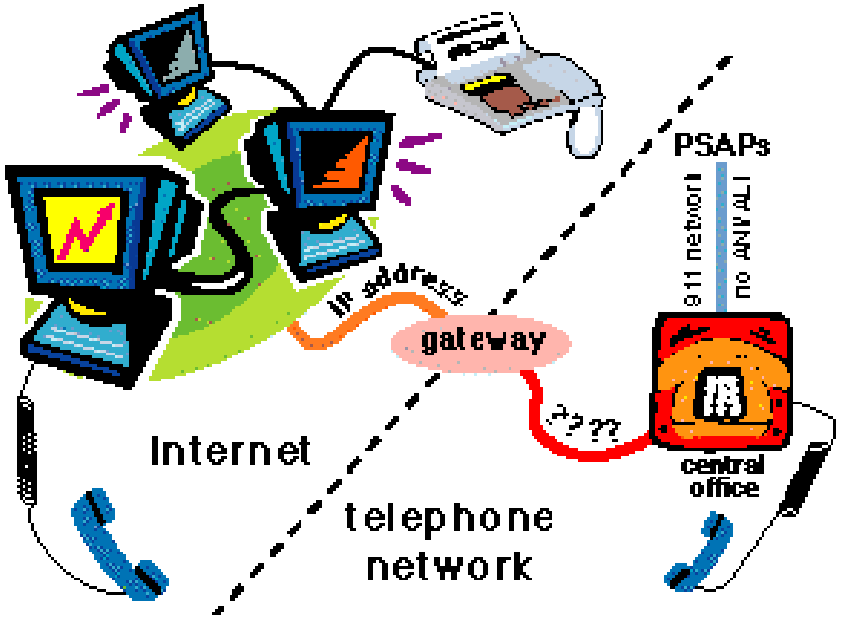
Overview of Voice over IP



February 2001 – University of Pennsylvania

Technical Report MS-CIS-01-31*

Princy C. Mehta
prmehta@seas.upenn.edu

Professor Sanjay Udani
udani@seas.upenn.edu

* This paper was written for an Independent Study course.

Table of Contents

Table of Figures

Acronyms and Definitions

Table 1 organizes an alphabetized listing of acronyms and their respective definitions used in this paper.

Table 1: List of Acronyms and Respective Definitions

Acronym	Definition
ACELP	Algebraic-Code-Excited Linear Prediction
AH	Authentication Header
ATM	Asynchronous Transfer Mode
CNG	Comfort Noise Generator
CS-ACELP	Conjugate-Structure, Algebraic-Code-Excited Linear Prediction
CSRC	Contributing Source Identifier
DCP	Device Control Protocol
DSP	Digital Signal Processor
DSSS	Direct Sequence Spread Spectrum
DTMF	Dual Tone Multi-Frequency
DTX	Discontinuous Transmission
ESP	Encapsulating Security Payload
EST	Eastern Standard Time
FEC	Forward Error Correction
FHSS	Frequency Hopping Spread Spectrum
GUI	Graphical User Interface
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IIS	Internet Integrated Service
IMTC	International Multimedia Telecommunications Consortium
IPDC	Internet Protocol Device Control
IPSec	Internet Protocol Security
IPvX	Internet Protocol version X
ISDN	Integrated Services Digital Network
ISM	Industry, Scientific, and Medical
ITU-T	International Telecommunication Union-Telecommunication
MAC	Media Access Control
MBE	Multi-Band Excitation
MCU	Multipoint Control Units
MELP	Mixed Excitation Linear Predictive
MGCP	Media Gateway Control Protocol
MMUSIC	Multiparty Multimedia Session Control
MOS	Mean Opinion Scores
MP-MLQ	Multi-Pulse-Maximum Likelihood Quantization
NIC	Network Interface Card
OFDM	Orthogonal Frequency Division Multiplexing
OS	Operating System
OSI/RM	Open Systems Interconnected Reference Model
PBX	Private Branch Exchange
PC	Personal Computer
PCF	Point Coordination Function
PCM	Pulse Code Modulation

Acronym	Definition
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RAM	Random Access Memory
RFC	Request For Comment
RSVP	Resource Reservation Protocol
RTCP	Real-Time Control Protocol
RTP	Real-Time Transport Protocol
SCP	Service Control Point
SGCP	Simple Gateway Control Protocol
SIP	Session Initiation Protocol
SS7	Signaling System 7
SSP	Service Switching Point
SSRC	Synchronization Source Identifier
STP	Service Transfer Point
TCP/IP	Transmission Control Protocol/Internet Protocol
UDP	User Datagram Protocol
UNII	Unlicensed National Information Infrastructure
VAD	Voice Activity Detector
VoIP	Voice over Internet Protocol
WAN	Wide Area Network
WEP	Wired Equivalent Privacy
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network

Introduction

The vast majority of information exchanged over the public telecommunication networks has been voice. The present voice communication networks, public telephone, and Integrated Services Digital Network (ISDN) networks utilize digital technology via circuit switching. Circuit switching establishes a dedicated path (circuit) between the source and destination. This environment provides fixed bandwidth and short and controlled latency (delay). It provides satisfactory quality service and does not require a complicated encoding algorithm. The capacity of the circuit, however, is not shared by other users, thereby hindering the system's overall efficiency.

In contrast, a packet-switched network such as the Internet switches data through a network by splitting data into packets containing destination identification that are sent and routed independently. It implements store-and-forward switching of discrete data units (packets), and implies statistical multiplexing. This is an ideal environment for non-voice data, where the performance of a best-effort delivery model in terms of throughput is more desirable than delivery of packets within bounded latency and jitter. Crudely sending voice data over such a network will lead to poor and even unacceptable quality.

What is required is a mechanism to transport voice over a packet-switched network: Voice over Internet Protocol (VoIP). A discussion on how VoIP is implemented will be presented. The goal of VoIP is to provide the efficiency of a packet-switched network while rivaling the quality of a circuit-switched network. The quality of VoIP does not yet match the quality of a circuit-switched telephone network, but there is an abundance of activity in developing protocols and speech encoders for the implementation of the high-quality voice service. One formidable problem is that the Internet was designed for data communications; consequently, packets suffer a long and variable delay that decreases voice quality. To overcome this problem, protocols are being developed to provide a certain share of network resources for each voice call through the network.

The paper will expound a complete picture describing the behind-the-scenes technology of VoIP, including the technology it comprises: signaling, encoding, transport, and gateway control. It will focus on audio codecs and how they impact voice quality. The trend in codec design appears to be towards encoding voice at progressively lower bit-rates, typically in the low- to mid-single digit kbps. The crux of the argument is, since most Internet users have at least a 28.8 kbps connection, is *this* effort really necessary, and more importantly, optimally focused? Would it not be more worthwhile to pursue matching the quality of Public Switched Telephone Network (PSTN) voice? Namely, make the codec robust enough to be relatively impervious to the random behavior of packet-switched networks, even if the bit-rate were higher, for example, in the tens of kbps.

Furthermore, emphasis will be made on using VoIP on a wireless network, specifically the IEEE 802.11x and Bluetooth standards. Wireless applications are inevitably the wave of the future; for VoIP to thrive, it is imperative that its operation not exhibit excess degradation in a wireless environment. The ensuing sections will examine these topics and present a qualitative discussion on current wireless VoIP applications. Moreover, qualitative results of VoIP experiments conducted over various networks will be furnished. Finally, a general discussion on VoIP, including examining the bit-rate versus quality issue, will be analyzed.

On the whole, many proprietary technologies for VoIP are available, and it is expected that these applications will expand as the technologies mature into certified standards – perhaps forming a single standard that is an amalgamation of current schemes. The Internet will also be widely used for facsimile calls and video-conferencing as standards evolve; however, these topics are outside the scope of this paper. Also outside the scope of this paper is transporting voice over other protocols, such as asynchronous transfer mode (ATM). Finally, as this is a technical paper, the economic impact of VoIP will not be discussed.

Implementation of Voice over IP

Voice over Internet Protocol (VoIP) is a means to talk over IP instead of solely over the PSTN. VoIP can be implemented in several ways, as shown in Figure 1. The first scenario depicts a voice call made from one PSTN telephone to another. This can call either be transmitted over traditional analog lines, or can be converted to IP, then back to the PSTN. This would be done to reduce cost by exploiting utilization optimization. The next scenario portrays a voice call made from a PSTN telephone to a voice application residing on a personal computer (PC). Finally, the third scenario illustrates a voice call initiated from the PC via its VoIP server, acting in a PSTN capacity, which is routed over the Internet to a telephone attached to an organization's call center, usually a Private Branch Exchange (PBX). Not shown in the diagram is a fourth scenario, which is simply a PC-to-PC call where the voice signal is transported via IP without accessing the PSTN.

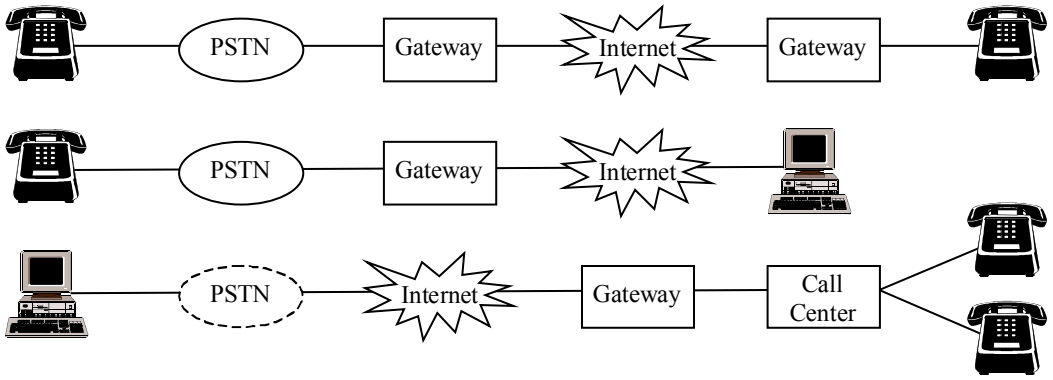


Figure 1: VoIP Applications

Before delving into the details of VoIP and the topics normally associated with it, such as vocoders and the slew of industry standards, an overview of the Transmission Control Protocol/Internet Protocol (TCP/IP) will be provided. VoIP interacts intimately with TCP/IP, be it for call setup, the actual conversation, or call teardown, so it is prudent to understand this concept.

Overview of TCP/IP

TCP/IP is defined as an industry standard suite of protocols that computers use to find, access, and communicate with each other over a transmission medium. In this context, a protocol is the set of standards and rules that a machine's hardware and software must follow in order to be recognized and understood by other computers. The protocol suite is implemented via a software package most commonly known as the TCP/IP stack, which breaks the job into a number of tasks. Each layer corresponds to a different facet of communication. The TCP/IP architecture consists of four "layers" performing certain functions: 1.) Application layer, 2.) Transport layer, 3.) Internet layer, and 4.) Physical (network interface) layer. Each layer contains protocols, which will be briefly summarized here. A full-scale description of each layer and its underlying functionality is well beyond the scope of this article, however, is more substantially covered in [Keshav] and [Peterson and Davie]. Figure 2 describes the TCP/IP reference model and how it maps to the Open Systems Interconnected reference model (OSI/RM), the standard that all other protocols follow.

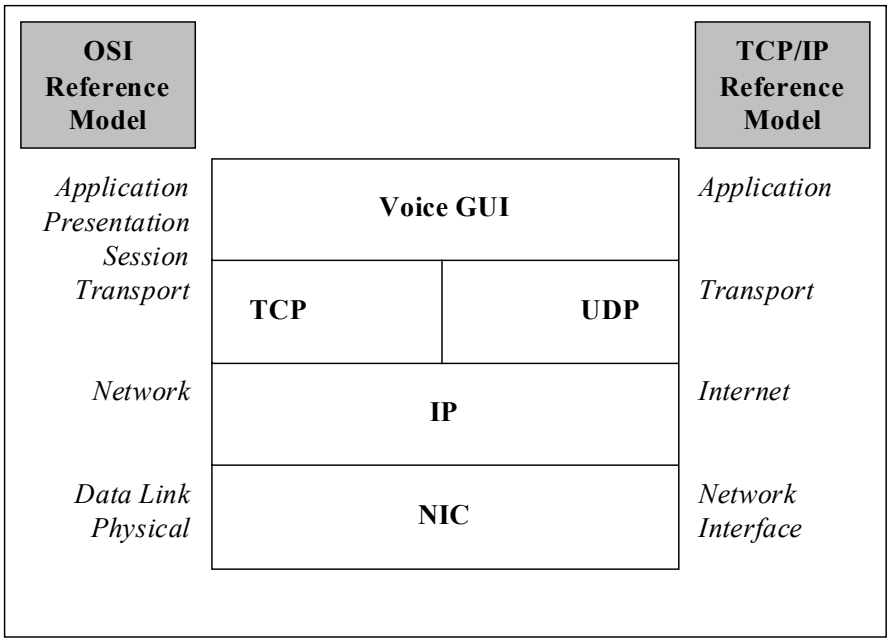


Figure 2: OSI and TCP/IP Reference Models

When transmitting voice over the Internet, the data being sent starts from the application layer (typically via the GUI), traverses down the "stack" to the network interface card (NIC) with each layer adding header and trailer frames. It is then sent to the receiver, where the data goes up the "stack" in reverse order, this time stripping the appropriate header and trailer frames.

Packetization

Given the nature of addition and removal of header/trailer data in each packet, there is an innate packetization and processing delay. For a latency-sensitive application such as voice, it is imperative that this delay be minimized. Conversely, it is desired to efficiently transmit packets over the Internet to fully utilize the bandwidth. There is clearly a trade-off in the desire to maintain small packets to minimize delay and the desire to send a large payload to minimize the overhead due to header content, thus maximizing payload efficiency. Packets are efficient for data transfer, but are not so attractive for real-time services such as voice. That is where the selection of an optimum voice coder is necessary, which is discussed later. Since this paper is focusing on voice over IP as opposed to voice over ATM, techniques on minimizing latency in the IP environment will be discussed.

The smallest packetization delay obviously occurs if only one sample of voice was sent at a time. However, that would cause a great number more of packets to be sent which would strain packetization processing as the single sample traversed the TCP/IP stack. If voice were digitized at 8,000 samples/s, where each sample is 1 byte, then a 500-byte packet would take 62.5 ms to fill. For a desired delay of no more than 100 ms, it would mean that 62.5 percent of the delay budget is spent in packetization!

Each voice packet incurs an uncompressed 40-byte header that comprises 20 bytes for the IP header, 8 bytes for the UDP header, and 12 bytes for the Real-Time Transport Protocol (RTP) header. The IP header consists of several fields, including version, its length, type of service, flags, time to live, protocol, header checksum, and source and destination IP addresses. The UDP header contains 8 bytes of Protocol Control Information (source and destination ports, UDP length and checksum). Finally, the RTP packet is used on

top of UDP to allow the transport of isochronous data across a packet network, which introduces jitter and may send packets out of order.

Components of VoIP

The PSTN is the collection of all the switching and networking equipment that belongs to the carriers that are involved in providing telephone service. In this context, the PSTN is primarily the wireline telephone network and its access points to wireless networks, such as cellular. VoIP is being promoted to augment, if not eventually replace, the current PSTN infrastructure. As previously mentioned, the overall technology requirements of an IP telephony solution can be split into four categories: signaling, encoding, transport, and gateway control. These are succinctly described below but are further discussed in the next four sections, respectively.

The purpose of the *signaling* protocol is to create and manage connections between endpoints, as well as to create and manage calls. Next, when the conversation commences, the analog signal produced by the microphone from the human voice needs to be *encoded* in a digital format suitable for transmission across an IP network. The IP network itself must then ensure that the real-time conversation is *transported* across the available media in a manner that produces acceptable voice quality. Finally, it may be necessary for the IP telephony system to be converted by a *gateway* to another format – either for interoperation with a different IP-based multimedia scheme or because the call is being placed onto the PSTN.

Signaling

Once a user dials a telephone number (or clicks a name hyperlinked to a telephone number), signaling is required to determine the status of the called party – available or busy – and to establish the call. There are multiple and complex levels of signaling that must take place in order to initiate and complete a call; these complexities escalate when VoIP users in packet networks communicate with PSTN subscribers. Neither H.323 nor the Session Initiation Protocol (SIP) alone makes up a complete set of IP telephony protocols; these protocols are merely competing standards for signaling. Both of these schemes will be explicated herein. Moreover, a means to achieve PSTN services from VoIP, namely interacting with Signaling System 7 (SS7), will be briefly examined.

H.323

H.323 is a set of protocols for voice, video, and data conferencing over packet-based networks, such as the Internet. The current recommendation, version 4.0, was ratified by the International Telecommunication Union – Telecommunication (ITU-T). The H.323 protocol stack is designed to operate above the transport layer of the underlying network. Thereby, H.323 can be used on top of any packet-based network transport, such as TCP/UDP/IP, to provide real-time multimedia communication. H.323 specifies protocols for real-time point-to-point audio communication between two terminals on a packet-based network that does not provide a guaranteed quality of service. The scope of H.323, however, is much broader and encompasses networking multipoint conferencing among terminals that support not only audio but also video and data communications, however, a discussion on this is outside the scope of the paper.

The following features can summarize the H.323 specification:

- Point-to-point and multipoint conferencing support
- Networking interoperability
- Heterogeneous client capabilities
- Audio and video codecs

- Management and accounting support
- Security
- Supplementary services

Logical Entities

In a general H.323 implementation, three logical entities (components) are required: gateways, gatekeepers, and multipoint control units (MCU). Terminals, gateways, and MCUs are collectively known as endpoints. It is possible to establish an H.323-enabled network with just terminals, which are H.323 clients. For more than two endpoints, an MCU is required, which can be combined into a terminal, gateway, or gatekeeper. Along with the gateway and gatekeeper (explained below), these components are essential to provide greater practical usefulness of VoIP services.

A gateway is an optional component in an H.323-enabled network. A gateway, in the context of VoIP, is a router that performs protocol conversion between different types of voice applications. When communication is required between different types of networks, a gateway is required at the interface. Through the provision of gateways, it is possible for H.323 terminals to interoperate with other conferencing terminals. A gateway provides data format translation, control signaling translation, audio and video codec translation, and call setup and termination functionality on both sides of the network.

A gatekeeper is an optional, but very useful, component of an H.323-enabled network, because it provides central management and control services. When a gatekeeper exists, all endpoints (terminals, gateways, and MCUs) must be registered with it. Registered endpoints' control messages are routed through the gatekeeper. Gatekeepers provide the following services: 1.) Address translation, 2.) Admission and access control of endpoints, 3.) Bandwidth management, and 4.) Routing capability.

H.323 Stack

H.323 encompasses many specific protocols, including Q.931, H.225, H.245, and ASN.1. To produce call-signaling functions, the H.323 partially merges the H.225 and Q.931 specifications. H.245 defines a variety of procedures that facilitates exchanging capabilities, master-slave determination, and channel signaling. Finally, the ASN.1 specification is employed to define how data is formatted in order to ensure interoperability among H.323-compliant endpoints. Figure 3 depicts a block diagram of an overall H.323 stack implementation.

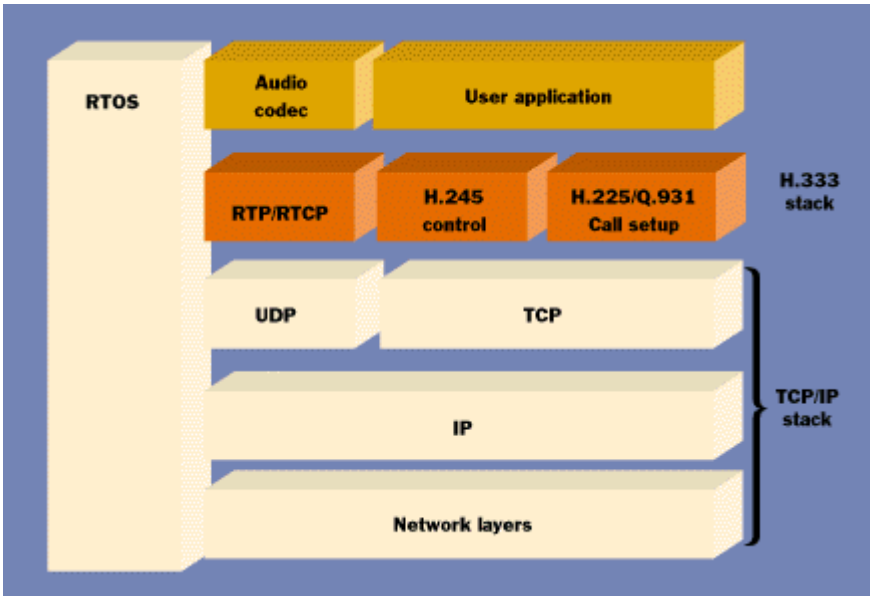


Figure 3: H.323 Stack Implementation

These protocols are executed in a strict order. First, setup occurs via Q.931 over TCP. Next, messages are passed between the caller ("User A") and callee ("User B") utilizing H.245, which is also over TCP. Finally, the RTP stream is sent over UDP, followed by the bi-directional Real-Time Control Protocol (RTCP). There are some conflicts and redundancies between H.245 and RTCP; however, those specifics are out of this paper's scope. Figure 4 underscores these messages to initiate an H.323 call between User A and User B.

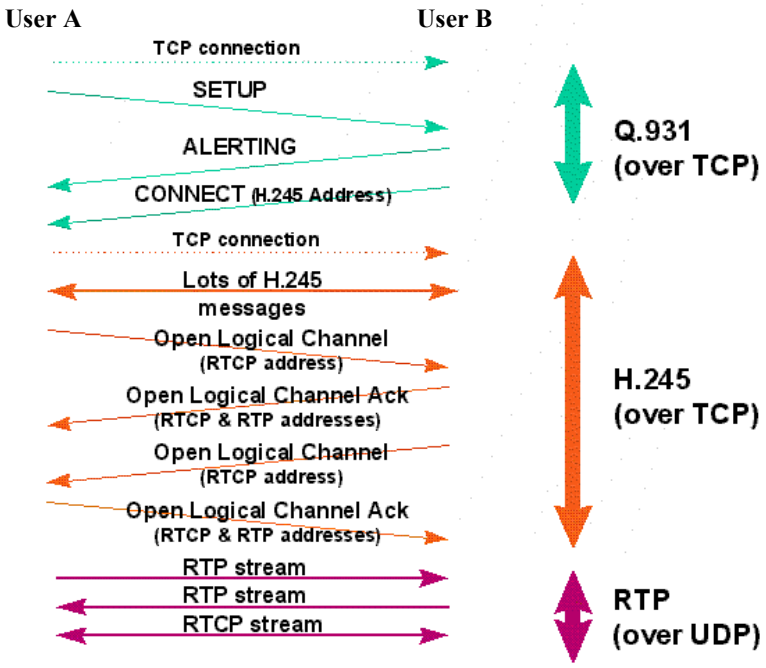


Figure 4: H.323 Call Setup

SIP

SIP is a signaling protocol for terminating phone calls over IP that is defined in RFC 2543 of the Multiparty Multimedia Session Control (MMUSIC) working group of the Internet Engineering Task Force (IETF). Unlike H.323, the traditional telephone protocol, however, SIP was designed specifically for the Internet. It not only exploits the manageability of IP, but is architecturally designed to make developing a telephony application relatively simple.

SIP is an application-layer control signaling protocol for creating, modifying, and terminating sessions with one or more participants. These sessions include Internet multimedia conferences, Internet telephone calls, and multimedia distribution. SIP can invite parties to both unicast and multicast sessions; the initiator does not have to be a member of the session to which it is inviting.

SIP Functionality

SIP can be utilized to initiate sessions and invite members to sessions that have been advertised by other means, such as via multicast protocols. The signaling protocol transparently supports name mapping and redirection services, allowing the implementation of intelligent network telephony subscriber services. These facilities also enable personal mobility [Handley, et al.]. In this context, personal mobility is the ability of end users to originate and receive calls and access subscribed telecommunication services on any terminal in any location. This mobility can be enhanced via wireless VoIP, which is described later in the paper.

SIP supports five facets of establishing and terminating multimedia communications:

- User location: determination of the end system to be used for communication.
- User capabilities: determination of the media and media parameters to be used.
- User availability: determination of the willingness of the called party to engage in communications.
- Call setup: "ringing", establishment of call parameters at both called and calling party.
- Call handling: including transfer and termination of calls.

SIP can also initiate multiparty calls using an MCU or fully-meshed interconnection instead of multicast. Gateways that connect PSTN parties can also use SIP to set up calls between them. The protocol is designed as part of the overall IETF multimedia data control architecture incorporating many protocols, such as Resource Reservation Protocol (RSVP) and RTP, for proper functionality and operation.

SIP can be used along with other call setup and signaling protocols where an end system that uses SIP exchanges to determine the appropriate end system address and protocol from a given address that is autonomous of the protocol employed. For example, SIP might be used to determine that the callee is reachable via the PSTN and indicate the phone number to be called, possibly suggesting the gateway to be used.

Initiating a SIP Call

SIP embarks on a four-step procedure to construct a VoIP call, from a signaling viewpoint. First, a caller locates the appropriate server, then sends a SIP request (usually "invite"). Typically, the request arrives at its destination, where the client accepts the call. Then the originating caller sends an acknowledgment back to the recipient. Likewise, the station that initiates the call also sends the acknowledgment. Figure 5 pictorially summarizes this procedure.

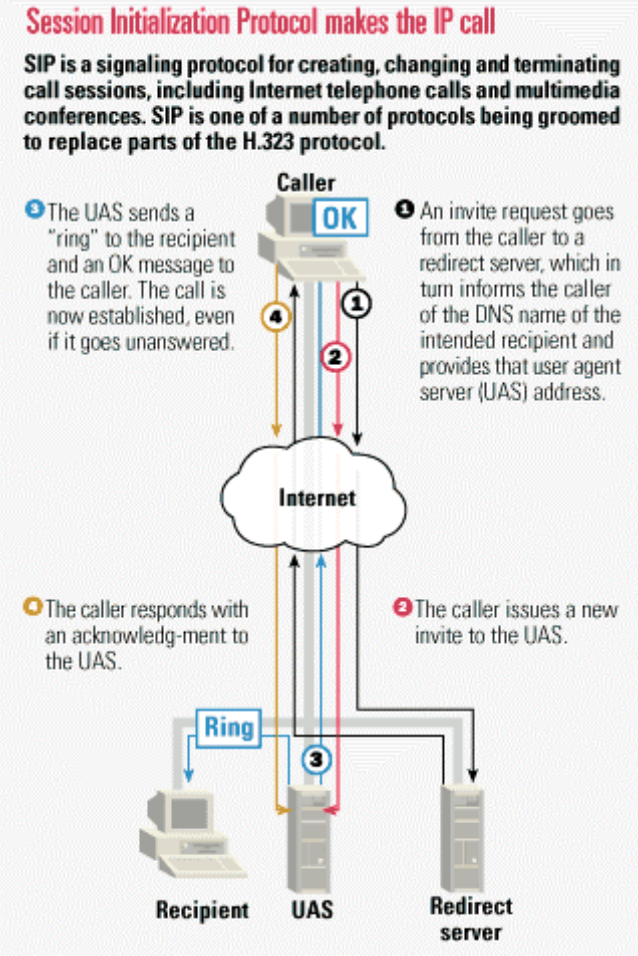


Figure 5: SIP Call Setup

H.323 vs. SIP

H.323 and SIP are both competing for the dominance of IP telephony signaling. There is much debate in the industry as to which protocol is superior, H.323, SIP, or perhaps another protocol that may be in the early stages of development. Currently, there is no clear-cut winner, however, the standards appear to be evolving such that the best features of each are being implemented in the other protocol. For example, the evolution of H.323 from versions 1 through 4 has focused on decreasing call setup delay from six or seven round trips to be on par with SIP's 1.5 round trips. Obviously, this convergence is highly desirable for interoperability issues between the two protocols and thereby reduces signaling overhead.

Both H.323 and SIP support the majority of required end-user functions comparatively equally, such as call setup and teardown, call holding, call transfer, call forwarding, call waiting, and conferencing [Carden]. Yet, functional differences remain, such as H.323's support for message waiting indication and SIP's support for third-party control. In addition, the third version of H.323 provides a more robust mechanism for capabilities exchange – the process by which it is determined whether a particular feature is supported by both participating entities – than does SIP.

Furthermore, H.323 and SIP differ in terms of advantage in Quality of Service (QoS) and management, scalability and flexibility, and interoperability, as described in Table 2. It appears that H.323 has exceptional QoS, management, and interoperability, due to H.323's support for the emerging Differentiated Services/Policy Management to QoS and the protocol's extensive history, respectively. On the other hand, because SIP is a significantly less complex protocol than its bloated counterpart, it scales much better.

Table 2: Advantages of H.323 and SIP in VoIP Features

Feature	Similar	Strengths of H.323 v. 3	Strengths of SIP
QoS and Management	Call setup delay, packet loss recovery, lack of resource reservation capability	Fault tolerance, admission control, policy control	Loop detection
Scalability and Flexibility	Stateless processing, UDP support, inter-server communications for endpoint location	Location of endpoints in other administrative domains	Less complexity, greater extensibility, ease of customization
Interoperability		PSTN signaling interoperability, inter-vendor interoperability	

In terms of impacting VoIP applications, vendors are implementing an assortment of protocols, ranging from the varieties of H.323 to SIP to a proprietary signaling protocol. Presumably, major vendors will support the two major protocols until it becomes clear that either one protocol will fade away or the two approaches will merge. The latter scenario is more likely unless either protocol makes significant advances that the other does not incorporate. Then again, if it remains to be H.323 vis-à-vis SIP, then both implementations of signaling must be supported indefinitely.

Signaling System 7

SS7 is the set of protocols used for call setup, teardown, and maintenance in the PSTN. It is the current suite of protocols used in the North American public network to establish and terminate telephone calls. SS7 is implemented as a packet-switched network, which typically uses dedicated links, nodes, and facilities. In general, SS7 is a non-associated, common channel, out-of-band signaling network – allowing switches to communicate during a call. However, SS7 signaling may traverse real or virtual circuits on links that also carry voice traffic.

Complete coverage of SS7 has appeared in the literature and its full treatment is a very detailed and lengthy subject [Douskalis]. The goal of this section is to provide a concise overview of the signaling functions and interfaces of SS7, because they impact the implementation of internetworking between IP-based telephony and the PSTN.

SS7 Network Topology

SS7 network topologies are constructed using three types of components that are arranged throughout the network in a manner that offers maximum reliability, flexibility, and speed for accomplishing several instrumental tasks in providing telephone service. These elements are Service Switching Points (SSPs), Signaling Transfer Points (STPs), and Service Control Points (SCPs). An SSP is the local telephone exchange, which employs subscriber circuits and trunks connecting to other exchanges. An STP offers transfer and routing services of SS7 messages originating at the SSP. An SCP offers access to the telephone companies' databases via the STP network.

Integrating SS7 and IP

An SS7-IP interface coordinates the SS7 view of IP elements and IP view of SS7 elements. There are three methods to integrate an IP-based network with SS7, each with its advantages and shortcomings. The first approach is to give the access concentrator the ability to interface directly to SS7. The advantage of this approach is that it keeps all the functionality of the SS7/IP integration contained within a single device, making it the most manageable solution. The limitation, however, is scalability, because each access concentrator would require its own connection to the SS7 network.

A simpler way to gain an SS7 connection for several access concentrators is to use an external converter to handle the translation of SS7 to PRI signaling. On the other hand, the converter is also limited in scalability. The final technique bridges the existing PSTN and IP networks, translating the signaling information between the two incompatible network types. Unlike simple converters, however, gateways provide added intelligence for security and control and can be equipped for greater redundancy, resiliency, and scalability. This disadvantage of an SS7 gateway is that it uses a special (and currently nonstandard) interface protocol to talk to the access concentrator.

The industry is moving toward converged network infrastructures to provide a more efficient and effective way of handling increased call volumes as well as delivering new, enhanced services. The integration of SS7 and IP is an important evolutionary step that will also provide significant short-term benefits. Figure 6 illustrates a type of VoIP network employing an SS7-to-IP gateway. SS7 provides the call control on either side of the traditional PSTN, while H.323 provides call control in the IP network. The media gateway (to be further discussed in a subsequent section) provides the circuit-to-voice conversion.

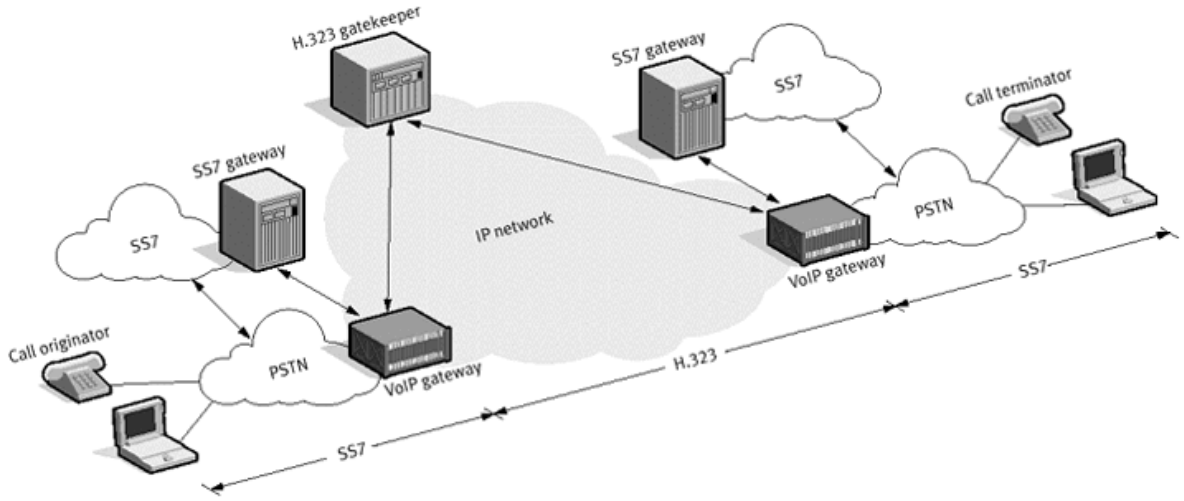


Figure 6: SS7-Based VoIP Network

Voice Coders

Given that packet-switched technology can deliver services far more cost efficiently than today's circuit-switched technology, an efficient voice encoding and decoding mechanism is vital. The purpose of a voice coder (vocoder) – also referred to as coder/decoder, or simply "codec" – is to use the analog signal whose provenance is human speech and transform and compress it into digital data. A number of factors must be taken into account to evaluate the "best" vocoder, where "best" refers to the optimal quality versus bandwidth trade-off. A set of such criteria includes bandwidth usage, silence compression, intellectual property,

look-ahead and frame size, resilience to loss, layered coding, and fixed point vs. floating point. This section will inspect this set more closely. Three popular ITU-T specifications will then ensue, namely G.711, G.723.1, and G.729. First, considerations that need to be taken into account for an optimum vocoder will be presented.

Criteria for Vocoder

The bit-rate of available narrowband codecs using today's technology ranges from 1.2 to 64 kbps, with an inevitable effect on the quality of restituted voice. There is ordinarily – but not always – a trade-off between voice quality and bandwidth used. Using the most efficient codec available today allows quasi-toll quality (where toll quality is equivalent to a PSTN telephone conversation) bandwidth usage to be as low as 5 kbps. As newer and more sophisticated algorithms are developed in the future, this figure will decrease, allowing more samples to be squeezed more efficiently while minimally sacrificing quality, if at all.

During a conversation, a speaker talks an average of 35 percent of the time [Hersent, et al.]; therefore, silence compression or suppression is an important feature. In a point-to-point conversation, this feature saves approximately half of the bandwidth; the savings are even greater in a decentralized multicast, where packets are efficiently destined to multiple nodes on the network. Silence compression includes three major components: voice activity detector (VAD), discontinuous transmission (DTX), and comfort noise generator (CNG). VAD is responsible for determining when the user is talking and when he is silent. DTX is the ability of a codec to stop transmitting frames when the VAD has detected a silence period. CNG is used to recreate background noise so that the line is not completely silent even if neither party is transmitting.

Intellectual property refers to the royalties manufacturers must pay to implement some codecs in their products. This can cause a certain vendor's VoIP application to use an inefficient codec that may be cheaper to license. Hence, factoring for royalties is another aspect that must be considered in evaluating vocoders – namely the trade-off between pecuniary cost and performance.

Most narrowband codecs compress voice in chunks (frames) and need look-ahead information. That is, these codecs require information about the samples immediately following the samples that they are currently encoding. The minimal delay introduced by a coding/decoding sequence is the frame length plus the look-ahead size (algorithmic delay). Codecs with a small frame length have a lesser delay than those with longer frame length, but introduce a larger overhead, as mentioned earlier. Most implementations choose to send multiple frames per packet and the real frame length to take into account is the sum of all frames stacked in a single IP packet. The smaller the frame size, the more frames in an IP packet, thereby there is minimal influence on latency. In fact, it is even better to employ codecs that have been designed for the longest frame length (within acceptable delay), since this allows even more efficient coding techniques.

In IP networks, packet loss will occur, which in turn causes codec frame loss. This loss is exacerbated on the Internet, where packet loss occurs in a correlated manner so that several consecutive packets may be lost. Thus, implementing rudimentary redundancy or recovery from packet loss may not be productive if the algorithm only protects against intermittent or nonconsecutive packet loss. It is possible to reduce the frame loss associated with packet loss through some advanced techniques. Forward error correction (FEC) styled redundancy can be used to recover from serious loss conditions, but at the expense of delay. Alternatively, multiple send redundancy can be utilized, but it would transmit more packets, and if router congestion were the cause, then using this technique would not be beneficial.

Most codecs today can only multicast voice at a single level of quality (bit-rate). Therefore, it would not be possible to transmit the same data at different quality rates to multiple listeners without sending separate streams. Some codecs still at the experimental stage can produce several data streams simultaneously: one

with the core information needed for "military quality" reception and others with more information to re-build a higher fidelity sound.

Digital signal processors (DSPs) are optimized for operations frequently encountered in signal processing algorithms. Floating-point DSPs are capable of operating on floating-point numbers. Fixed-point DSPs can operate on two fixed-point operands only if the power of two is the same on both operands. Therefore, the latter is less powerful, but also computationally less expensive.

ITU-T Specifications

The ITU has a rigorous process in approving vocoders. Before a codec is chosen, the ITU evaluates the mean opinion scores (MOS) and usually requires "toll quality" or better, where toll quality is defined in G.726 via RFC 2422. Among many criteria, the codec must meet the following:

- Acceptable quality for men and women of varying ages, accents, and languages
- Resilience to background noise
- Minor degradation of voice quality after several successive coding/decoding processes
- Ability to pass dual tone multi-frequency (DTMF) signals transparently
- Ability to easily transcode the coded signal into other ITU standard coders
- Satisfactory quality even after some frame loss

The following three ITU-based audio codecs are frequently used in VoIP applications.

G.711

The G.711 describes a relatively simple way to digitize analog data by using a semi-logarithmic scale, called the companded pulse code modulation (PCM). Its goal is to increase the resolution for small signals, while large signals are treated proportionally. The encoded stream is 64 kbps, consisting of 8 kHz sampling of 8 bit signals. The frame length is eight 125 μs samples, or 1 ms.

G.723.1

The G.723.1 codec has been selected as the baseline codec for narrowband H.323 communications by the International Multimedia Telecommunications Consortium (MTC) VoIP forum. G.723.1 is a coded representation that can be used for compressing the speech component of multimedia services at a low bit rate (compared to G.711's 64 kbps). The vocoder has two bit rates associated with it, 5.3 and 6.3 kbps, whose mode of operation can change dynamically at each frame. Its frame length is 30 ms, however, another 7.5 ms delay is necessary for its look-ahead buffer, resulting in a total algorithmic delay of 37.5 ms.

The G.723.1 vocoder encodes speech in frames using linear predictive analysis-by-synthesis coding. The excitation for the high rate coder is multi-pulse-maximum likelihood quantization (MP-MLQ), whereas the low rate coder is algebraic-code-excited linear prediction (ACELP). The codec is capable of providing silence compression: VAD, DTX, and CNG.

G.729A

The G.729/G.729A vocoder uses conjugate-structure, algebraic-code-excited linear prediction (CS-ACELP) coding technique. It produces a speech rate of 8 kbps and costs an algorithmic delay of 15 ms (10 ms frame length and 5 ms of look-ahead time). G.729A is a reduced-complexity version of the original G.729 specification. The codec, like G.723.1, is also capable of providing silence compression via VAD, CNG, and DTX schemes.

These ITU-approved codecs are summarized in Table 3, where the expected MOS can range from a scale of 1 (bad) to 5 (excellent). The frame length does not include the 7.5 ms and 5 ms of look-ahead buffer latency that G.723.1 and G.729A respectively impose. It could be anticipated that the MOS would decrease with bit rate – an ostensible trade-off. However, G.729A had a lower MOS than the 6.3 kbps version of G.723.1. This can be ascribed to the notion that the MP-MLQ algorithm can better reproduce voice than an ACELP-derived one.

Table 3: Summary of ITU Vocodecs

Voice Coder	Bit Rate	Frame Length	Expected MOS
G.711 (PCM)	64 kbps	1 ms	4.1
G.723.1 (MP-MLQ)	6.3 kbps	30 ms	3.9
G.723.1 (ACELP)	5.3 kbps	30 ms	3.65
G.729A (CS-ACELP)	8 kbps	10 ms	3.7

Future Coders

In the industry, there is work in developing coders. One recently established coder is the Mixed Excitation Linear Predictive (MELP) vocoder, which utilizes a miniscule 2.4 kbps. Surprisingly, informal experiments suggest that the enhanced 2.4 kbps MELP coder performs as well as the higher bit rate 4.8 kbps FS1016 CELP standard [Causal]. It is professed that the coder has been optimized for performance in acoustic background noise and in channel errors, as well as for efficient real-time implementation. Listening to the samples do corroborate that the voice is comprehensible, but the quality is that of analog cellular telephone service. Even more, there has been development of a high quality speech coder based on the Multi-Band Excitation (MBE) model operating at both 2.4 kbps and 1.2 kbps, which is half of MELP's bit-rate consumption.

The trend in industry appears to be developing coders that utilize less bandwidth than their predecessors. Since the early 1990s, the ITU has forged ahead from the 64 kbps G.711 to the more recent G.723.1 specification that consumes less than one-twelfth of that bandwidth. According to [GIPS], this bandwidth savings comes at the cost of lower quality and lower robustness to hostile network environments. Given the inevitable increase in the average user's bandwidth over time, is the effort in industry being optimally utilized to improve VoIP by creating lower bit-rate coders? Alternatively, would this effort be better directed at improving quality first, then addressing bandwidth, as long as the bit-rate consumed is in the tens of kilobits per second (to satisfy the 28.8 kbps Internet connection)? These questions will be further examined later in the paper.

Transport

Once signaling and encoding occur, RTP and RTCP – both defined in RFC 1889 – are utilized to transport the voice packets. Media streams are packetized according to a predefined format and placed in RTP packets. RTP provides delivery monitoring of its payload types through sequencing and timestamping. RTCP offers insight on the performance and behavior of the media stream, such as voice stream jitter. Fortunately, RTP and RTCP are designed to be independent of the signaling protocol, encoding schemes, and network layers implemented.

In essence, RTP provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio (voice). RTP does not address resource reservation and does not guarantee QoS for

real-time services. The data transport is augmented by a control protocol, RTCP, to allow monitoring of the data delivery in a manner scalable to large multicast networks, and to provide minimal control and identification functionality.

RTP

RTP, as mentioned, provides end-to-end delivery services for data with real-time characteristics. Those services include payload type identification, sequence numbering, timestamping, and delivery monitoring. Applications typically run RTP on top of UDP to make use of its multiplexing and checksum services. In fact, both protocols contribute parts of the transport protocol functionality; however, RTP may be used with other suitable underlying network or transport protocols [RFC 1889].

RTP does not intrinsically provide any mechanism to ensure timely delivery or provide other QoS guarantees, but relies on lower-layer services to do so. It also requires the use of a signaling protocol to set up the connection and negotiate the media format that will be used. RTP does not guarantee delivery or prevent out-of order delivery, nor does it assume that the underlying network is reliable and delivers packets in sequence. The sequence numbers included in RTP allow the receiver to reconstruct the sender's packet sequence, but sequence numbers might also be used to determine the proper location of a packet without necessarily decoding packets in sequence.

RTP is intended to be malleable to provide the information required by a particular application, and will often be integrated into the application processing rather than being implemented as a separate layer. It is intended to be tailored through modifications and/or additions to the headers as needed.

The basic (standard) RTP header consists of only twelve bytes. To satisfy application-specific requirements, H.225.0 specifies modifications and nonstandard extensions to RTP header – a list of contributing source identifiers (CSRCs). For the purposes herein, only the standard RTP header fields will be described. The RTP header comprises: two bits for version, one-bit padding, another bit for extension, four-bit CSRC count, one bit for marker, seven bits for payload type, 16-bit sequence number, 32-bit timestamp, and 32 bits for the synchronization source identifier (SSRC).

RTCP

RTCP is based on the periodic transmission of control packets to all participants in the session, using the same distribution mechanism as the data packets. The underlying protocol must provide multiplexing of the data and control packets, for example, using separate port numbers with UDP. On this note RTP must be assigned an even UDP port number and the corresponding RTCP is assigned the next higher (odd) numbered UDP port. RTCP performs the following four functions:

- Provides feedback on the quality of the data distribution (primary function)
- Carries a persistent transport-layer identifier for an RTP source, canonical name
- Controls the rate in order for RTP to scale up to a large number of participants
- Conveys minimal session control information

The first three functions are mandatory when RTP is used in an IP multicast environment, and are recommended for all environments, whereas the fourth function is optional. These functions are exhaustively performed in the five types of RTCP packets: sender report, receiver report, source description, hang-up from a session, and application-specific packets. Quintessentially, RTCP is primarily useful as a heartbeat monitoring whether data delivery is occurring at all, and for endpoints to decide whether parts of the corresponding RTP stream are being lost in cases of network malfunction.

Gateway Control

Gateways are responsible for converting packet-based audio formats into protocols understandable by PSTN systems. The aforementioned signaling protocols (such as H.323 and SIP) provide more services than are necessary, such as service creation and user authentication, which are irrelevant for gateways. Vendors have gravitated towards simplified Device Control Protocols (DCPs), rather than all-encompassing signaling protocols [deCarmo].

Figure 7 displays processing that must occur in a gateway to convert PSTN to IP and vice versa. The network interface in a gateway includes any hardware or software that connects the gateway to the telephone system or network. Digital signal processing is typically achieved with dedicated hardware and associated software algorithms that perform voice coding described in a previous section. Specifically, the DSP subsystem (de)compresses speech, detects tones and silence, generates tones and comfort noise, and cancels echo. To efficiently perform vocoding, DSP implementations depend on processing entire frames of data at once. Finally, between the DSP processing and passing the data to the WAN, there are a number of packet-handling processes that must be encountered. A nontrivial amount of gateway-incurred latency is present, which affects perceived voice quality.

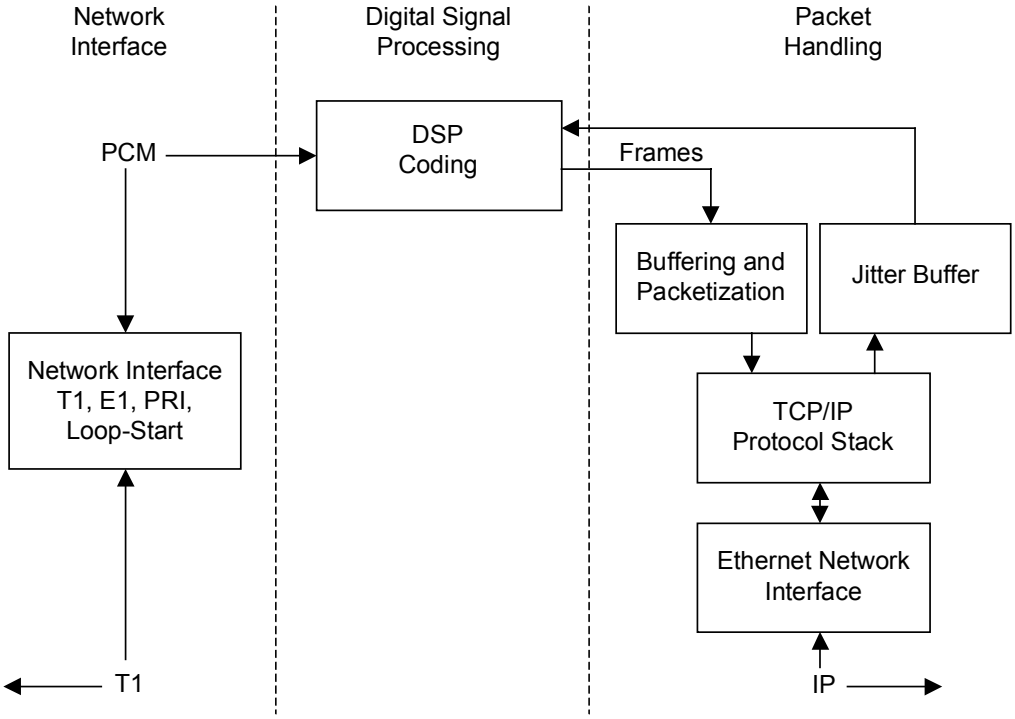


Figure 7: Gateway Processing

The Internet Protocol Device Control (IPDC) was a first-generation DCP whose goal was to create "dumb" endpoints (or gateways). It separated services from the gateway and placed network intelligence on a server. The Simple Gateway Control Protocol (SGCP) was created at approximately the same timeframe as IPDC, and it similarly revolves around intelligent servers and simple endpoints. The two standards merged to combine the best features of each and formed the Media Gateway Control Protocol (MGCP).

Media Gateway Control Protocol

A media gateway is a network element that provides conversion between the information carried on telephone circuits and data packets carried over the Internet or over other IP networks. MGCP is an IETF standard that defines gateway control. It is the lightweight telephony protocol that aims to reduce complexity and increase reliability and interoperability for Internet telephones. MGCP also enhances security since all critical information is stored on trusted servers, thereby, MGCP devices are treated as untrusted network elements. Unfortunately, MGCP partially overlaps with signaling protocols, which obscures the boundary between signaling and gateway control.

Megaco

While MGCP was evolving, a parallel effort was underway at ITU, which was developing H.GCP – a protocol that contains the minimal features necessary to create gateway. The ITU and IETF pooled their efforts and created the Megaco protocol (H.248). Although Megaco is still being refined, it contains all of MGCP's functionality, plus superior controls over analog telephone lines and the ability to transport multiple commands in a single packet.

The Megaco framework could potentially enable service providers to offer a wide variety of converged telephone and data services. Media gateways will be the junctions that provide a path between switched and packet networks for voice. Megaco implementations can also be enhanced using extension methods: packages. These packages are sets of commands, related events, and statistics that can be added to a basic Megaco device. When the media gateways are initially set up for communication, a vocoder approach will normally be used. Megaco-related standards will enable support of existing and new applications of telephone service over hybrid telephone networks that will contain a mix of switched, IP, and ATM technology.

Table 4 compares the popular DCPs. Megaco appears to incorporate the desired features of gateway control.

Table 4: Comparison of DCPs

Feature	IPDC	SGCP	MGCP	Megaco
ASCII-based	No	Yes	Yes	Yes
Binary	Yes	No	No	No
Trunking controls	No	No	No	Yes
Event controls	Yes	No	Yes	Yes
Packages & Extensibility	Yes	No	Yes	Yes

Wireless Networks

An emerging trend for implementing VoIP – and in general, connecting computing devices – is in wireless networks. A wireless LAN (WLAN) is a data transmission system designed to provide location-independent network access between computing devices by using radio waves rather than a cable infrastructure. WLANs give users wireless access to the full resources and services of the LAN across a building or campus environment. There are some fundamental concerns that WLANs introduce, however, that are not present in the typical wireline systems discussed thus far. These issues include a higher frequency of dropped packets, larger latency, and more jitter.

Why is wireless connectivity and communications important? There are numerous benefits of utilizing WLANs. In any network environment, users would be able to access the network far beyond their personal desktops, giving these mobile users much-needed freedom in their network access. Specifically, they can access information from anywhere in the building or campus. A WLAN system provides a powerful combination of wireline network throughput, mobile access, and configuration flexibility. It liberates users from

tethered access to the network backbone, giving them anytime, anywhere network access. Applications include VoIP from mobile personal communications devices.

Potential uses of VoIP in a wireless network will be discussed; however, two popular wireless standards will be described first. These key wireless technologies, IEEE 802.11x (which collectively refers to the 802.11, 802.11b, and 802.11a standards) and Bluetooth will be described, followed by a discussion on how each affects VoIP. More emphasis will be given on 802.11b, because, as will soon be mentioned, it is the better candidate for using VoIP on a wireless device and is currently widely available. A thorough discussion and detailed analysis on 802.11 and Bluetooth is not within the scope of this document; rather, that information can be found in [3Com] and [Champness], and [Coffee], respectively. Instead, this section will serve as the foundation to provide an overview and discussion on their affect on VoIP.

IEEE 802.11x

The IEEE ratified the original 802.11 specification in 1997 as the standard for WLANs. That version of 802.11 provided for 1 Mbps and 2 Mbps data rates and a set of fundamental signaling methods and other services. Recognizing the critical need to support higher data-transmission rates, the IEEE ratified the 802.11b standard (also called 802.11 High Rate) for transmissions of up to 11 Mbps two years afterward. More recently, the 802.11a specification has been standardized, which permits transmissions of up to 54 Mbps. When implementing this updated standard, WLANs can achieve performance, throughput, and availability comparable to wireline Ethernet.

The 802.11 standards focus on the first two layers of the OSI/RM shown in Figure 2, namely, the Physical and Data Link Layers, respectively. Thereby, any LAN application, including VoIP, can run on an 802.11-compliant WLAN as easily as they run over Ethernet. The original 802.11 standard defines the basic architecture, features, and services of 802.11b. The 802.11b and 802.11a specifications affect only the physical layer, adding higher data rates and more robust connectivity.

Overview of 802.11

802.11 defines two pieces of equipment: a wireless station – typically a PC equipped with a wireless NIC – and an access point, which bridges the wireline and wireless networks. The access point acts as the base station for the wireless network, aggregating access for multiple wireless stations onto the wireline network. Hence, mobile devices can roam between access points seamlessly and transparently. The standard defines two modes: infrastructure mode and ad hoc mode. In the former mode, the wireless network consists of at least one access point connected to the wireline network infrastructure and a set of wireless end stations. The ad hoc mode is merely a set of 802.11 wireless devices that communicate directly with one another without using an access point or any connection to a wireline network.

802.11 WLANs communicate using radio waves because these waves penetrate through many indoor structures and can reflect around obstacles; thereby, line-of-sight communication between the access point and wireless station is not required. The three physical layers originally defined in the standard include two spread-spectrum radio techniques and a diffuse infrared specification. The radio-based standards operate within the 2.4 GHz Industry, Scientific, and Medical (ISM) band. Spread-spectrum techniques increase reliability, boost throughput, and allow unrelated products to share the spectrum without explicit cooperation. Still, the possibility of interference between these devices exists.

802.11 defines data rates of 1 Mbps and 2 Mbps via radio waves using frequency hopping spread spectrum (FHSS) or direct sequence spread spectrum (DSSS). FHSS divides the 2.4 GHz band into seventy-five 1-MHz sub-channels. The conversation between two endpoints occurs over a different hopping pattern; patterns are designed to minimize the chance of two senders using the same sub-channel simultaneously. In

contrast, DSSS divides the 2.4 GHz band into fourteen 22-MHz channels. Data is sent across one of these 22 MHz channels without hopping to other channels.

Enhancements via 802.11b

FHSS is severely impeded by regulations – sub-channels are restricted to a bandwidth of 1 MHz. This forces FHSS systems to spread their usage across the entire 2.4 GHz band. Thus, these systems must hop often, which leads to a nontrivial amount of overhead. Given the bureaucratic limitations of FHSS, 802.11b data is encoded using DSSS technology to achieve higher speeds. The physical layer of this specification can support two speeds, 5.5 Mbps and 11 Mbps. This indicates that 802.11b systems will interoperate and thus be fully backward compatible with 802.11 systems that employ DSSS, but not with FHSS radios.

Of inconsequential importance to voice applications due to its orders of magnitude lower bit-rate, but of significance to other applications, is that 802.11b allows transmission speeds to fluctuate as a function of range and noise between the access point and wireless device. 802.11b WLANs use dynamic rate shifting, allowing data rates to be automatically adjusted (either increased or decreased, in increments of 11, 5.5, 2, and 1 Mbps), to compensate for the changing nature of the radio channel. Moreover, a WLAN will always have slower performance than an equivalently rated Ethernet-based LAN. WLANs incur extra overhead due to the way those senders handle packet collisions and demand additional acknowledgment packet receipts from the receiver. In fact, 802.11b is at best only 85 percent efficient at the physical layer [Conover].

Enhancements via 802.11a

An extension to the 802.11 architecture – 802.11a – defines different multiplexing techniques that can achieve data rates up to 54 Mbps by exploiting the 5 GHz Unlicensed National Information Infrastructure (UNII) frequency band. This standard has only recently been ratified, and no products have been yet marketed, however such products are expected to be available by mid-2001. Thereby, its performance cannot be experimentally measured until the appropriate equipment (access points and wireless NICs) is available.

Instead of using spread spectrum technology used in 802.11b, 802.11a taps into a complex technology called Orthogonal Frequency Division Multiplexing (OFDM), which helps improve speed and signal quality. OFDM can significantly improve utilization of the wireless channel, in both driving data rates up and getting higher performance [Moore]. This is attainable because unlike spread spectrum, which has to send signals in direct sequence, OFDM can separate the wireless channel into sub-frequencies to be transmitted at low data rates in parallel. This enables simultaneous transmission of high bandwidth video with low rate voice.

The impact on delay, packet loss, and jitter, is unknown since there are no products currently available. Nevertheless, outside the quintupled transmission rate, it is expected to behave similarly to 802.11b, with another key exception: less interference. Since 802.11a operates at the 5 GHz frequency band, it will avoid interference with Bluetooth and other equipment that operate at the 2.4 GHz band – unless other appliances start to impinge on the 5 GHz band.

Support for Time-Sensitive Data

Time-sensitive data, such as voice, is supported in the 802.11 Media Access Control (MAC) specification through the Point Coordination Function (PCF). In this mode, one access point controls access to the media. The access point will poll each station for data, and after a set time, move on to the next station. No station is allowed to transmit unless it is polled, and stations receive data from the access point only when they are polled. In essence, PCF provides a time-division duplexing capability to accommodate time-bounded, connection-oriented services, such as voice.

Since PCF gives every station a turn to transmit in a predetermined fashion, a maximum latency is guaranteed. An obvious shortcoming to PCF is that it is not scalable since a single point needs to have control of the media access and must poll all stations. Nonetheless, for the latency-sensitivity innate in voice applications, jitter is controlled because the maximum delay is known. Potential use for VoIP via 802.11b is described in a succeeding section.

802.11x Security

802.11 provides for both MAC layer access control (authentication) and encryption mechanisms, which are collectively known as Wired Equivalent Privacy (WEP). The objective of WEP is to provide WLANs security equivalent to their wireline counterparts. For access control, an identification value is programmed into each access point and is mandatory knowledge in order for a wireless client to associate with an access point. In addition, there is a provision for a table of MAC addresses to be included in the access point, restricting access to clients whose MAC addresses are on the list. For data encryption, the standard provides for optional encryption using a 40-bit RC4 PRNG shared-key algorithm [Zyren and Petrick]. All data sent and received while the end station and access point are associated can be encrypted using this key.

Beyond Layer 2, 802.11b WLANs support the same security standards supported by wireline LANs for access control and encryption. For example, network operating system (OS) logins and Internet Protocol Security (IPSec, discussed herein) can be utilized for additional access control and encryption, respectively. These higher-layer technologies can be used to create end-to-end secure networks encompassing both wireline and wireless LAN components, with the wireless piece of the network gaining unique additional security from the 802.11 feature set.

Bluetooth

Bluetooth is the specification used as a blueprint for IEEE's 802.15 wireless personal area network (WPAN) initiative. It is typically used for providing device-to-device connectivity on an ad hoc basis, whereas WLAN systems target as a wireless replacement or extension of the LAN infrastructure. Bluetooth is meant to be more than just a radio channel. It is proposed to be an intelligent and robust method for allowing devices to seek and provide one another services in ways that streamline mobile computing and enable more responsive behavior from wireline networks. Bluetooth operates in a band of radio frequencies that is just above 2.4 GHz, like IEEE 802.11b, and can thus cause interference. This will be explained shortly.

Bluetooth was specifically designed to accommodate both synchronous (such as voice) communications and asynchronous (data) communications. This technology is meant more as a wire replacement than a LAN topology; thereby, it is likely that Bluetooth will coexist with other standards that are more LAN-oriented [Flood]. The Bluetooth approach aims to dramatically reduce the complexity of the protocol and reduce the transmitter power, and consequently, the coverage range to lower cost and simplify operation.

To accomplish these goals, Bluetooth uses an arrangement of very small "piconets" that can support only eight nodes at a time. All network connectivity is ad hoc, which means there is no network established until a device chooses to communicate. When communications are established, devices within the piconet determine a master node, which synchronizes timing and controls communications. Communication rates between Bluetooth devices can reach 1 Mbps at a radius of 10 meters, but this is highly dependent on how Bluetooth is implemented on those devices. For example, some vendors are pushing to expand the range of Bluetooth connections to 100 meters [Mannion].

Bluetooth uses spread spectrum, in which multiple users share a single spectrum slice but use sophisticated information processing to identify their own signals while ignoring others, like 802.11. Specifically, Bluetooth uses frequency hopping, wherein senders and receivers follow preplanned sequences of moves be-

tween narrow channels within an agreed-upon range. This rapid movement – 1600 hops per second – is essentially to avoid collisions with other packets.

Bluetooth Security

In any wireless implementation, security is paramount. Like 802.11x, Bluetooth addresses the area of security. Devices connecting via Bluetooth enjoy automatically negotiated link-level security, with key sizes up to 128 bits. However, Bluetooth's protocols only establish the identity of a device, not its user. Security negotiations take place only when a connection is first established, not on subsequent connection exchanges. This means that Bluetooth alone cannot enforce one-way transfers of data. Therefore, any applications that run on top of Bluetooth connectivity must implement user authentication and database or service access control to enhance overall security. This can be considered as a trade-off between security and convenience from the user's point of view.

Interference with 802.11b

Since applications for both IEEE 802.11b and Bluetooth are targeted for similar users and environments, it is likely that both radios will come in close proximity to each other. Moreover, both technologies operate at the 2.4 GHz, along with the nascent HomeRF standard and some microwave ovens and cordless telephones, making it possible for the wireless radios to be adversely affected from these devices. Studies have been conducted to determine the degree of harmful, mutual interference caused by the radios [Zyren]. The degree in which an 802.11 device is susceptible to interference from nearby Bluetooth transmitters is clearly dependent upon the strength of the desired DSSS signal from the access point, which in turn is dependent on the range. According to the cited study, 802.11b WLANs show graceful degradation and acceptable reliability in presence of significant levels of Bluetooth interference.

Another study was conducted to determine the impact of an 802.11 DSSS WLAN system on a Bluetooth link [Haartsen and Zürbes]. That study assumed an office environment with few WLAN access points but many WLAN devices. Furthermore, their study differentiated between the impact on Bluetooth data and impact on Bluetooth voice. The voice link was disturbed in fewer than one percent of the cases when the Bluetooth operating distance remained below 2 meters, however, if the operating distance was increased to 10 meters, the probability escalated to eight percent. For the data link, a throughput reduction of more than ten percent occurred with a 24 percent probability at a distance of 10 meters. Because of the limited frequency overlap of the WLAN and Bluetooth systems, the worst case throughput reduction is 22 percent.

Summary of Wireless Technologies

Table 5 summarizes the two wireless protocols, along with (Fast) Ethernet, in terms of speed and range. Not shown in this table is 802.11a, which has a maximum transmission rate of 54 Mbps. However, because no products have been yet released implementing this standard, its range is not yet known.

Table 5: Summary of Wireless and Wireline Protocols

Protocol	Transmission Rate	Approximate Range
Bluetooth	1 Mbps	33 feet
802.11b	11 Mbps	300 feet
Ethernet/Fast Ethernet	10/100 Mbps	Not applicable

IP Security

Both 802.11b and Bluetooth wireless technologies are amenable to higher-level security protocols. One such protocol gaining rapid popularity is Internet Protocol Security (IPSec). IPSec is a framework of open standards for ensuring secure private communications over the Internet. This IETF-developed standard provides security at the network or packet-processing layer of network communication.

IPSec provides two choices of security service: Authentication Header (AH), which essentially allows authentication of the sender of data, and Encapsulating Security Payload (ESP), which supports both authentication of the sender and encryption of data. The specific information associated with each of these services is inserted into the packet in a header that follows the IP packet header. Consequently, IPSec ensures confidentiality, integrity, and authenticity of data communications across a public network. Consequently, wireless or wireline VoIP can occur securely, notwithstanding the cost of greater delay and jitter.

Impact of Wireless VoIP

There are numerous potential uses for VoIP with either of these two wireless technologies. Most mobile devices today are shipping with PC Card slots that can house 802.11b NICs, whereas there remains a dearth of Bluetooth PC cards. This fact, coupled with 802.11x's superior operating range, leads many to believe that 802.11b will find more functions in wireless VoIP. Plausible scenarios with either wireless technology will be discussed next.

VoIP via Wireless LAN

VoIP over wireline networks can be migrated to the wireless side via 802.11b seamlessly with few additional issues. Devices that can currently use 802.11b products include laptop computers, personal digital assistants, and wearable computers. Figure 8 depicts a scenario where handheld computers communicate with their respective access points over IP.

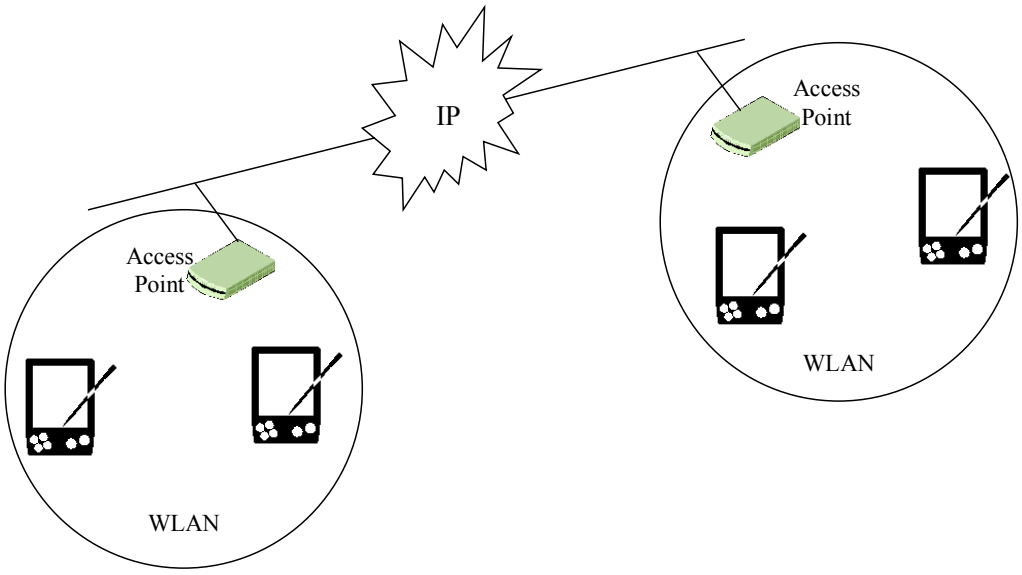


Figure 8: Wireless Communications via 802.11b

As long as a VoIP client can be installed on the hardware and OS tandem, wireless VoIP is feasible. The benefit here is that, it is envisioned that in the near future, a significant number of users will utilize a form of handheld computer (such as a wearable computer or personal digital assistant), which will have wireless network access via 802.11x. Thus, if mobile users need to communicate with each other, they are not re-

stricted to leaving voice mail messages on the campus PSTN telephone system or relying on a text-based paging system. They can simply dial the person's VoIP number (or click on the person's name, depending on the configuration) and enjoy ubiquitous communication. Of course, the handheld device must be capable of supporting sound input and output.

On the Bluetooth side, a perspective on the viability of mobile VoIP is permitting users to use Bluetooth-enabled devices to make VoIP calls. These devices would communicate with the access point(s), which would link the callers to the wireline LAN and ultimately to the gateway to call a PSTN telephone. Naturally, the gateway is unnecessary to simply conduct a call to another Bluetooth device, similar to the 802.11b pictorial shown in Figure 8. The described scenario could serve as a restricted alternative to using cellular telephones – restricted in the sense that it is only feasible in environments where Bluetooth communications devices are within 10 meters of the access point(s).

One of many implementation issues is that the handheld device must be within proximity of an access point, where the range and throughput can vary depending on the environment, such as number and material of walls and other obstacles the waves must propagate through or deflect off. Of course, the wireless protocol utilized is an essential factor on range and throughput. Providing a sufficient number of access points throughout the operating area can mitigate this issue of non-overlapping coverage.

Quality of VoIP

The basic routing philosophy on the Internet is "best-effort", which serves most users well enough but isn't adequate for the time-sensitive, continuous stream transmission required for VoIP. It is imperative for an implementation of VoIP to remain cognizant of quality. Quality encompasses many factors; the ones that will be examined here are QoS, packet loss, jitter, and latency.

Quality of Service

QoS refers to the ability of a network to provide better service to selected network traffic over various underlying technologies, including IP-routed networks. QoS features are implemented in network routers to provide better and more predictable network service by:

- Supporting dedicated bandwidth
- Improving loss characteristics
- Avoiding and managing network congestion
- Shaping network traffic
- Setting traffic priorities across the network

This has an immediate impact on VoIP. In order to achieve toll quality voice, the application necessitates high QoS support, such as reserving enough bandwidth (as determined by the codec) and proactively avoiding congested networks. To configure an IP network for real-time voice traffic, the appropriate QoS needs to be selected for both edge and backbone routers in the network. Edge routers perform packet classification admission control, and configuration management; in contrast, backbone routers perform congestion management and congestion avoidance.

Real-time voice applications have different characteristics and requirements from those of traditional data applications. Because they are real-time based, voice applications tolerate minimal variation of delay affecting delivery of their voice packets. Voice traffic is also intolerant of packet loss, out-of-order packets, and jitter, all of which gravely degrade the quality of the voice transmission delivered to the recipient end user.

To effectively transport voice traffic over IP, mechanisms are required that ensure reliable delivery of packets with low and controlled latency.

Another approach utilizes RSVP, which is a relatively new protocol developed to enable the Internet to support QoS. Using RSVP, a VoIP application can reserve resources along a route from source to destination. RSVP-enabled routers will then schedule and prioritize packets to fulfill the QoS. RSVP is part of the Internet Integrated Service (IIS) model, which ensures best-effort service, real-time service, and controlled link-sharing.

IPv6 QoS Support

While QoS is an extension to the current version of IP (IPv4), the succeeding protocol, IPv6, will inherently support QoS. However, IPv6 also has a much larger packet, so it is possible that while QoS will alleviate much of the jitter and congestion voice packets currently suffer, it could come at the cost of increased latency. Because IPv6 headers are 40 bytes long, compared with 20-byte IPv4 headers, the overhead per packet is doubled. This may pose a problem for codecs that only succeed with diminutive packets. Nevertheless, this larger packet overhead can be partially offset if IPv6 provides for efficient compression schemes for the header.

Packet Loss

UDP/IP networks cannot provide a guarantee that packets will be delivered at all, much less in order. Packets will be dropped under peak loads and during periods of congestion. Due to time sensitivity of voice transmissions, the normal TCP-based retransmission schemes are not appropriate. Approaches used to compensate for packet loss include interpolation of speech by replaying the last packet, and sending of redundant information. Packet losses greater than ten percent are generally intolerable unless the encoding scheme implemented provides extraordinary robustness.

Jitter

Because IP networks cannot guarantee the delivery time of data packets (or their order), the data will arrive at very inconsistent rates. The variation in inter-packet arrival rate is jitter, which is introduced by variable transmission delay over the network. Removing jitter to allow an equable stream requires collecting packets and storing them long enough to permit the slowest packets to arrive in time to be played in the correct sequence. Each jitter buffer, which is used to remove the packet delay variation that each packet is subjected to as it transits the network, adds to the overall delay.

Latency

Latency is the time delay incurred in speech by the IP telephony system. One-way latency is the amount of time measured from the moment the speaker utters a word until the listener actually hears the word. Round trip latency, of course, is the sum of the two one-way latency figures that compose the user's call. The lower the latency, the more natural interactive conversation becomes and the additional delay incurred by the VoIP system is less discernable. In PSTN calls, the round trip latency of calls originating and terminating within the continental United States is under 150 ms.

In a VoIP implementation that is primarily used in a cost-reduction or toll bypass application, studies suggest that users will tolerate one-way latency of up to 200 ms [Brooktrout]. Furthermore, user perception of the link quality can be mapped in terms of one-way latency, as shown in Figure 9.

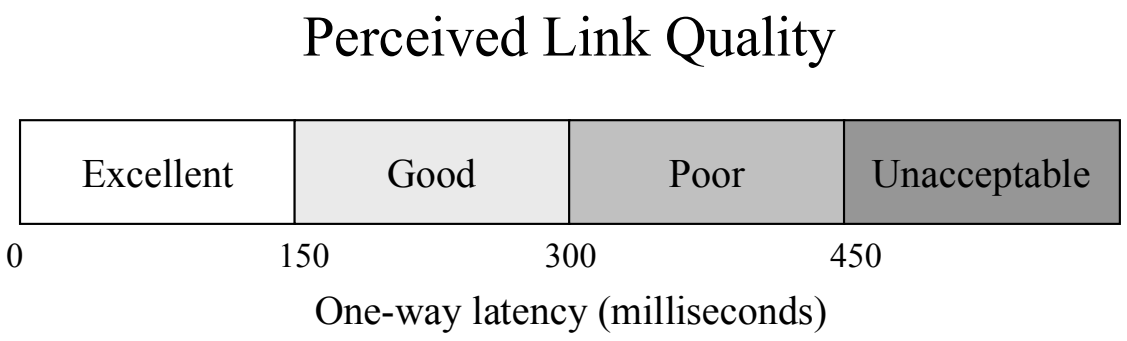


Figure 9: Quality Perception vs. Latency

While literature on the effects of delay is relatively extensive, it is sometimes slightly inconsistent. For example, the 1996 ITU Recommendation G.114 for one-way end-to-end transmission time limits is:

- Under 150 ms: acceptable for most user applications
- 150 to 400 ms: acceptable provided that administrators are aware of the transmission time impact on the transmission quality of user applications
- Over 400 ms: unacceptable for general network planning purposes

More will be mentioned in a subsequent section on how certain VoIP implementations can trade-off latency for an improvement or deterioration in another feature or performance of another voice quality-related parameter, as dictated by the codec.

There are several sources of delay that can contribute to the one-way (and thus round trip) latency. Generally, the VoIP system is constructed using gateways to interface existing telephone equipment together over a WAN, as shown in Figure 1. Thereby, much of the latency is introduced in two primary sources: in the gateways at either end, and by the network that connects the gateways, as depicted in Figure 7. Naturally, there is latency in the vocoder utilized as well, but this is usually limited to tens of milliseconds.

Consequential Issues

Two problems that result from a high end-to-end delay in a voice network are echo and talker overlap. Echo becomes a problem when the round-trip delay is more than 50 ms. Since echo is perceived as a significant quality problem, the VoIP system must address the need for echo control and implement echo cancellation. Talker overlap – the problem of one caller stepping on the other talker's speech – is exacerbated when the one-way delay is greater than 250 ms. The end-to-end delay budget, therefore, is the major constraint and driving requirement for reducing latency through a packet network.

VoIP Experiments

Experiments were performed to qualitatively evaluate VoIP performance using available technology within a LAN and when traversing from a WAN to PSTN. For the former test, Cisco Systems, Inc.'s IP SoftPhone and HardPhone were employed. For the trials of making IP-to-PSTN voice calls, Dialpad.com, Inc.'s service was utilized.

Cisco IP Phone

Several experiments were conducted to qualitatively test voice over IP with some products in existence today. Cisco's VoIP product was utilized to evaluate IP telephony from a Microsoft Windows-based applica-

tion ("SoftPhone") and a hardware-based IP telephone ("HardPhone"). The SoftPhone experiments were conducted when the PC was tethered to a LAN and when the laptop was on a wireless (802.11) LAN. The HardPhone resembles a traditional corded telephone, except it has an RJ-45 interface for an Ethernet LAN instead of an RJ-11 jack. Figure 10 depicts two of the scenarios used to conduct VoIP experiments via Cisco's VoIP solution.

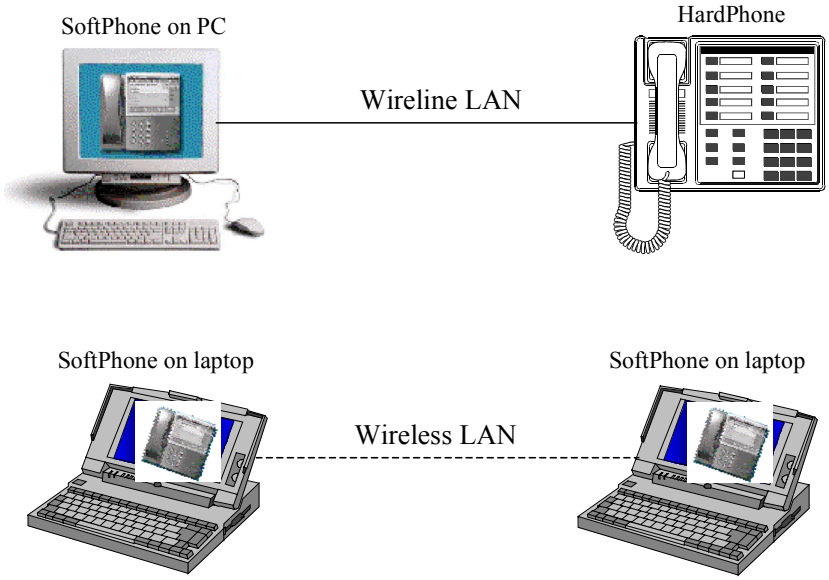


Figure 10: SoftPhone/HardPhone Configuration

This figure hides the behind-the-scenes details, where a VoIP server (given the moniker "CallManager" by Cisco) handles the calls between two or more parties via H.323. There are two vocoders implemented by the Cisco solution: G.711 and G.729.1 [Cisco].

For the experiment itself, laptops with an Intel Pentium II 400 MHz processor and 128 MB of random access memory (RAM) were utilized; the desktop PC was even more powerful, equipped with dual Intel Pentium II-733 MHz processors. For simple VoIP, however, this extra "horsepower" is unnecessary, as long as processor-intensive applications are not running in the background. The OS of choice was Microsoft Windows 2000 (the Cisco solution only runs on Windows 9x, NT 4, and 2000). Those laptops that were connected to the WLAN used 802.11 (2 Mbps); the other PCs and wireline laptops used a more customary Fast Ethernet NIC (100 Mbps). Cisco's HardPhone consists of an Ethernet (10 Mbps) connection.

Voice calls were initiated and received by the various components, and it did not appear that which device called mattered, whether the device was the caller or the callee. What did matter was network traffic load and processor load. The codec used in the Cisco equipment was G.711 (64 kbps). This bandwidth is negligible when compared to (Fast) Ethernet connections and consumes less than four percent of the 802.11 link.

When no other applications were actively running in the background and the LAN is well-behaved, i.e. lightly loaded, the voice quality was very good, close to PSTN quality and better than cellular reception. However, when applications that consumed a sufficient amount of processing were executed, the quality deteriorated rapidly. In normal speech, words were arbitrarily dropped or poorly transmitted by the client device. Once the processor-intensive application was suspended or the priority of the VoIP application was increased via the OS, the voice quality returned to normal.

Loading the network by blasting it with randomly created packets (Spirent Communications SmartBits was employed to generate this traffic) also wreaked havoc on the voice quality. Specifically, once the network reached full utilization, all calls were dropped. After suspending the network traffic generator, calls had to be reestablished and no ill effects were detected. One-way latency was observed around 250 or 300 ms, depending on if the device was on the (Fast) Ethernet LAN or used 802.11, respectively. Incidentally, when QoS was enabled for both the SoftPhone and HardPhone via the switches (Cisco 6500 series) – while SmartBits was on – the torrent of network traffic did not appear to disturb the voice calls. Experiments were not conducted to observe voice quality when manipulating jitter and artificially inflating delay because the necessary equipment was not readily available.

Another experiment using the SoftPhone and HardPhone was conducted over a WAN. These VoIP devices were separated by almost 3,000 miles and were linked via a fractional T1 link. Moreover, the laptop running SoftPhone utilized a wireless NIC and initiated a call to the HardPhone. Clarity was adequate and latency was minor, roughly one-third of one second. Some audio problems that were noticed, such as jitter and echo, could be attributed to a cheap headset tethered to the laptop running SoftPhone.

It can be presumed through other studies that as long as one-way latency is under 300 ms, delay will not impact the quality as much as other factors. Jitter can drastically impact voice quality [Verizon]. Furthermore, as the conducted experiments indicated, dropped packets are not well-tolerated. Coupled with G.711's inability to correct for dropped or error-laden packets, this further deteriorates voice quality.

Because Cisco implements H.323 for call signaling in its VoIP products, they can theoretically communicate with other H.323-comforming applications, such as Microsoft NetMeeting; this compatibility was corroborated with another demonstration. No affect on quality was revealed when comparing a SoftPhone-to-SoftPhone call to a SoftPhone-to-NetMeeting session.

Dialpad VoIP

In another set of experiments, a voice call was made from a PC to a PSTN telephone, utilizing Dialpad's technology. Its service is based on Dialpad's proprietary, but H.323-compliant, Java applet based client technology. Dialpad has created a new architecture called Split-323 (U.S. patent pending) that makes its service scalable purportedly to millions of users – and flexible by accommodating several flavors of H.323 [Dialpad]. The third pictorial of Figure 1 depicts a scenario of how the VoIP client communicates with a PSTN device. In essence, the packetized voice is translated by the gateway to G.711 signaling, which is compatible with PSTN. The reverse occurs for voice sent from the telephone destined for the PC. In all Dialpad experiments, the PC used contained an Intel Pentium II 450 MHz processor and 128 MB of RAM.

Using Dialpad, a call was initially set up from the PC to a standard PSTN telephone. Approximately sixty miles separated the caller and callee. During the three-minute conversation, it was estimated that the round trip latency was approximately one to two seconds. Moreover, the voice quality was unsatisfactory. In addition to the unacceptable length of latency, packets were frequently dropped, causing the speaker on both sides (PC and telephone) to constantly repeat himself. The PC connected to the Internet via 28.8 kbps modem dialup at 11:00 p.m. EST – an off-peak time in terms of worldwide Internet usage. It is evident that VoIP traversing many networks leads to woeful quality of service.

Next, the feasibility of using the Internet for PC-to-PSTN calls was further stretched. From the PC, a conversation commenced to a nearby user on his digital cellular telephone ("cell phone"). Typically, the quality of a cell phone is worse than that of a landline telephone. In keeping with this theme, it was determined that the quality of the conversation was indeed exacerbated when the callee operated his cell phone as opposed to the traditional telephone. A ballpark one-way latency was measured, and it is surprising to note that the per-

ceived latencies were asymmetrical. From the PC, a one-way latency of approximately one-half second was observed. However, audio initiated from the cell phone endured a one-way latency of two seconds.

A possible explanation of this disproportional latency could be on how the voice is packetized from the PC and how the cell phone sends its signals to the cellular towers. In the former example, the voice would have to be packetized into the Dialpad vocoder. Then, the packets would be converted to the G.711 format, by way of a gateway, to prepare it for the PSTN. Next, the signal is routed to the cellular service provider, which converts the signal to its protocol. Finally, the message is eventually forwarded to the cell phone. This appears to be less computationally intensive than the reverse case. Namely, it is more expensive in terms of time to convert the signal from the cell phone to PSTN, then finally to IP.

Another call was established between the PC and a PSTN telephone at 1:30 a.m. EST. This time, almost 400 miles separated the two speakers. The voice quality on the PC side was determined to be much better than that heard by the speaker on the telephone. The voice from the telephone participant was fairly clear, the voice was recognizable, and few packets were dropped. On the other hand, the voice initiated from the PC was of poor quality. Over 10 percent of the packets were dropped, as measured by quantifying the percentage of conversation that was successfully understood. The conspicuous weak point on both ends was the lengthy round trip latency. This intolerable delay was deemed to be three seconds, as measured by producing a sound on one end, then having the other user immediately reply so that the first user could calculate the time passed.

Practical issues regarding VoIP, including its feasibility to compete and possibly replace the traditional telephone system, will be discussed next.

Summary

Certain key topics highlighted herein will be presented. Namely, the trade-off between bit-rate and voice quality will be discussed, followed by re-examining wireless VoIP. The paper will conclude by studying current trends on VoIP and the impact that this technology is expected to make, if it has not already.

Bit-rate vs. Voice Quality

Can VoIP emerge from a specialized application or niche to mainstream voice communication? It is palpable that while VoIP technology may have progressed admirably, as gauged by protocol and vocoder development, it still has plenty of room for improvement, such as:

- Quality of voice transmissions
- Reliability of the Internet
- Standards battles
- Competition/confusion with wireless
- Human factors/usability

Reliability cannot be overemphasized. The PSTN operates with at least 99.999 percent specified availability and is available even during power outages. Neither can be said for today's VoIP, but it must in the near future if VoIP is to gain wide acceptance.

As mentioned previously, the focus of many developers in the industry has been to design vocoders that consume progressively lower bandwidth, as is evident with the creation of MELP and then MBE. This effort may be misguided. Most applications of VoIP rely on connectivity to the Internet; the vast majority of these users have at least a 28.8 kbps connection, if not more. Nonetheless, developers still pursuing ultra-low bandwidth coders instead of improving the quality of low bandwidth coders already in existence. Per-

haps this effort is to allow users to concurrently enjoy other bandwidth-consuming applications, such as browsing the World Wide Web. Even more, many of these algorithms, such as the three ITU-T specifications formerly described, were created for using voice over a reliable circuit-switched connection rather than the packet-based network the Internet utilizes.

Some developers are recognizing this fact by constructing higher quality codecs that consume more bandwidth. They are amenable to trading-off bandwidth to achieve this quality. For example, Global IP Sound's coding maintains a 64 kbps bit-rate, but it improves robustness and reduces delay and complexity [GIPS]. Furthermore, it is designed for the Internet, and has mechanisms to degrade speech quality gracefully when encountering increasing packet loss, while moderate packet loss is trivial. In general, it is critical for the vocoder to tolerate mishandled, dropped, and out-of-order packets – short of a cornucopia of packets loss – intrinsic in UDP. Of equal importance, one-way latency should be confined to one-quarter of one second. Regrettably, the leitmotif of the PC-to-PSTN VoIP experiments was a protracted delay. Finally, the codec should address other QoS issues, such as maintaining an optimally-sized buffer to restrain jitter and echo.

Given these constraints, it is believed that in order for VoIP to gain higher acceptance, the focus should be on high quality instead of low bit-rate, as this will be crucial for the acceptance of IP telephony. Obviously, 64 kbps is too high for users who dialup via analog modem to connect to the Internet, nevertheless, a higher quality codec consuming 24 kbps will be preferable to a low quality codec using 5 kbps. In a corporate and broadband environment, even 64 kbps is just noise in the line when the average user is allotted hundreds, if not thousands, of kilobits per second.

Another possibility could be in developing even higher bandwidth vocoders to allow something that the traditional telephone system can never do: transport high fidelity stereo audio. One possibility would be allowing users to call another VoIP application to listen to high quality, compressed music, such as MP3, which would consume a mere 128 kbps. Of course, there are other issues involved, such as the server's ability to provide music at this fidelity while being able to scale exponentially.

Revisiting Wireless VoIP

From simple experimentation, it is evident that VoIP quality is very acceptable in the Cisco VoIP experiments, whether done via wireline or untethered. In fact, when running VoIP via 802.11 wireless technology, the laptop was brought near an operating microwave oven to see if any interference could be observed. No disruptions or voice quality degradation was noticed. Wireless connectivity, as previously mentioned, is an inescapable trend for the future; this is further hastened by the proliferation of wireless communications-enabled devices. For VoIP to succeed, it must be tolerant of even higher packet losses usually associated with wireless technologies, such as 802.11x and Bluetooth.

For both 802.11b and Bluetooth to succeed, each needs to recognize its strength, and avoid encroaching the other's territory, since they are both on the same 2.4 GHz frequency band. 802.11x is best suited to replace or supplement wireline LANs, hence, it is the superior means of transmission for VoIP to utilize. Bluetooth was created and should be focused as a WPAN. Voice applications, because of their low bandwidth usage, can also operate here, thereby permitting a Bluetooth access point to transport voice to another access point, and ultimately to the Bluetooth client. The sore spot, again, is the implementation of current vocoders' inability to tolerate moderate packet loss, provide a stable QoS, and maintain jitter without requiring an excessive buffering mechanism.

Impact of VoIP

The growth in IP-based services the past few years has been explosive. It is projected that this market will continue to grow at an even higher rate for several years to come. IP telephony is expected to benefit from

this deluge of IP services. There is a paradigm shift beginning to occur since more communications is in digital form and transported via packet networks, such as IP – and data traffic is far out-accelerating traditional voice telephony traffic. While there is more than a century of experience in designing, operating, and managing conventional circuit-switched networks, relatively limited data is available about IP-based networks. The success of VoIP hinges primarily on a clear understanding of the overall technology and service requirements.

Frost & Sullivan prognosticate that the annual growth rate of global IP telephony service will exhibit triple-digits: VoIP products manufactured increase from under four million in 2000 to over one-half billion in 2006 [Guizani, et al.]. Another survey [Feldman] has ascertained that almost half of industry experts anticipate that 15 to 20 percent of total voice traffic will run over data networks, within a two-year timeframe. That number leaps to 91 percent when the time horizon is expanded to three to five years. This suggests that in the immediate future, VoIP usage will be modest. However, within a few years, more business and residential customers will adapt to VoIP as its quality and reliability improves. Eventually (anywhere from the end of this decade to the century's end) circuit-switched telephony will be a memory, regulated to museums alongside the telegraph.

VoIP will impact real-time voice traffic in three different ways.

- Voice trunks can replace the analog or digital circuits that are serving as voice trunks or PSTN-access trunks.
- PC-to-PC voice can be provided for multimedia PCs operating over an IP-based network without connecting to the PSTN, including ubiquitous wireless VoIP access.
- Telephony communications appears as a normal telephone to the caller, but may actually consist of various forms of VoIP, all interconnected to the PSTN.

VoIP networks are already incorporating IP-based PBXs that emulate the functions of a traditional PBX. These allow both standard telephones and multimedia PCs to connect to either the PSTN or the Internet, providing a seamless migration path to VoIP. Moreover, traditionally telephone service can be enhanced, such as combining real-time and non-real-time communications, high fidelity audio, conference calling, and scores of other features.

The [Feldman] study also determined that currently, just 40 percent of those who have used VoIP believe it to be "the same" or "superior" to conventional dialing. However, as long as the underlying technologies of VoIP improve to address the issues presented earlier in this section, VoIP is poised to rocket. This will be further accelerated as IPv4 matures to IPv6 (predominantly due to its built-in QoS support), a more reliable network infrastructure (as the Internet evolves to Internet2), and demand for voice communications via wireless devices. In fact, today's VoIP services are merely a harbinger of the high performance integrated voice, video, and data services that will be available in the not-too-distant future.

References

Anttalainen, Tarmo. *Introduction to Telecommunications Network Engineering*. Artech House, Boston, MA, 1999.

Carden, Philip. *Building Voice over IP*. Network Computing, May 8, 2000.

Cermak, Gregory W. *Measuring Subjective Quality of Speech over packet Networks*. Verizon Laboratories, 2000.

Champness, Angela. *IEEE 802.11 DSSS: The Path to High Speed Wireless Data Networking*. `http://www.wirelessethernet.org/whitepapers.asp`.

Cisco Telephony. `http://www.cisco.com`, Cisco Systems, Inc., 2000.

Cisco Voice Applications. `http://www.cisco.com`, Cisco Systems, Inc., 2000.

Coffee, Peter. *Bluetooth Goal: Civilize Wireless*. eWeek, October 30, 2000.

Conover, Joel. *Anatomy of IEEE 802.11b Wireless*. Network Computing, August 7, 2000.

Dialpad Technology. `http://www.dialpad.com/company/technology.html`, Dialpad.com, Inc., 2000.

deCarmo, Linden. *The Media Gateway Control Protocol: A Simpler and More Reliable Voice over the Internet*. Dr. Dobb's Journal, May 2000.

Douskalis, Bill. IP Telephony: *The Integration of Robust VoIP Services*. Prentice Hall, Upper Saddle River, NJ, 2000.

Flood, Kevin. *Staging a Comeback: Emerging Standards Position Wireless Networks for Broad Appeal*. Global Knowledge, October 18, 1999.

GIPS Speech Coding and Speech Quality in IP Telephony. Global IP Sound, Stockholm, Sweden.

Guizani, Moshen, Rayes, Ammar, and Atiquzzaman, Mohammed. *Internet Telephony*. IEEE Communications, April 2000.

Handley, M., Schulzrinne, H., Schooler, E., and Rosenberg, J. *SIP: Session Initiation Protocol*. RFC 2543, The Internet Society, March 1999.

Haartsen, Jaap C. and Zürbes, Stefan. *Bluetooth Voice and Data Performance in 802.11 DS WLAN Environment*. Ericsson, May 31, 1999.

Hersent, Oliver, Gurle, David, and Petit, Jean-Pierre. *IP Telephony: Packet-Based Multimedia Communications Systems*. Addison-Wesley, Harlow, England, 2000.

IEEE 802.11b Wireless LANs: Wireless Freedom at Ethernet Speeds. 3Com Corp., 2000.

Introduction: Quality of Service Overview. Cisco Systems, Inc., 2000.

IP Security – IPSec: Overview. Cisco Systems, Inc., 1998.

ITU-T. *G.711: Pulse Code Modulation (PCM) of Voice Frequencies*. International Telecommunication Union, November 1988.

ITU-T. *G.723.1: Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s*. International Telecommunication Union, March 1996.

ITU-T. *G.729: Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-code-Excited Linear-Prediction (CS-ACELP)*. International Telecommunication Union, March 1996.

Keshav, S. *An Engineering Approach to Computer Networking*. Addison-Wesley, Reading, MA, 1997.

Lewis, Chris. *Transporting Encapsulated Voice Traffic*. Network Computing, November 29, 1999.

Leveraging the Intelligence of SS7 to Improve IP-Based Remote Access and Other IP Services. 3Com Corp., May 19, 1999.

Low-Rate Speech Coding. `http://www.causalproductions.com/TEMP/INDEX/IC96S106.HTM`.

Machi, Jim. *For Voice Coders, Economics Matter More than Compression Rate*. Technology Marketing Corp., November 1999.

Mannion, Patrick. *Motorola's V.92 Card Takes Bluetooth on New Path*. Electronic Engineering Times, November 20, 2000.

Moore, Cathleen. *Faster Wireless LANs on Tap*. InfoWorld Media Group, Inc., September 15, 2000.

Nijhawan, Vinit. *Wireless Connectivity for Mobile PCs*. Circuit Cellular, March 2000.

Pachomiski, Jason. *TechRepublic's TCP/IP Primer*. TechRepublic, 2000.

Peterson, Larry L. and Davie, Bruce S. *Computer Networks: A Systems Approach*. Morgan Kaufman Publishers, Inc., San Francisco, CA, 1996.

Percy Alan. *Understanding Latency in IP Telephony*. Brooktrout Technology, February 1999.

Ploskina, Brian. *Climbing Aboard the SIP Bandwagon*. Inter@active Week, September 25, 1999.

QoS Features for Voice. Cisco Systems, Inc., 2000.

Raikar, Amit. *Voice over IP Networks (VoIP)*. `http://www.umr.edu/~araikar/courses/cs285/VoIP.html`.

Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V. *RTP: A Transport Protocol for Real-Time Applications*. RFC 1889, January 1996.

Speech Coding Solutions for IP Telephony. Global IP Sound, Stockholm, Sweden.

TechRepublic's Telecom and Data Networking Glossary. TechRepublic.

Voice over IP (VoIP). `http://www.tothai.net/Thai/download/White%20Paper/voip.htm`.

VoIP Survey. Feldman Communications, Inc., Annapolis, MD, November 27, 2000.

Wells, Jim. *Will Bluetooth Make Your Network Security Toothless?*. TechRepublic, September 18, 2000.

Willis, David. *The Future is SIP*. Network Computing, September 20, 1999.

Woods, Darrin. *Connecting to the Voice World*. Network Computing, April 17, 2000.

Woods, Darrin. *Translating Menus at the VoIP Café*. Network Computing, December 27, 1999.

Yanowitz, Jason. *Under the Hood of the Internet: An Overview of the TCP/IP Protocol Suite*. ACM Crossroads, January 20, 2000.

Young, Larry. *New Protocol Connects Voice over IP*. Network World Fusion, December 13, 1999.

Zyren, Jim. *Reliability of IEEE 802.11 Hi Rate DSSS WLANs in a High Density Bluetooth Environment*. Intersil Corp., June 8, 1999.

Zyren, Jim and Petrick, Al. *IEEE 802.11 Tutorial*. `http://www.wirelessethernet.org/whitepapers.asp`.