer 2 (Ethernet) headers add additional overhead. Voice packets are typically very small and often contain no more than 20 bytes of information, so it is obvious that overhead can quickly overwhelm the bandwidth requirements.

Systems designers have several tools to help reduce the problem. First, voice activity detection (VAD) is used at the source to regulate the flow of packets by stopping transmission if the analog voice level falls below a threshold. This has the net result of reducing the bandwidth requirements by about half because most human conversations are silent at least half the time as the other person talks (unless there is a serious argument going on).

There are a couple of problems with this solution. First, switch on/switch off times must be carefully tuned to avoid clipping. Cisco solves this problem by continuously

sampling and coding, and then dropping the packet at the last moment if voice energy fails to exceed a certain minimum within the allotted time. In effect, a mostly empty voice packet is queued and prepared for transport, and will precede the speaker's first utterance, if necessary. The other problem created with VAD is the lack of noise at the receiver end. Human users of these early systems frequently complained that it sounded like they had been disconnected during the call because they no longer heard noise from the other end while they were talking. This proves that VAD is working but is evidently not user-friendly.

Cisco and other manufacturers have solved this problem by adding *comfort noise* to the receive end of the conversation. When a receiver is in buffer underflow condition —that is, it is not receiving packets—the system generates a low-level pink or white noise signal to convince listeners that they are still connected. More advanced systems actually sample the ambient background noise at the far end and reproduce it during periods of silence.

Another tool often used by network designers is to compress the RTP headers. A great deal of information in RTP headers is duplicated or redundant in a stream. Cisco routers can compress the RTP headers on a hop-by-hop basis, reducing required bandwidth by a significant amount.

The end result of these steps is illustrated in Table 19-2. This table shows the relative bandwidth requirements of various codec implementations, along with additional overhead associated with typical network transport layers.

Table 19-2: VoIP/Channel Bandwidth Consumption

Algorithm	Voice BW kbps	MOS	Codec Delay msec	Frame Size (Bytes)	Cisco Payload (Bytes)	Packets per Second	IP/ UDP/ RTP Header (Bytes)	CRTP Header (Bytes)	L2	Layer2 header (Bytes)	Total Bandwidth kbps no VAD	Total Bandwidth kbps VAD
G.729	8	3.9	15	10	20	50	40		Ether	14	29.6	14.8
G.729	8	3.9	15	10	20	50		2	Ether	14	14.4	7.2
G.729	8	3.9	15	10	20	50	40		PPP	6	26.4	13.2
G.729	8	3.9	15	10	20	50		2	PPP	6	11.2	5.6
G.729	8	3.9	15	10	20	50	40		FR	4	25.6	12.8
G.729	8	3.9	15	10	20	50		2	FR	4	10.4	5.2
G.729	8	3.9	15	10	20	50	40		ATM	2 cells	42.4	21.2
G.729	8	3.9	15	10	20	50		2	ATM	1 cell	21.2	10.6

Algorithm	Voice BW kbps	MOS	Codec Delay msec	Frame Size (Bytes)	Cisco Payload (Bytes)	Packets per Second	IP/ UDP/ RTP Header (Bytes)	CRTP Header (Bytes)	L2	Layer2 header (Bytes)	Total Bandwidth kbps no VAD	Total Bandwidth kbps VAD
G.711	64	4.1	1.5	160	160	50	40		Ether	14	85.6	42.8
G.711	64	4.1	1.5	160	160	50		2	Ether	14	70.4	35.2
G.711	64	4.1	1.5	160	160	50	40		PPP	6	82.4	41.2
G.711	64	4.1	1.5	160	160	50		2	PPP	6	67.2	33.6
G.711	64	4.1	1.5	160	160	50	40		FR	4	81.6	40.8
G.711	64	4.1	1.5	160	160	50		2	FR	4	66.4	33.2
G.711	64	4.1	1.5	160	160	50	40		ATM	5 cells	106.0	53.0
G.711	64	4.1	1.5	160	160	50		2	ATM	4 cells	84.8	42.4
G.729	8	3.9	15	10	30	33	40		PPP	6	20.3	10.1
G.729	8	3.9	15	10	30	33		2	PPP	6	10.1	5.1
G.729	8	3.9	15	10	30	33	40		FR	4	19.7	9.9
G.729	8	3.9	15	10	30	33		2	FR	4	9.6	4.8
G.729	8	3.9	15	10	30	33	40		ATM	2 cells	28.3	14.1
G.729	8	3.9	15	10	30	33		2	ATM	1 cell	14.1	7.1

G.723.1	6.3	3.9	37.5	30	30	26	40		PPP	6	16.0	8.0
G.723.1	6.3	3.9	37.5	30	30	26		2	PPP	6	8.0	4.0
G.723.1	6.3	3.9	37.5	30	30	26	40		FR	4	15.5	7.8
G.723.1	6.3	3.9	37.5	30	30	26		2	FR	4	7.6	3.8
G.723.1	6.3	3.9	37.5	30	30	26	40		ATM	2 cells	22.3	11.1
G.723.1	6.3	3.9	37.5	30	30	26		2	ATM	1 cell	11.1	5.6
G.723.1	5.3	3.65	37.5	30	30	22	40		PPP	6	13.4	6.7
G.723.1	5.3	3.65	37.5	30	30	22		2	PPP	6	6.7	3.4
G.723.1	5.3	3.65	37.5	30	30	22	40		FR	4	13.1	6.5
G.723.1	5.3	3.65	37.5	30	30	22		2	FR	4	6.4	3.2
G.723.1	5.3	3.65	37.5	30	30	22	40		ATM	2 cells	18.7	9.4
G.723.1	5.3	3.65	37.5	30	30	22		2	ATM	1 cell	9.4	4.7

Delay

Network designers planning to implement VoIP must work within a delay budget imposed by the quality of the system to the users. As a typical rule, total end-to-end delay must be kept to less than about 150 ms.

Propagation delay is determined by the medium used for transmission. The speed of light in a vacuum is 186,000 miles per second, and electrons travel about 100,000 miles per second in copper. A fiber network halfway around the world (13,000 miles) would theoretically induce a one-way delay of about 70 milliseconds. Although this delay is almost imperceptible to the human ear, propagation delays in conjunction with handling delays can cause noticeable speech degradation. Users who have talked over satellite telephony links experience a delay approaching 1 second in some cases, with typical delays of about 250 ms being tolerable. Delays greater than 250 ms begin to interfere with natural conversation flow, as speakers interrupt each other.

Handling delays can impact traditional circuit-switched phone networks, but they are a larger issue in packetized environments because of buffering of packets. Therefore, delay should be calculated to determine whether it stays below the threshold of 150 to 200 ms.

G.729 has an algorithmic delay of about 20 milliseconds because of look ahead. In typical Voice over IP products, the DSP generates a frame every 10 milliseconds. Two of these speech frames are then placed within one packet; the packet delay, therefore, is 20 milliseconds.

There are other causes of delay in a packet-based network: the time necessary to move the actual packet to the output queue, and queue delay. Cisco IOS software is quite good at moving and determining the destination of a packet. (This fact is mentioned because other packet-based solutions [PC-based and others] are not as good at determining packet destination and moving the actual packet to the output queue.) The actual queue delay of the output queue is another cause of delay. This factor should be kept to less than 10 milliseconds whenever possible by using whatever queuing methods are optimal for that network.

Table 19-3 shows that different codecs introduce different amounts of delay.

Table 19-3: Codec-Introduced Delay

Compression Method	Bit Rate (kbps)	Compression Delay (ms)
G.711 PCM	64	0.75
G.726 ADPCM	32	1
G.728 LD-CELP	16	3 to 5
G.729 CS-ACELP	8	10
G.729a CS-ACELP	8	10
G.723.1 MPMLQ	6.3	30
G.723.1 ACELP	5.3	30

In addition to steady state delay, discussed previously, VoIP applications are sensitive to variations in that delay. Unlike circuit-based networks, the end-to-end delay over a packet network can vary widely depending on network congestion. Short-term variations in delay are called *jitter, defined as the variation from when a packet was expected and when it actually is received. Voice devices have to compensate for jitter by setting up a playout buffer to play back voice in a smooth fashion and to avoid discontinuity in the voice stream. This adds to the overall system delay (and complexity). This receive buffer can be fixed at some value or, in the case of some advanced Cisco Systems devices, is adaptive.*

Note that jitter is the primary impediment to transmitting VoIP over the Internet. A typical VoIP call over the Internet would traverse many different carrier systems,

with widely varying latency and QoS management. As a result, VoIP over the public Internet results in poor quality and is typically discouraged by VoIP vendors. Nevertheless, many software applications exist to provide free voice services over the Internet. The common characteristic of these Voice over Internet systems is very large receive buffers, which can add more than 1 second of delay to voice calls. Free voice is attractive, but to business users, the poor quality means that these systems are worthless. However, some residential users are finding them adequate—especially for bypassing international toll charges.

In the future, as Internet service providers enhance the QoS features of their networks, Voice over Internet solutions will become more popular. In fact, many analysts predict that voice will eventually become free, as a bundled service with Internet access.

Quality of Service for VoIP

As seen previously, the quality of voice is greatly affected by latency and jitter in a packet network. Therefore, it is important for network designers to consider implementation of QoS policies on the network. In addition to protecting voice from data, this has the added benefit of protecting critical data applications from bandwidth starvation because of oversubscription of voice calls.

The elements of good QoS design include provisions for managing packet loss, delay, jitter, and bandwidth efficiency. Tools used to accomplish these goals are defined here:

- **Policing**—Provides simple limiting of packet rate, often by simply dropping packets that exceed thresholds to match capacities between different network elements. Policing can be performed on either input or output of a device. Examples include random early detection (RED) and WRED (weighted RED). These techniques help identify which packets are good candidates to drop, if necessary.

- **Traffic shaping**—Provides the capability to buffer and smooth traffic flows into and out of devices based on packet rate. Unlike policing, however, traffic shaping tries to avoid dropping packets, but it tends to add latency and jitter as they are buffered for later transmission.

- **Call admission control**—Provides the capability to reject requests for network bandwidth from applications. In the case of VoIP, an example might be the use of Resource Reservation Protocol (RSVP) to reserve bandwidth prior to completion of a call. Similarly, an H.323 gatekeeper might be used in signaling to manage a portion of available bandwidth on a per-call basis.

- **Queuing/scheduling**—These are used with buffering to determine the priority of packets to be transmitted. Separate queues for voice and data, for example, allow delay-sensitive voice packets to slip ahead of data packets. Examples useful for VoIP include weighted fair queuing and IP RTP priority queuing, among others.

- **Tagging/marking**—Includes various techniques to identify packets for special handling. In the case of VoIP packets, for example, the packets can be identified by RTP format, IP precedence bits (ToS bits), and so on. Tagging is also critical to preserve QoS across network boundaries. For example, tag switching preserves IP tagging across an ATM network, allowing VoIP to traverse an ATM network.

- **Fragmentation**—Refers to the capability of some network devices to subdivide large packets into smaller ones before traversing a narrow bandwidth link. This is critical to prevent voice packets from getting "frozen out" while waiting for a large data packet to go through. Fragmentation allows the smaller voice packets to be inserted within gaps in the larger packet. The large packet is subsequently reassembled by a router on the other end of the link so that the data application is unaffected.

H.323 Overview

H.323 is a derivative of the H.320 videoconferencing standard, but it assumes LAN connectivity rather than ISDN between conferencing components. As such, QoS is not assumed and is not implicitly supported. When used to support a VoIP application, the calls are treated as audio-only videoconferences.

Standards-based videoconferencing is generally governed by the International Telecommunications Union (ITU) "H-series" recommendations, which include H.320 (ISDN protocol), H.323 (LAN protocol), and H.324 (POTS protocol). These standards specify the manner in which real-time audio, video, and data communications takes place over various communications topologies. Standards compliance promotes common capabilities and interoperability between networked multimedia building blocks that may be provided by multiple vendors.

The H.323 standard was ratified in 1996 and consists of the following component standards:

- **H.225**—Specifies messages for call control, including signaling, registration and admissions, and packetization/synchronization of media streams.

- **H.245**—Specifies messages for opening and closing channels for media streams and other commands, requests and indications.

- **H.261**—Video codec for audiovisual services at P¥64 kbps.

- **H.263**—Specifies a new video codec for video POTS.

- **G.711**—Audio codec, 3.1 kHz at 48, 56, and 64 kbps (normal telephony).

- **G.722**—Audio codec, 7 kHz at 48, 56, and 64 kbps; ratified.

- **G.728**—Audio codec, 3.1 kHz at 16 kbps.

- **G.723**—Audio codec, for 5.3 and 6.3 kbps modes.

- **G.729**—Audio codec (G.729a is a reduced complexity variant).

Following are H.323 device descriptions:

- **Terminal**—An H.323 terminal is an endpoint on the local-area network that provides for real-time, two-way communications with another H.323 terminal, gateway, or multipoint control unit. This communication consists of control, indications, audio, moving color video pictures, and data between the two terminals. A terminal may provide speech only, speech and data, speech and video, or speech, data, and video.

- **Gateway**—An H.323 gateway (GW) is an endpoint on the local-area network that provides for real-time, two-way communications between H.323 terminals on the LAN and other ITU terminals on a wide-area network, or to another H.323 gateway. Other ITU terminals include those complying with recommendations H.310 (H.320 on B-ISDN), H.320 (ISDN), H.321 (ATM), H.322 (GQOS-LAN), H.324 (GSTN), H.324M (mobile), and V.70 (DSVD).

- **Proxy**—The proxy is a special type of gateway that, in effect, relays H.323 to another H.323 session. The Cisco proxy is a key piece of the conferencing infrastructure that can provide QoS, traffic shaping, and policy management for H.323 traffic.

- **Gatekeeper**—The gatekeeper, which is optional in an H.323 system, provides call control services to the H.323 endpoints. More than one gatekeeper may be present and they can communicate with each other in an unspecified fashion. The gatekeeper is logically separate from the endpoints, but its physical implementation may coexist with a terminal, MCU, gateway, MC, or other non-H.323 LAN device.

- **Multipoint control unit**—The multipoint control unit (MCU) is an endpoint on the local-area network that provides the capability for three or more terminals and gateways to participate in a multipoint conference. It may also connect two terminals in a point-to-point conference, which may later develop into a multipoint conference. The MCU generally operates in the fashion of an H.231 MCU, but an audio processor is not mandatory. The MCU consists of two parts: a mandatory multipoint controller and optional multipoint processors. In the simplest case, an MCU may consist of only an MC with no MPs.

- **Multipoint controller**—The multipoint controller (MC) is an H.323 entity on the local-area network that provides for the control of three or more terminals participating in a multipoint conference. It may also connect two terminals in a point-to-point conference, which may later develop into a multipoint conference. The MC provides for capability negotiation with all terminals to achieve common levels of communications. It also may control conference resources, such as who is multicasting video. The MC does not perform mixing or switching of audio, video, and data.

- **Multipoint processor**—The multipoint processor (MP) is an H.323 entity on the local-area network that provides for the centralized processing of audio, video, and data streams in a multipoint conference. The MP provides for the mixing, switching, or other processing of media streams under the control of the MC. The MP may process a single media stream or multiple media streams, depending on the type of conference supported.

- **Point-to-point conference**—A point-to-point conference is a conference between two terminals. It may be either directly between two H.323 terminals or between an H.323 terminal and an SCN terminal via a gateway. It is a call between two terminals.

- **Switched-circuit network (SCN)**—A public or private switched telecom-munications network such as the GSTN, N-ISDN, or B-ISDN.
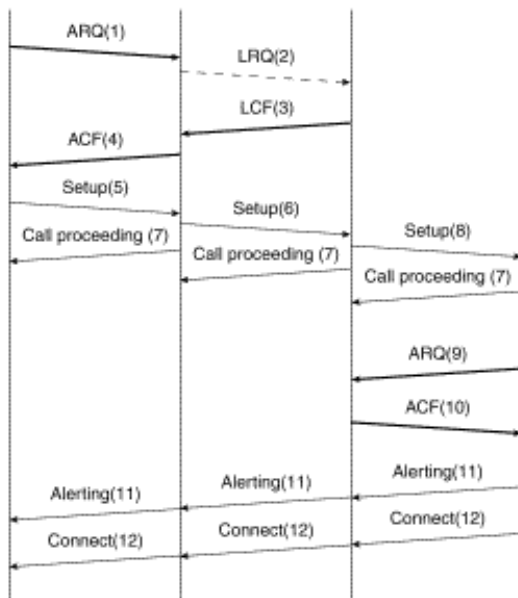
H.323 provides for fairly intelligent endpoints, which are responsible for maintaining their own call state. In its simplest form, H.323 is a peer-to-peer signaling system. Endpoints can call each other directly using the procedures provided by the standards if they know each other's IP address. Initial call setup signaling messages follow the traditional ISDN Q.931 model, using ASN.1-formatted information packets over TCP. As such, the signaling protocol relies on TCP retransmissions for QoS. After the call setup phase, the two endpoints do a capabilities exchange to negotiate which of several standard audio codecs to use, and finally they elect RTP port numbers to use for the voice media itself. Note that because RTP port numbers are assigned dynamically by the endpoints within a wide range, there are some difficulties operating through firewalls unless they maintain the call setup process itself.

H.323 Call Flow and Protocol Interworking

The provision of the communication is made in the steps shown in Figure 19-5.

Figure 19-5: Call Flow Between H.323 Devices

```
ct841905.eps
87050013
09/25/00
steve gifford
```

As can be seen from Figure 19-5, H.323 is designed to be robust and flexible, but at the cost of less efficiency.

General MGCP Overview

Media Gateway Control Protocol (MGCP) represents a relatively new set of client/server VoIP signaling protocols. These protocols have evolved in answer to the need for stateful, centralized management of relatively dumb endpoint devices. This capability greatly extends the utility of the system by making the VoIP system easier to design, configure, and manage because all major system changes occur at the server.

At the time of this writing, MGCP is an IETF draft. It may never be ratified as is by the IETF. Instead, a more advanced derivative protocol called MEGACO will probably be the ultimate solution. However, market demand has encouraged several vendors (including Cisco Systems) to announce support for MGCP in prestandard form. This has created the situation of a de facto standard with interoperability demonstrations among various vendors. This is generally good for the market because it

has resulted in products with real customer value from various vendors.

As with most standards, MGCP has a colorful history. Initially, a client/server protocol called Simple Gateway Control Protocol was proposed jointly by Bellcore (now Telcordia) and Cisco Systems. This was the first step toward a truly stateless client. During the same period, another client/server protocol, called Internet Protocol Device Control (IPDC), was being developed by Level 3 in conjunction with Cisco Systems and other vendors. IPDC was conceived as a more generic control system for various IP multimedia devices. As the two protocols matured in the standards committees, they eventually merged to form MGCP.

MGCP Concepts

As stated before, MGCP uses simple endpoints called media gateways (MGs). An intelligent media gateway controller (MGC) or call agent (CA) provides services. The endpoint provides user interactions and interfaces, while the MGC provides centralized call intelligence. A master/slave relationship is preserved at all times between the MGC and the MGs. In fact, all changes of state are forwarded to the MGC via a series of relatively simple messages. The MG can then execute simple actions based on commands from the MGC.

It is important to understand the stateless nature of the MG endpoints. They have no local call intelligence. For example, in the case of an FXS type interface supporting an analog telephone, when the user goes off-hook, the gateway notifies the MGC, which then instructs the MG to play the dial tone. When the user enters digits (DTMF) to dial a number, each digit is relayed to the MGC individually because the MG has no concept of a dial plan. It doesn't know when the user has dialed enough digits to complete a call. In a sense, the MG becomes a logical extension of the MGC. If any new services are introduced (such as call waiting), they need be introduced only into the MGC.

Typically, MGCP messages are sent over IP/UDP between the MG and the MGC. Any special telephony signaling interfaces (such as the D channel of a Primary Rate Interface) are simply forwarded directly to the MGC for processing rather than terminating them in the MG. This means that for typical applications, the data connection between the MG and the MGC is critical to keep calls up.

The media connection (voice path) itself is usually over IP/RTP, but direct VoATM and VoFrame Relay can also be used. (In fact, MGCP does not specify the media.) For security, MGCP uses IPSec to protect the signaling information.

MGCP Advantages

MGCP offers several advantages over typical H.323 implementations. Although MGCP has not been ratified as an official standard, enough vendors have demonstrated interoperability that it can be safely deployed by customers without fear of being locked in. It leverages existing IETF protocols (SDP, SAP, RTSP). Probably most importantly, the centralized call control model in MGCP allows for much more efficient service creation environments, including billing, call agents, messaging services, and so on. Depending on vendor implementation, the MGC can support standard computer telephony integration (CTI) interfaces such as Telephony Application Programming Interface (TAPI) used on PBXs.

MGCP Protocol Definitions

The MGCP model specifies the following:

- **Endpoints**—Specific trunk/port or service, such as an announcement server.

- **Connections**—The equivalent of a session. Connections offer several modes: send, receive, send/receive, inactive, loopback, and a continuity test.

- **Calls**—Groupings of connections.

- **Call agents**—The media gateway controller (MGC).

MGCP messages are composed from a short list of primitives:

- **NotificationRequest (RQNT)**—Instructs the gateway to watch for specific events.

- **Notify (NTFY)**—Informs the MGC when requested events occur.

- **CreateConnection (CRCX)**—Creates a connection to an endpoint inside the gateway.

- **ModifyConnection (MDCX)**—Changes the parameters associated with an established connection.

- **DeleteConnection**—Deletes an existing connection. Ack returns call statistics.

- **AuditEnpoint (AUEP)**—Audits an existing endpoint.

- **AuditConnection (AUCX)**—Audits an existing connection.

- **RestartInProgress (RSIP)**—Is a gateway notification to the MGC that an MG or an endpoint is restarting or stopping.

Of specific interest are the notification messages. The media gateway uses these messages to tell the MGC of a change of state. They typically involve signaling or events. Some examples of each are listed here:

- **Signals**—Ringing, distinctive ringing (0 to 7), ringback tone, dial tone, intercept tone, network congestion tone, busy tone, confirm tone, answer tone, call waiting tone, off-hook warning tone, pre-emption tone, continuity tone, continuity test, DTMF tones

- **Events**—Fax tones, modem tones, continuity tone, continuity detection (as a result of a continuity test), on-hook transition, off-hook transition, flash hook, receipt of DTMF digits

MGCP has a number of features that make it attractive for deployment of VoIP systems. First, messaging is UDP-based rather than TCP-based, which makes it more efficient. The centralized control model is subject to a single point of failure, so media gateways can be designed to revert to a standby MGC upon failure of the primary controller. This can result in the model being as reliable as any other call control model. MGCP scales well, typically depending only on the processing power of the MGC. When that becomes the limiting factor, the network can be subdivided into separate MGC domains. Therefore, an MGCP call control model can scale to millions of endpoints.

The protocol is also reliable, with an acknowledgment for each request consisting of one
of three options: success, transient error, and permanent error. Requests that are not acknowledged can be retried. MGCP also relies on DNS to resolve names to IP addresses. This means that the IP address can be abstracted to multiple nodes, or a single node can have multiple IP addresses. Again, all this adds to the flexibility of the
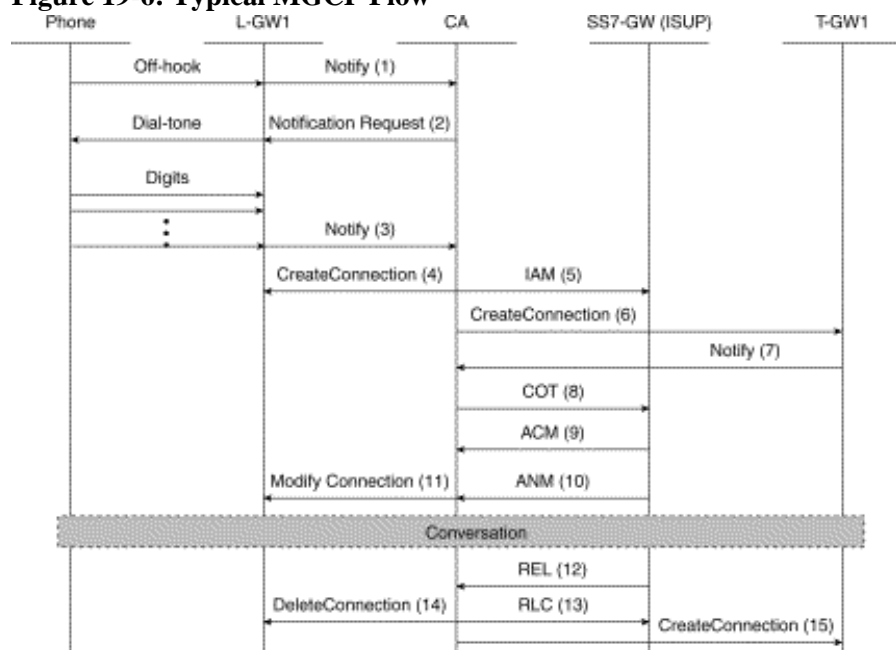
protocol.

Typical MGCP call flow is shown in Figure 19-6.

General SIP Tutorial

Session Initiation Protocol (SIP) is a new entry into the signaling arena, with a peer-to-peer architecture much like H.323. However, unlike H.323, SIP is an Internet-type protocol in philosophy and intent. It is described in RFC 2543, which was developed with the IETF MMUSIC Working Group in September 1999. Many technologists regard SIP as a competitor to H.323 and complementary to client/server protocols such as MGCP. As such, it will probably see deployment in mixed environments composed of combinations of SIP end points along with MGCP devices.

Figure 19-6: Typical MGCP Flow

steve gifford

SIP depends on relatively intelligent endpoints, which require little or no interaction with servers. Each endpoint manages its own signaling, both to the user and to other endpoints. Fundamentally, the SIP protocol provides session control, while MGCP provides device control. This provides SIP with a number of advantages. First, the simple message structure provides for call setup in fewer steps than H.323 so that performance is better than H.323 using similar processing hardware. SIP is also more scalable than H.323 because it is inherently a distributed and stateless call model. Perhaps the key difference (and advantage) of SIP is the fact that it is truly an Internet-model protocol from inception. It uses simple ASCII messaging (instead of ASN.1) based on HTTP/1.1. This means that SIP messaging is easy to decode and troubleshoot—but more importantly, it means that web-type applications can support SIP services with minimal changes. In fact, SIP fully supports URL (with DNS) naming in addition to standard E.164 North American Numbering Plan addressing. That means that in a SIP model, a user's e-mail address and phone address can be the same. It also means that the session is abstracted so that very different endpoints can communicate with each other.

SIP is modeled to support some or all of five facets of establishing and terminating multimedia communications. Each of these facets can be discovered or negotiated in a SIP session between two endpoints.

- User location

- User capabilities

- User availability

- Call setup

- Call handling

Although SIP is philosophically a peer-to-peer protocol, it is made up of logical clients and servers, often collocated within an endpoint. For example, a typical SIP client may be an IP phone, PC, or PDA; it contains both a user agent client (UAC) to originate SIP requests and a user agent server (UAS) to terminate SIP requests. Also supported are SIP proxy servers, SIP redirect servers (RS), registrars, and location servers. These servers are all optional but also very valuable in actual SIP implementations.

SIP servers are defined here:

- **Proxy server**—Acts as a server and client; initiates SIP requests on behalf of a UAC.

- **Redirect server (RS)**—Receives a SIP request, maps the destination to one or more addresses, and responds with those addresses.

- **Registrar**—Accepts requests for the registration of a current location from UACs. Typically is colocated with a redirect server.

- **Location server**—Provides information about a callee's possible locations, typically contacted by a redirect server. A location server/service may co-exist with a SIP redirect server.

SIP Messages

SIP messages consist of a simple vocabulary of requests and responses. Requests are called *methods* and include these:

- **REGISTER**—Registers current location with the server.

- **INVITE**—Is sent by the caller to initiate a call.

- **ACK**—Is sent by the caller to acknowledge acceptance of a call by the callee. This message is not responded to.

- **BYE**—Is sent by either side to end a call.

- **CANCEL**—Is sent to end a call not yet connected.

- **OPTIONS**—Is sent to query capabilities.

SIP Addressing

As mentioned previously, SIP addressing is modeled after mailto URLs. For example, a typical SIP address might look like:

sip: "einstein" aeinstein@smartguy.com; transport=udp

However, standard E.164 addressing can also be supported by embedding it in the same URL format, like this:

+14085553426@smartguy.com; user=phone

The address structure also indicates parameters such as transport type and multicast address.

SIP Call Flow

As seen in Figure 19-7, call setup with SIP is much simpler than H.323, even with a proxy server involved. Without the proxy server, the endpoints must know each other. However, call setup proceeds from a simple INVITE message directly from one endpoint to the other.

Figure 19-7: Call Flow for Session Initiation Protocol (SIP)
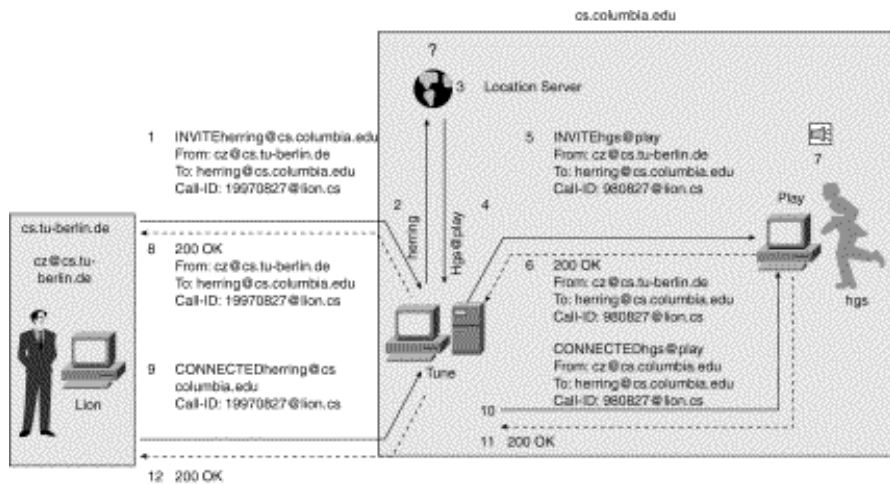
```
ct841907.eps
87050013
09/25/00
steve gifford
```

Comparison and Contrast of the Various VoIP Signaling Alternatives

The various signaling alternatives each offer advantages and disadvantages for system designers. A few highlights are presented here.

First, regarding MGCP and H.323, the scope of the protocols is different. MGCP is a simple device-control protocol, while H.323 is a full-featured multimedia conferencing protocol. H.323 is currently approved up to version 3, while MGCP has not been and may never be fully ratified; it is merely a de facto standard adopted by some manufacturers. As such, MGCP interoperability has been demonstrated, but not industry-wide. Likewise, the complexity of H.323 has inhibited interoperability as well.

MGCP can set up a call in as few as two round-trips, while H.323 typically requires seven or eight round-trips. (Note: H.323v2 provides for a fast start process to set up some calls in only two round-trips, but this is not widely implemented.) Call control is little more than device control for MGCP, while H.323 derives call flow from Q.931 ISDN signaling as a media control protocol. This control information is transmitted over UDP for MGCP, and over TCP for H.323.

SIP and H.323 are more direct competitors. They are both peer-to-peer, full-featured multimedia protocols. SIP is an IETF RFC, while H.323v3 has been approved by the ITU. Interoperability of both protocols has been demonstrated. SIP is more efficient than H.323, allowing some call setups in as little as a single round-trip. In addition, SIP uses existing Internet-type protocols, while H.323 continues to evolve new elements to fit into the Q.931 ISDN model.

Comparison of SIP to MGCP is similar to the comparison of H.323 to MGCP, in that SIP (like H.323) is a media-control protocol and MGCP is a device-control protocol. The same differences emerge as before between client/server and peer-to-peer. The fundamental difference is that peer-to-peer protocols such as H.323 and SIP tend to scale more gracefully, but client/server protocols such as MGCP are easier to design and maintain.

Evolution of Solutions for Voice over Data

The first products to integrate voice and data were targeted at eliminating long-distance telephone toll charges by providing tie lines between PBXs over a WAN infrastructure. These products were typically integrated into a router or another data device and provided simple point-to-point tie line service using simple analog trunk ports. As the products matured, more interface types were supported, including digital interfaces, E&M, and other types.

Later, as capabilities improved, support for analog telephone sets was introduced. This application was initially targeted at off-premises extensions from the PBXs using Private Line Automatic Ringdown (PLAR) circuits, but later DTMF detection was added within these gateway devices along with support for basic dial plans. Ultimately, this resulted in the capability of the WAN network devices to provide not only transport, but also tandem switching for the attached PBXs.

Over time, enterprise-wide call logic began to migrate toward the WAN data network elements. Each individual PBX at the edge of the WAN cloud needed only to forward intersite calls into the WAN gateways, without regard for further detailed trunk route calculations. Dial plans provisioned in data gateways such as Cisco Systems-integrated voice/routers were sufficient to manage trunking between many sites.

This model worked very well, especially for smaller networks of 10 or fewer sites. However, as installations grew increasingly larger with greater numbers of sites, it became difficult to administer. Every time a new site was added or the dial plan was otherwise changed, network engineers would need to manually log in to every router in the network to make corresponding dial plan changes. This process with unwieldy and error-prone. Ultimately, vendors began introducing tools that made this job easier. For example, the Cisco Voice Manager (CVM) product provides a GUI interface for dial plan configuration and management, and allows network engineers to manage hundreds of voice gateways.

Again, these solutions were sufficient for many applications, but scaling again became an issue at even larger system sizes with many hundreds to thousands of nodes. As large enterprises and service providers began to evaluate the technology, they discovered
scaling issues in two general areas: connection admission control (CAC) and dial plan centralization.

Connection admission control became more important as voice traffic grew. It became obvious that although a gateway could see another gateway across a logical flat mesh network, it was not always possible to complete a call. A method was needed for some central intelligence to act as traffic cop and to regulate the number of calls between critical nodes. Calls exceeding the defined number would be dropped or rerouted as necessary.

Dial plans also became too large to administer on small network elements. The flat mesh topology essentially made it necessary to store dial plan information about all sites in each node. Memory and processor limitations soon became the limiting factor to further growth.

The solution to both of these problems was the introduction of centralized call control. In the case of Voice over Frame Relay and Voice over ATM, virtual switch controller-type systems were introduced to centralize the call logic and intelligence. Likewise, for VoIP, the H.323 gatekeeper function was used to provide this

centralized control function. In the case of Cisco Systems, for example, the Multimedia Conferencing Manager (MCM) H.323 gatekeeper application was deployed to support voice networks as well as the videoconferencing networks for which it was developed.

Note that centralized call control logic does not mean centralization of voice paths. Only the dial plan administration and call control are centralized. The actual switching of voice packets still occurs in the data network elements as it always has, so the inherent economies and efficiency of packet voice solutions remain intact.

The Future: Telephony Applications

As integrated voice/data solutions continue to mature, a new wave of applications has emerged from various vendors. Instead of providing simple transport and switching functions for PBXs, packet voice solutions can now begin to replace those PBXs with an end-to-end solution. This means that packet voice technologies are no longer a service provided by the network, but they become an application running on the network. The distinction is critical in terms of how these products are marketed and administered. These products can be categorized by architecture and consist of the following general types:

- **Un-PBX**—In this architecture, a PC-based server contains both trunk gateway ports and analog telephone ports. Typically, special software and drivers running on an NT operating system provide all standard key system functions to the analog telephones. Supplementary functions such as **hold** and **transfer** are activated via hookflash and * commands. The systems typically scale up to as many as 48 telephones. Note that there is no redundancy, but the overall cost of the system can be much less than that of older key systems. Many products include integrated voicemail by saving digitized voice messages on the hard disk.

- **LAN-PBX**—This is a general category of products that are based on LAN telephony all the way to the desktop. Some products offer LAN telephony services through the use of a software client on the user's PC, while others actually offer telephone instruments that plug into the LAN. Of the latter, products can be based on the MAC layer (Ethernet), ATM, or IP. Products at Layer 3 (those that are IP-based) offer greater flexibility and scaling because IP is a routable protocol. That means that these products can be used on different LAN segments. Products based on lower-layer protocols offer an attractive price point because client complexity is lower.

Over the long run, the greatest challenges facing LAN telephony are reliability and scalability. These issues must be addressed if voice/data integration is ever to replace the traditional PBX architecture. Products address these issues in a number of ways. For example, the Cisco Systems IP telephony solution provides for redundant call processing servers so that if one fails, the IP telephones switch to a backup unit. In addition, call control models that reduce server complexity provide for better scalability. In this case, the Cisco Systems products use a client/server call control model similar to MGCP, called Skinny Station Protocol. This allows a single server to manage thousands of telephone endpoints (telephones and gateway ports).

Incentives Toward Packet Telephony Applications

LAN-based telephony solutions offer attractive business models to consumers today. Typical "un-PBX" systems cost less than the key systems that they replace. Likewise, LAN-based PBX systems provide superior return on investment to traditional PBX systems. Although initial equipment costs are comparable, LAN PBXs typically cost much less than PBXs to install because they use the existing data infrastructure (Category 5 cabling) rather than separate voice wiring. Administration is also less burdensome because LAN and
server administrators can manage the system without the need for dedicated telephony technicians. Finally, toll-bypass savings are also a byproduct of the system because calls between offices stay on the data network from end to end. Over time, these savings add up to the point that a LAN-based telephony system can offer considerable savings over traditional PBXs.

This is not to say that PBX systems will disappear overnight. Instead, traditional PBX vendors are actively migrating the existing products to become packet-enabled. Starting with simple data trunk cards to provide toll-bypass capability, PBX vendors are adding H.323 VoIP cards to allow the PBXs to manage H.323 clients as well.

They see the PBX evolving into a voice server, much as the LAN PBX vendors are building from the ground up. Only time will tell which solution will be superior, but one thing is clear: Customers will have more choices than ever.

Perhaps the most compelling reason to consider IP telephony-type applications is the future integration of applications with voice. Over the years, a significant amount of work has gone into computer telephony integration (CTI) in traditional PBXs. These systems began to offer application programming interfaces such as Telephony API (TAPI), Telephony Services API (TSAPI), and Java Telephony API (JTAPI). This work has resulted in advanced call center functions, including screen pops for agents and active call routing between call centers.

However, technologists believe that this is only the beginning. Integrated voice/data applications will revolutionize the way people use these systems. For example, Unified Messaging enables users to access voicemail, e-mail, and fax from one common server, using whatever media they choose. A user can retrieve voicemails on a PC (as .wav files) or, conversely, can retrieve written messages from a telephone utilizing text-to-speech capability in the system.

Fundamental to all these examples is a rethinking about the way people access and use information. It will become possible for the receiver of a message to determine the media rather than the sender. In addition, integration with intelligent assistant-type software from various vendors will enable users to set up rules for management of all incoming calls. In the call center, complex business rules (for example, checking credit before accepting new orders) can be applied to all forms of incoming communications (voice, e-mail, and so on) uniformly. The final result will be not only cost savings, but also increased efficiency for organizations that can learn to leverage this technology.

Summary

This chapter has provided an overview of technologies and applications of integrated voice/data networking. Specific protocol and architectural definitions for voice over Frame Relay, voice over ATM, and voice over IP were provided. However, more importantly, emphasis was placed on the reasons why these technologies have become prevalent. These technologies support a range of applications with very real business benefits for users. These benefits include cost savings from applications such as toll bypass through total replacement of PBXs with VoIP technology. More importantly, new integrated applications can benefit from packet voice technologies.

Along with these technologies comes the pressure of deciding which one is appropriate for specific situations. The value of various solutions was reviewed, with Voice over ATM and Voice over Frame Relay shown as most appropriate for simple toll bypass and tandem switching; Voice over IP provides support for end-to-end voice applications to the desktop at the expense of greater complexity.

Review Questions

Q—*What are the three main packet voice technologies?*

A—Voice over Frame Relay, Voice over ATM, and Voice over IP are the three main packet voice technologies.

Q—*How are packet voice technologies used to provide toll bypass cost savings?*

A—Voice traffic between locations can be routed over a wide-area network with data instead of using long-distance carriers. Depending on distance and toll charges, cost savings can be substantial.

Q—*What are the primary voice-signaling protocols?*

A—These are H.323, Session Initiation Protocol (SIP), and Media Gateway Control Protocol (MGCP).

Q—*Describe how peer-to-peer voice signaling protocols are different from client/server protocols.*

A—Client/server signaling protocols depend upon a central call control entity to maintain the state of the endpoints. This model makes it easier to support advanced call features. Peer-to-peer protocols utilize smarter endpoints and do not require a central call control entity, so they scale better.

For More Information

Books

Davidson. *Voice over IP Fundamentals*. Cisco Press.

Newton, Harry. *Newton's Telecom Dictionary*.

Dodd, Annabel Z. *The Essential Guide to Telecommunications*.

| HOME | CONTENTS | PREVIOUS | NEXT | GLOSSARY | FEEDBACK | SEARCH | HELP |