Voice Over IP

A Comprehensive Study

By
Anoop Vetteth
Dilip Antony Joseph (CS00294)
Vincent

Voice Over IP – A Comprehensive Study

Anoop Vetteth, Dilip Antony Joseph (CS00294), Vincent

1 Introduction 5
2 What is VoIP? 5
3 Why VoIP? 6
4 Can you make free calls? 6
5 Major Concerns 6
6 VoIP – The Technology 6
6.1 Making a PSTN Call 7
6.2 Making a VoIP call 7
6.3 VoIP Components 8
6.3.1 Codecs 8
6.3.2 VoIP Protocols 8
6.3.3 IP Telephony Servers and PBXs 9
6.3.4 VoIP Gateways 9
6.3.5 IP Phones and Softphones 9
7 Strategies for Setting up VoIP 10
7.1 Setting up a simple Toll Bypass 11
7.2 Setting up a Total IP Telephony network 13
7.3 IP enabled PBXs 13
7.4 Direct PC to PC Communication 14
8 Audio/Video Coding 14
8.1 Video Coding 15
8.1.1 H.263 Video Coding 15
8.2 Audio Coding 18
8.2.1 G.711 19
8.2.2 Linear Predictive Coding (LPC) 20
8.2.3 G.728 21
9 The H.323 Suite of Protocols 22
9.1 H.323 Versions 22
9.2 H.323 Components 23
9.2.1 Terminals 23
9.2.2 Gateways 23
9.2.3 Gatekeepers 24
9.2.4 Multipoint Control Units 24
9.2.5 H.323 Zone 25
9.3 H.323 Protocols 25
9.3.1 H.225 Registration, Admission and Status (RAS) 26
9.3.2 H.225 Call Signaling 26
9.3.3 H.245 Control Signaling 27
9.3.4 Real-Time Transport Protocol 27
9.3.5 Real-Time Transport Control Protocol 27
9.4 Making a H.323 Call 28

10 Session Initiation Protocol (SIP) 29
10.1 Functions of SIP 29
10.2 SIP in Action 30
10.3 Registration 30
10.4 Structure of SIP 30
10.5 Session Description Protocol (SDP) 31
11 H.323 vs. SIP – What should we use? 31
11.1 Differences between H.323 and SIP 31
11.2 What should we use? 32
12 Media Gateway Control Protocol - MGCP 32
12.1 What is MGCP? 32
12.2 The Working of MGCP 33
12.3 MGCP Commands 34
13 Quality of Service in VoIP networks 35
13.1 Why is QoS required? 35
13.2 DiffServ 35
13.3 MPLS 35
13.4 RSVP 36
13.5 IP Phone configuration 36
13.6 Other Measures to achieve QoS 37
14 Current Status of VoIP 37
15 Conclusion 37
16 References 37
17 Glossary 38

Table of Figures

FIGURE 1 PSTN BASED TELEPHONE SYSTEM (LEFT) AND A VOIP BASED SYSTEM (RIGHT). THE PSTN AND VOIP ARE INTERCONNECTED USING A GATEWAY. ..... **ERROR! BOOKMARK NOT DEFINED.**
FIGURE 2 VOIP COMMUNICATION BETWEEN DIFFERENT KINDS OF ENDPOINTS ..... **ERROR! BOOKMARK NOT DEFINED.**
FIGURE 3 VOIP PROTOCOL STACK. THIS FIGURE SHOWS WHERE THE VARIOUS VOIP PROTOCOLS FIT IN. THE PROTOCOLS ARE DESCRIBED IN THIS TERM PAPER. 10
FIGURE 4 H.323 ENABLES COMMUNICATION BETWEEN 2 H.323 ENDPOINTS OVER A PACKET BASED NETWORK SUCH AS AN IP NETWORK. 22
FIGURE 5 A GATEWAY CONNECTS AN IP BASED BASED H.323 NETWORK (LEFT) WITH A SWITCHED CIRCUIT NETWORK (RIGHT) 24
FIGURE 6 A H.323 ZONE 25
FIGURE 7 THE SHADED BOXES REPRESENT THE VARIOUS PROTOCOLS UNDER H.323 26

ABSTRACT

Voice over IP or ***VoIP*** is a fast emerging technology which provides cost effective telephony solutions. As against traditional PSTN networks, VoIP converts voice into suitably encoded digital packets and transmits those across existing IP based networks like the internet. Apart from simple voice-based telephone services, VoIP can provide advanced functionality like unified multimedia messaging, real-time data transfer and multiparty conferencing.

To a user, a VoIP call does not appear to be much different from a normal PSTN call. VoIP depends on many components – audio/video codecs, VoIP protocols, IP servers, gateways and of course, the IP phones, to make this possible. G.711 and G.729 are popular audio codecs while H.323 and SIP are common protocols. All the various components of a VoIP system are described in detail in this term paper.

QoS or Quality of Service is an integral requirement of any VoIP system. The various methods to achieve QoS like DiffServ, MPLS, RSVP and optimal IP phone configuration are studied in detail. There exist different strategies for deploying VoIP in an organization. Depending on the resource constraints and user requirements, organizations can go in for simple toll bypasses, IP enabled Private Branch Exchanges or even a total IP Telephony solution. Readily available software like NetMeeting enable direct multiparty conferences between computers.

VoIP is being used on a large scale today, mainly due to the large cost savings it offers. An often asked question is: *Can I make free phone calls using VoIP?*

The answer is **YES**!

1 Introduction

In today's world, computer networks have changed the way in which people communicate. The internet, which began as a US defense project, is now a ubiquitous household entity. Applications like the World Wide Web and email have drastically cut down the delays in global communication and information gathering. However, the Public Switched Telephone Networks (PSTN) continue to dominate the area of voice communication.

Data to be transferred over the internet is split into multiple packets, each of which is transmitted independently. The packets are reassembled in the correct order at the destination. IP (Internet Protocol) is the technology which makes this possible. Voice is just another kind of data. Voice too can be packetized and transmitted using IP. This is where VoIP comes in.

2 What is VoIP?

VoIP stands for Voice over Internet Protocol. VoIP describes a set of facilities for managing the delivery of voice information using IP. In other words, VoIP is the ability to make telephone calls and send faxes over IP-based data networks with a suitable quality of service (QoS) and superior cost/benefit. This can also be called IP Telephony.

VoIP is very different from PSTN (Public Switched Telephone Network). When an "ordinary" PSTN call is made, a virtual switched circuit is set up between the source and the destination for the duration of the call. No such circuit is established in the case of VoIP. The analog voice signal, suitably encoded in a digital form, is split into various packets. Each packet is independently transmitted across private networks or the public internet using IP, and reassembled at the destination. Thus VoIP is not, in any way, about re-engineering the PSTN. It can be considered to be a direct competitor to the PSTN networks in the telephone services industry.

The origins of VoIP can be traced back to a simple shareware program for simple PC to PC voice chat over the internet, released in 1994. Today the scope of VoIP is not limited to transmitting only audio signals between two end-points. It supports multiparty conferencing as well as real-time video, data and fax transmissions.

Although IP networks are prevalent today, implementing VoIP on them requires further effort. There are many issues to be addressed and problems to be overcome. This term paper aims to describe the various issues involved in the implementation of VoIP and to study the various protocols and procedures developed to tackle them.

3 Why VoIP?

PSTN telephones have been in operation for more than a hundred years. VoIP must offer significant advantages over PSTN if it is to be accepted globally. Cost reduction is the driving force behind the adoption of VoIP. For example, international calls can be routed to its destination country, through the public internet; thus, bypassing toll regimes and saving on PSTN bills.

VoIP can help companies in reducing overall equipment costs by integrating the voice and data networks. Multiparty conferencing and real-time multimedia and data transfer are some other services easily provided by VoIP. These factors have been fuelling the development and deployment of VoIP in various parts of the world today.

4 Can you make free calls?

Does there exist something like a free VoIP call?
The answer is YES. VoIP uses the public internet to transport packets. As against that in PSTN, distance is not the criterion for billing in the public VoIP. Therefore, VoIP calls can be made for free, using the existing infrastructure of the internet (access to the internet is assumed). However, there are caveats. VoIP often doesn't guarantee the same Quality of Service as the traditional PSTN. Voice quality in VoIP may be degraded. Connection setup times may be more than that in the PSTN calls.

5 Major Concerns

The quality of voice received by an endpoint is a major concern in VoIP. Poor voice quality is attributed to the unreliability inherent in packet based IP networks. The packets carrying voice data can get dropped by the network or may arrive out of order. Various protocols like RTP and RTCP have been devised to decrease this problem.

PSTN will continue to be a major player in the telephone industry. Hence, VoIP should be able to seamlessly interoperate between PSTN as well as with other existing multimedia transmission networks. As the voice data is transmitted over the public internet, privacy of the call is an issue in VoIP. VoIP must provide methods to encrypt the call, as well as do authentication and admission control of users.

VoIP involves a large number of co-operating processes. Various strategies and technologies have been developed to address these concerns. This paper analyzes the various concerns and their solutions in the following sections.

6 VoIP – The Technology

6.1 Making a PSTN Call

Knowledge about the working of PSTN helps in understanding VoIP. Making an ordinary telephone call involves a large number of steps. These basic steps in a PSTN call are enumerated below

1. The caller picks up the telephone handset hears a dial tone.
2. A telephone number is entered, specifying the address of the callee.
3. Signals are sent through the PSTN to set up a circuit for the call. Capacity and bandwidth are reserved for the call.
4. The destination phone rings, indicating to the callee that a call has arrived.
5. The callee picks up the telephone handset and begins a conversation. The audio, voice conversation is translated to digital format in the center of the network, and then back to analog at the edge.
6. The conversation ends, call billing occurs, the circuit is taken down, and resources are released.

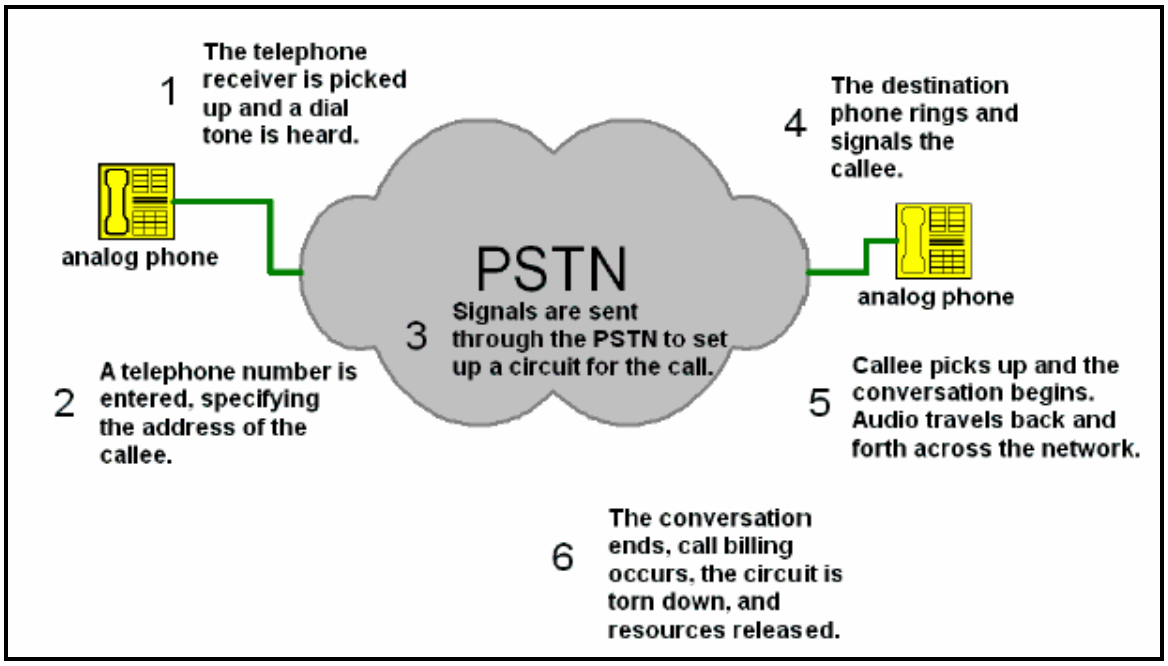


Figure 1 The six steps in a typical telephone call

6.2 Making a VoIP call

To a user, making a VoIP call is no different from making a PSTN call. However in a VoIP call, the call setup portion of the calling sequence has to be simulated—dial tone, ringing, busy signals. The audio portion of the call itself needs to be converted from analog to digital, cut into packets, sent across the network still in packet format,

reassembled, and converted from digital back to analog. Codecs at either end do the conversion from analog to digital and back.

Here's what happens when a call is made using VoIP:

1. The caller picks up the telephone handset and hears a dial tone.
2. The caller enters a telephone number, which will be mapped to the IP address of the callee.
3. Call setup protocols are invoked to locate the callee and send a signal to produce a ring.
4. The destination phone rings, indicating to the callee that a call has arrived.
5. The callee picks up the telephone handset and begins a two-way conversation. The audio transmission is encoded using a codec and travels over the IP network using a voice streaming protocol.
6. The conversation ends, call teardown occurs, and billing is performed.

6.3 VoIP Components

VoIP involves transferring voice data in packet form over IP networks. However, as seen above, the user should not notice any difference between VoIP call and a PSTN call, except possibly for the advanced features offered by VoIP. The following are the typical components which are needed to achieve this in a VoIP system.

6.3.1 Codecs

A codec (which stands for "compressor/decompressor" or "coder/decoder") is the hardware or software that samples analog sound and converts it to digital bits, which it outputs at a predetermined data rate. The codec often performs compression as well, to save bandwidth. There are dozens of available codecs, each with its own characteristics.

G.711, G.726 and G.729 are examples of popular codecs. In the VoIP systems, these codecs are used to convert the user's voice into digital signals for transmission across the network. Codecs are discussed in more detail in the following section.

6.3.2 VoIP Protocols

Protocols are required during the various phases of a VoIP call. These protocols define the interactions between the various VoIP terminals. For example, protocols are required to set up and tear down calls between two terminals. After a call has been set up, other protocols are needed to transmit the real-time voice data across the network.

The protocols H.323, SIP (Session Initiation Protocol) and MGCP (Media Gateway Control Protocol) used in call setup will be described in the following sections. RTP

(Real-time Transport Protocol) and RTCP (Real-time Transport Control Protocol) used in transmitting the voice data to the receiver is also discussed.

6.3.3 IP Telephony Servers and PBXs

In VoIP, IP Telephony Servers and PBXs handle the services (for example – call forwarding, call transfer) provided by the PBX in the traditional PSTN. In addition to these basic functions, the IP PBX can provide services like a unified messaging service – convergence of voice mail and email. An IP telephony server can also function as a video conferencing server allowing multiparty conferences. They can also take the role of Gatekeepers (as in H.323 protocol) which provide the address conversion and admission controls facilities.

While the traditional PBX is often a closed-box proprietary system, the IP PBX can be implemented on a computer. These IP PBXs can be managed through a standard Application Programming Interface (API). IP PBXs are often clustered to provide increased scalability and reliability.

6.3.4 VoIP Gateways

VoIP gateways are an important part of VoIP system. Gateways connect the VoIP system to the traditional PSTN or to any other non-VoIP network. The gateway must be able to handle the signaling protocols of both the VoIP and PSTN while making a call. For example, a gateway uses the H.323 protocol while talking to a VoIP terminal on one side of a call and uses SS7 (Signaling System 7) while talking to the PSTN phone on the other side. Gateways also perform *transcoding*, i.e. conversion from one codec to another, if the codecs used on either sides of the gateway differ.

6.3.5 IP Phones and Softphones

Traditional phones are analog. The voice signals are converted to digital signals at the IP PBX before forwarding on the network. A new breed of phones which implement the codecs themselves are called IP phones. These IP phones convert the voice into digital form and then contact a IP telephony server for call processing. These phones have an Ethernet connection rather than the standard 4-line telephone connector.

The personal computer can also easily take the role of an IP phone. The computer's CPU can process the voice signals received from the headset and microphone plugged into the audio card. The computer's LAN connection is used for the signaling and data transfer.

The next sections deal with the various codecs and protocols used in implementing VoIP. The following figure shows the complete VoIP protocol stack

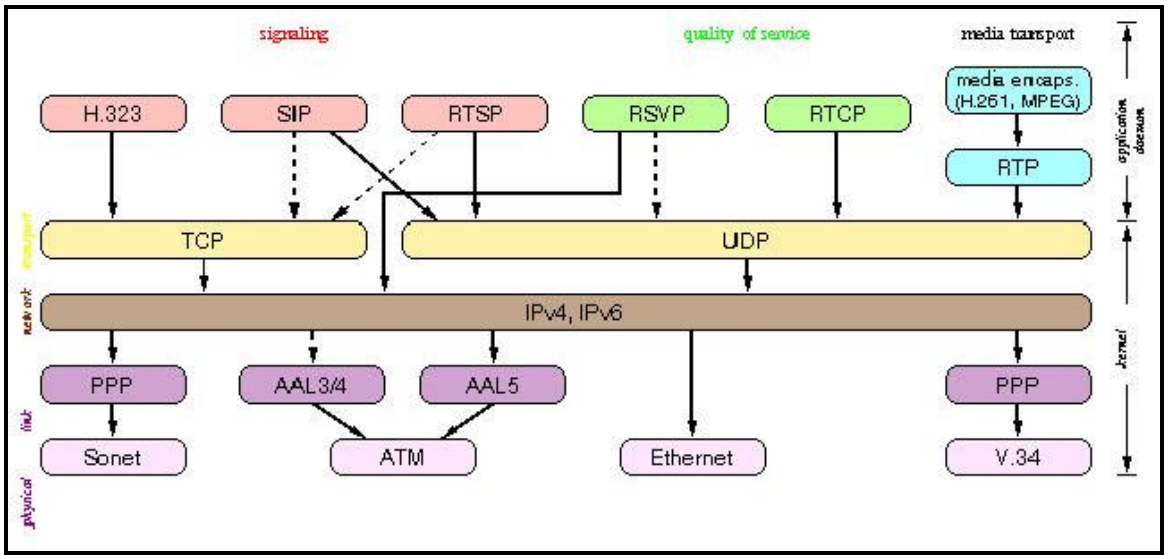


Figure 2 VoIP protocol stack. This figure shows where the various VoIP protocols fit in. The protocols are described in this term paper.

There are two competing sets of protocols for VoIP – *H.323* and *SIP*. MGCP is the most popular protocol used to control VoIP gateways. These protocol sets completely describe the working of VoIP. RTP and RTCP are used to transmit the voice between the terminals. All these protocols will be described in detail in the following sections. Common audio/video codecs are used for both H.323 and SIP. The details about the audio and video encoding are presented in the next section.

7 Strategies for Setting up VoIP

VoIP enables communication between a variety of endpoints. Software like MS *NetMeeting* is used for PC to PC communication. Web sites like *net2phone.com* offer PC to Phone services. Direct phone to phone as well as phone to PC services can be provided by VoIP.

Different strategies for setting up VoIP are adopted on the basis of the user requirements and the telephone system currently in usage. Various strategies for implementing VoIP are discussed in this section.

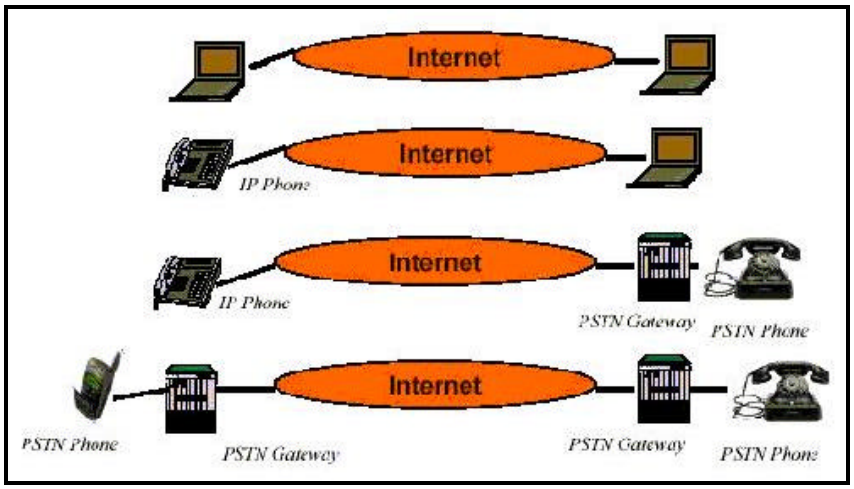


Figure 3 Various modes of using VoIP

7.1 Setting up a simple Toll Bypass

The communication budgets of companies are often dominated by the long distance calls made. Toll bypass is a simple method to reduce this cost by a large amount.

In the existing telephone networks, two offices of a company that wish to talk with each other directly often use leased lines to connect their respective PBX. These leased lines are expensive and have to be paid for by the company, irrespective of their usage. In a simple toll bypass scheme, the existing tie lines are replaced by connections to an IP based network. The public internet is most often used for this purpose. The simplest way to do this to simply unplug the existing PBXs and connect them to a VoIP *Relay* that converts that converts the converts the call signaling and voice data to a form suitable for transmission across an IP network. The VoIP relay which directly connects to a router on the IP network may be a standalone device or may be directly implemented in the PBX itself. The following figures show an existing PSTN based network and the VoIP toll bypass set up on the network.

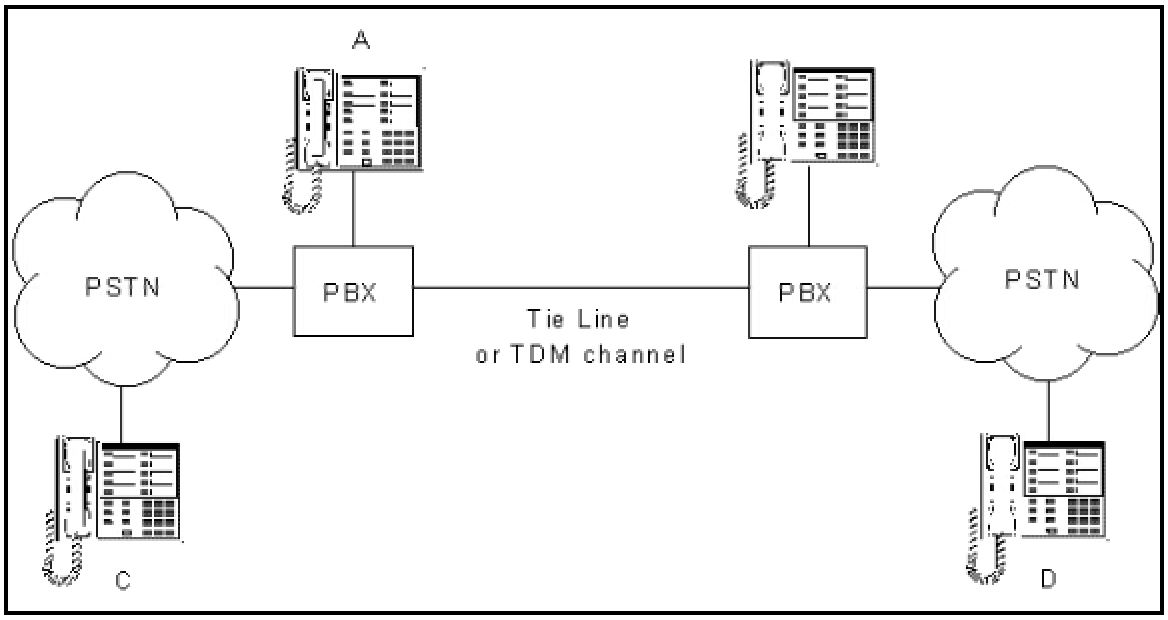


Figure 4 A corporate telephone network with a tie line connecting 2 remote offices

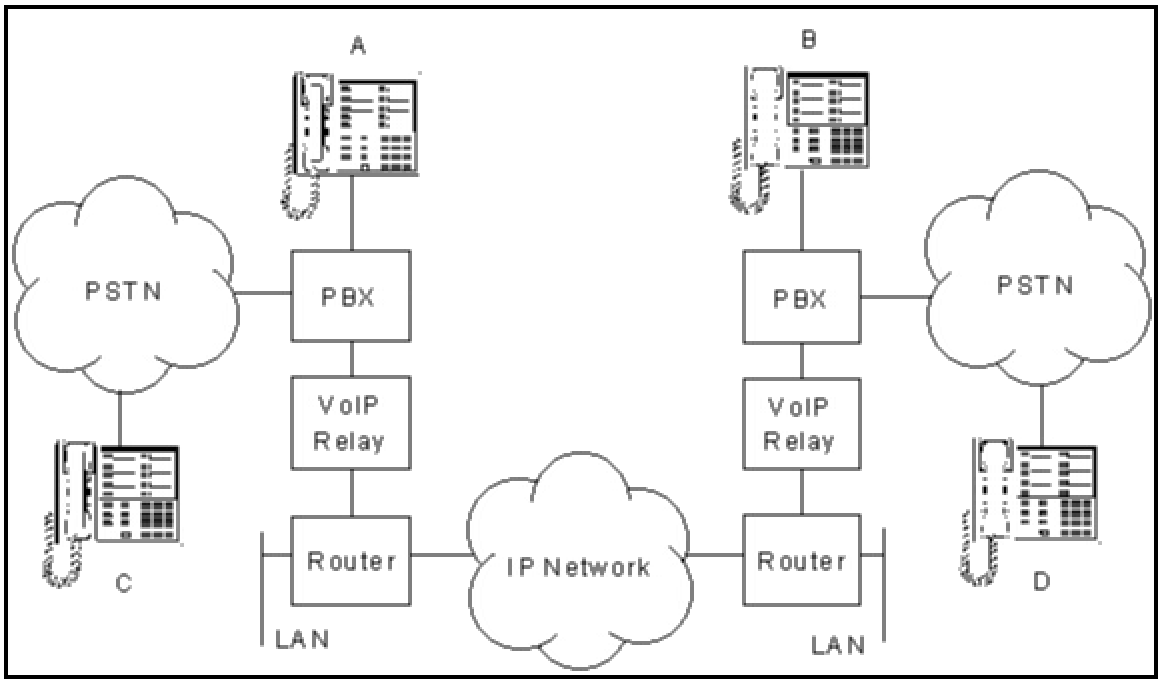


Figure 5 A simple VoIP based Toll Bypass

There are three main issues in setting up a VoIP based toll bypass. Firstly, the VoIP relay should be able to handle sufficient signaling information to support the features in use on PBXs. Secondly, the network elements used to set up the VoIP relay should confirm to the accepted standards like H.323 or SIP. The last issue is concerned with the encoding scheme used and the ability of the IP network and the VoIP relay to provide the desired

level of QoS. Setting up a simple bypass in an existing PSTN based telephone system is the easiest way to make VoIP work for you.

7.2 Setting up a Total IP Telephony network

In a total IP telephony approach, all end-user devices are connected directly to the network via LAN connections. The end-user devices can be of two types – hardware IP phones or PC based softphones. These devices can communicate with each other directly using IP and RTP. Communication with the PSTN phones is enabled through a VoIP gateway which handles the different signaling and voice data formats. IP telephony servers (for example, the H.323 gatekeeper) or clusters of servers, provide the basic call set up functions as well as advanced functions like call routing, E.164 telephone number to IP address mapping, billing etc. The following figure shows a complete IP telephony based system.

A total IP telephony solution involves replacement of the almost the entire existing telephone equipment. This is a very costly approach. It is also not very easy to set up all the components in the correct way. Hence the intermediate approach described next can be used.

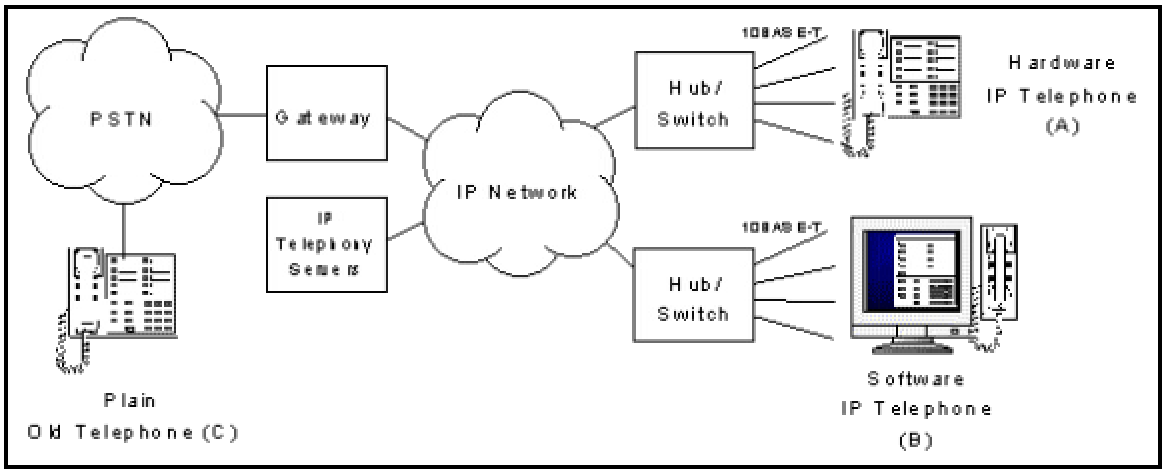


Figure 6 Total IP Telephony - IP phones and softphones directly plug into the Ethernet. The VoIP is network is connected to the PSTB through a gateway.

7.3 IP enabled PBXs

IP enabled Private Branch Exchanges or PBXs provide an intermediate route for VoIP deployment. The existing cabling and handsets need not be changed. Only the PBXs have to upgraded so that they are able to understand IP protocols and are connected to an IP network like the internet. The phones connected to the PBX will appear like IP

phones to the outside world. The PBX can communicate with the external PSTN world through a gateway interface.

Many PBX vendors offer upgrades which make a PBX IP enabled. However pure IP PBXs also known as IPBXs are also available. Thus these PBXs provide an intermediate route to VoIP deployment.

7.4 Direct PC to PC Communication

VoIP can be used for direct PC to PC connections. A total VoIP based system enables this direct communication. However software like the H.323 client NetMeeting allow direct PC to PC based voice and multimedia communication. NetMeeting is not limited to single party connections but can support multiparty conferences as well. Many modern operating systems come with various protocol (for example H.323) implementations. Thus, direct PC to PC communication can be achieved through software.

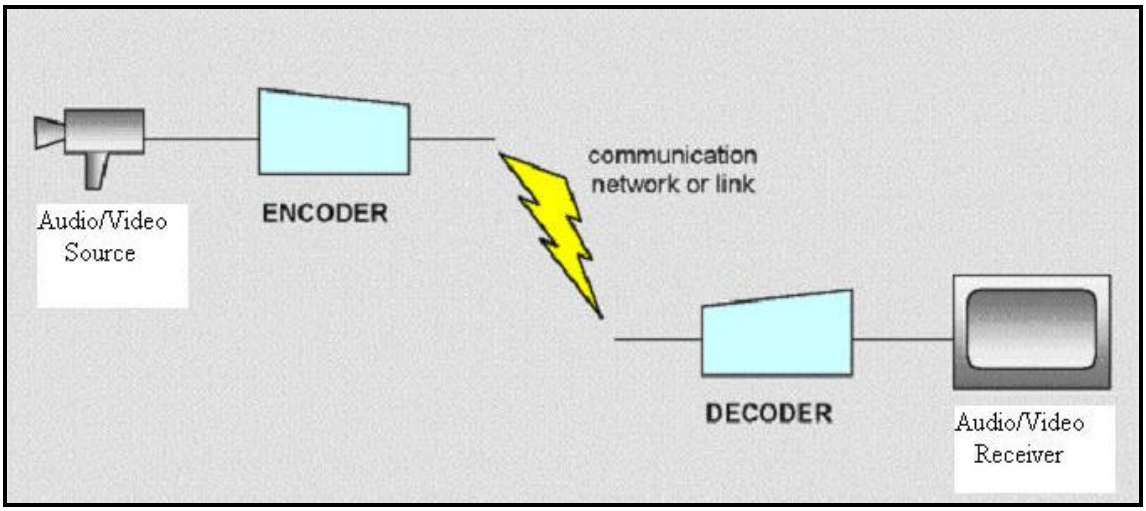
8 Audio/Video Coding

Audio/Video telephony has a wide range of applications including:

- desktop and room-based conferencing
- audio/video over the Internet and over telephone lines
- surveillance and monitoring
- telemedicine (medical consultation and diagnosis at a distance)
- computer-based training and education

In each case video and audio information are transmitted over telecommunications links, including networks, telephone lines, ISDN and radio. The transmitted information has a high "bandwidth" (i.e. many bytes of information per second) and so these applications require **video compression** or **video coding** technology to reduce the bandwidth before transmission.

A typical system is shown in the following Figure:



8.1 Video Coding

Frames of video information are captured at the source and are encoded (compressed) by a video encoder. The compressed "stream" is transmitted across a network or telecommunications link and decoded (decompressed) by a video decoder. The decoded frames can then be displayed. There are several types of video codecs like H.261and H.263 that compress and decompress media streams. In this term paper we shall consider only that codec that is most commonly used for VOIP i.e. the H.263 video coding.

8.1.1 H.263 Video Coding

H.263 is the ITU-T standard and was published sometime in 1995/1996. It was designed for low bit-rate communication and early drafts specified data-rates less than 64 Kbits/s; however this limitation has now been removed. It is expected that the standard will be used for a wide range of bit-rates, not just low bit-rate applications.

A number of video coding standards exist, each of which is designed for a particular type of application: for example, JPEG for still images, MPEG2 for digital television and H.261 for ISDN video conferencing. H.263 is aimed particularly at video coding for low bit rates (typically 20-30kbps and above).

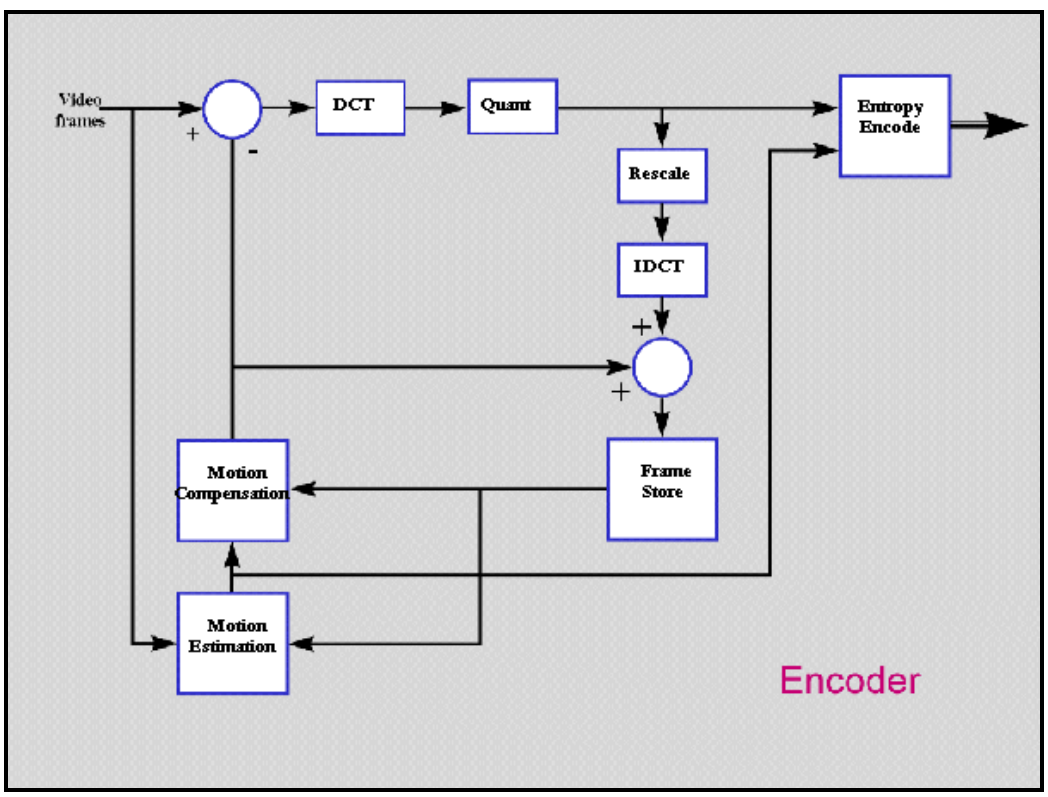
H.263 supports five resolutions. In addition to QCIF and CIF that were supported by H.261 there is SQCIF, 4CIF, and 16CIF. SQCIF is approximately half the resolution of QCIF. 4CIF and 16CIF are 4 and 16 times the resolution of CIF respectively. The support of 4CIF and 16CIF means the codec could then compete with other higher bitrate video coding standards such as the MPEG standards.

Picture Formats Supported					
Picture	**Luminance**	**Luminance**	**H.261**	**H.263**	**Uncompressed**

format	pixels	lines	support	support	bitrate (Mbit/s)			
					10 frames/s		30 frames/s	
					Grey	Colour	Grey	Colour
SQCIF	128	96		Yes	1.0	1.5	3.0	4.4
QCIF	176	144	Yes	Yes	2.0	3.0	6.1	9.1
CIF	352	288	Optional	Optional	8.1	12.2	24.3	36.5
4CIF	704	576		Optional	32.4	48.7	97.3	146.0
16CIF	1408	1152		Optional	129.8	194.6	389.3	583.9

The H.263 standard specifies the requirements for a video encoder and decoder. It does not describe the encoder or decoder themselves: instead, it specifies the format and content of the encoded (compressed) stream. A typical encoder and decoder are described in the following sections.

8.1.1.1 H.263 Encoder



The first step in reducing the bandwidth is to subtract the previous transmitted frame from the current frame so that only the difference or residue needs to be encoded and transmitted. This means that areas of the frame that do not change (for example the background) are not encoded. Further reduction is achieved by attempting to estimate

where areas of the previous frame have moved to in the current frame (**motion estimation**) and compensating for this movement (**motion compensation**). The motion estimation module compares each 16x16 pixel block (**macroblock**) in the current frame with its surrounding area in the previous frame and attempts to find a match. The matching area is moved into the current macroblock position by the motion compensator module. The motion compensated macroblock is subtracted from the current macroblock. If the motion estimation and compensation process is efficient, the remaining "residual" macroblock should contain only a small amount of information.

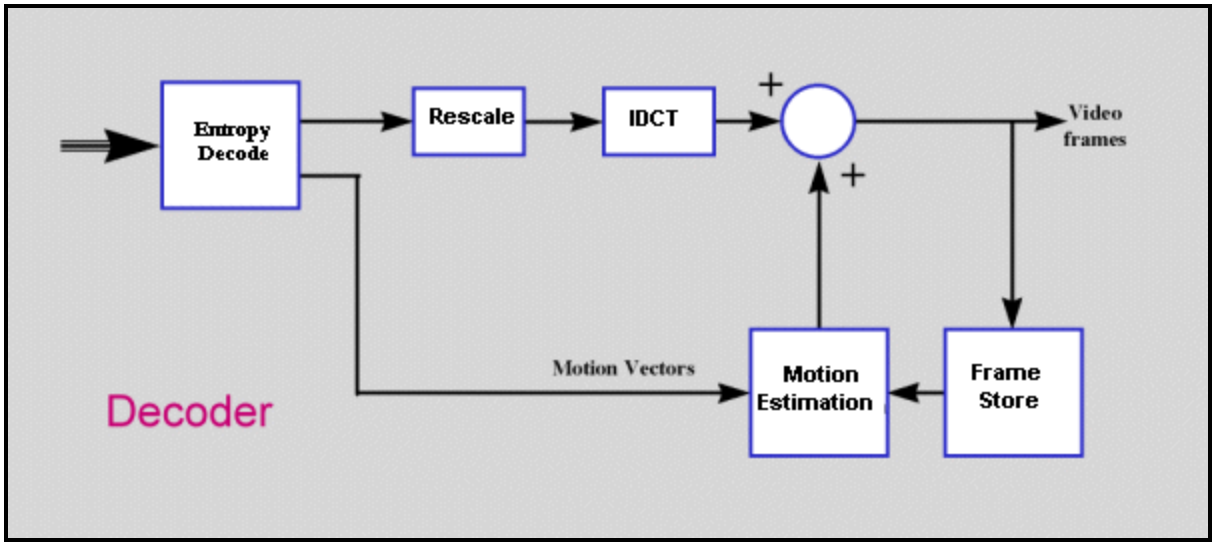
The Discrete Cosine Transform (DCT) transforms a block of pixel values (or residual values) into a set of "spatial frequency" coefficients. This is analogous to transforming a time domain signal into a frequency domain signal using a Fast Fourier Transform. The DCT operates on a 2-dimensional block of pixels (rather than on a 1-dimensional signal) and is particularly good at "compacting" the energy in the block of values into a small number of coefficients. This means that only a few DCT coefficients are required to recreate a recognizable copy of the original block of pixels.

For a typical block of pixels, most of the coefficients produced by the DCT are close to zero. The quantizer module reduces the precision of each coefficient so that the near-zero coefficients are set to zero and only a few significant non-zero coefficients are left. This is done in practice by dividing each coefficient by an integer scale factor and truncating the result. It is important to realize that the quantizer "throws away" information.

An entropy encoder replaces frequently-occurring values with short binary codes and replaces infrequently-occurring values with longer binary codes. The entropy encoding in H.263 is based on this technique and is used to compress the quantized DCT coefficients. The result is a sequence of variable-length binary codes. These codes are combined with synchronization and control information (such as the motion "vectors" required to reconstruct the motion-compensated reference frame) to form the encoded H.263 bit-stream.

The current frame must be stored so that it can be used as a reference when the next frame is encoded. Instead of simply copying the current frame into a store, the quantized coefficients are re-scaled; inverse transformed using an Inverse Discrete Cosine Transform and added to the motion-compensated reference block to create a reconstructed frame that is placed in a store called the **frame store**. This ensures that the contents of the frame store in the encoder are identical to the contents of the frame store in the decoder. When the next frame is encoded, the motion estimator uses the contents of this frame store to determine the best matching area for motion compensation.

8.1.1.2 H.263 Decoder



The variable-length codes that make up the H.263 bit-stream are decoded in order to extract the coefficient values and motion vector information.

This is the "reverse" of quantization: the coefficients are multiplied by the same scaling factor that was used in the quantizer. However, because the quantizer discarded the fractional remainder, the rescaled coefficients are not identical to the original coefficients.

The IDCT reverses the DCT operation to create a block of samples: these (typically) correspond to the difference values that were produced by the motion compensator in the encoder.

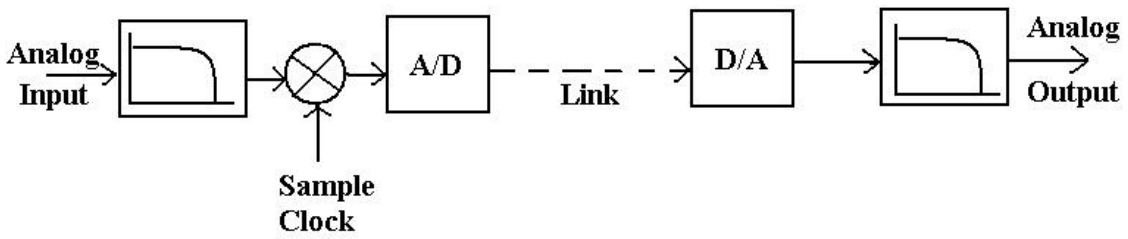
The difference values are added to a reconstructed area from the previous frame. The motion vector information is used to pick the correct area (the same reference area that was used in the encoder). The result is a reconstruction of the original frame: note that this will not be identical to the original because of the "lossy" quantization stage, i.e. the image quality will be poorer than the original. The reconstructed frame is placed in a frame store and it is used to motion-compensate the next received frame.

8.2 Audio Coding

Several types of video coding are available but each technique is used for specific purposes. Some of the commonly used audio codecs are the G.711, G.723.1 and G.728. In this term paper we shall briefly describe a basic coding technique i.e. the G.711 and also an advanced coding technique, the G.728. Both these work on entirely different principles and both can be used for efficient audio coding and compression.

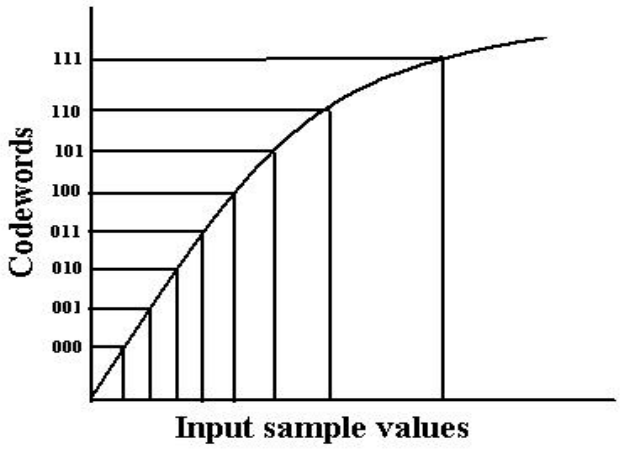
8.2.1 G.711

G.711 is the international standard for encoding telephone audio on an 64 kbps channel. It is a pulse code modulation (PCM) scheme operating at a 8 kHz sample rate, with 8 bits per sample. According to the Nyquist theorem, which states that a signal must be sampled at twice its highest frequency component, G.711 can encode frequencies between 0 and 4 kHz. The figure below is a self explanatory block diagram of the PCM scheme.

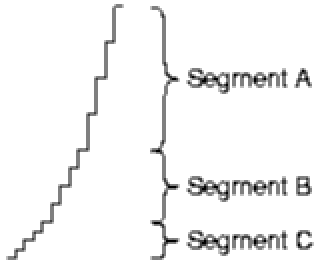


An encoder that uses equal length quantization intervals for all samples is called uniformly encoded PCM. But this scheme treats both the low and high amplitude signals equally. This leads to loss of information in low amplitude signals and hence the Signal to Noise Ratio is high.

This led to the introduction of a new scheme called companding which in turn led to two different variants of G.711: A-law and mu-law. A-law is the standard for international circuits. Each of these encoding schemes is designed in a roughly logarithmic fashion. Lower signal values are encoded using more bits; higher signal values require fewer bits. This ensures that low amplitude signals will be well represented, while maintaining enough range to encode high amplitudes. The figure below shows how companding is achieved.



The actual encoding however doesn't use logarithmic functions as shown above. The input range is broken into segments, each segment using a different interval between decision values. Most segments contain 16 intervals, and the interval size doubles from segment to segment. The illustration shows three segments with four intervals in each.

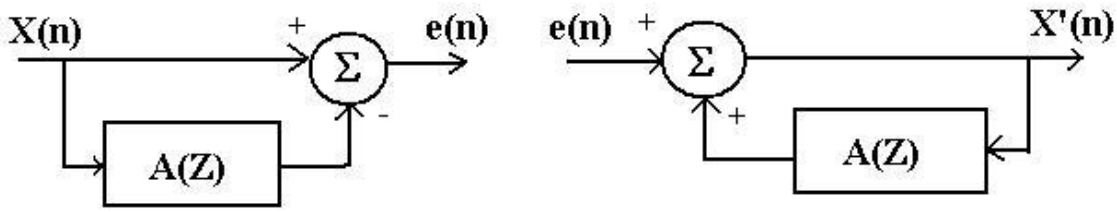


Both encodings are symmetrical around zero. mu-law uses 8 segments of 16 intervals each in each of the positive and negative directions, starting with a interval size of 2 in segment 1, and increasing to an interval size of 256 in segment 8. A-law uses 7 segments. The smallest segment, using an interval of 2, is twice the size of the others (32 intervals). The remaining six segments are "normal", with 16 intervals each, increasing up to an interval size of 128 in segment 7. Thus, A-law is skewed towards representing smaller signals with greater fidelity.

But in order to reduce the bandwidth of transmitted signal we have to extract features from the signal and send that rather than send the signal itself represented as codewords. This brings us to the scheme G.728. But to understand this scheme we shall first look at the Linear Predictive Coding.

8.2.2 Linear Predictive Coding (LPC)

A linear predictive coder is a popular vocoder that extracts perceptually significant features of speech directly from a time waveform rather that from frequency spectra. Fundamentally LPC analyzes a speech waveform to produce a time-varying model of the vocal tract excitation and transfer function. A synthesizer in the receiving terminal re-creates the speech by assign the specified excitation through a mathematical model of the vocal tract. By periodically updating the parameters of the model and specification of the excitation the synthesizer adapts to changes in either. During any one specification interval, however, the vocal tract is assumed to represent a linear time-invariant process. The figure below shows the analysis and synthesis block diagrams.



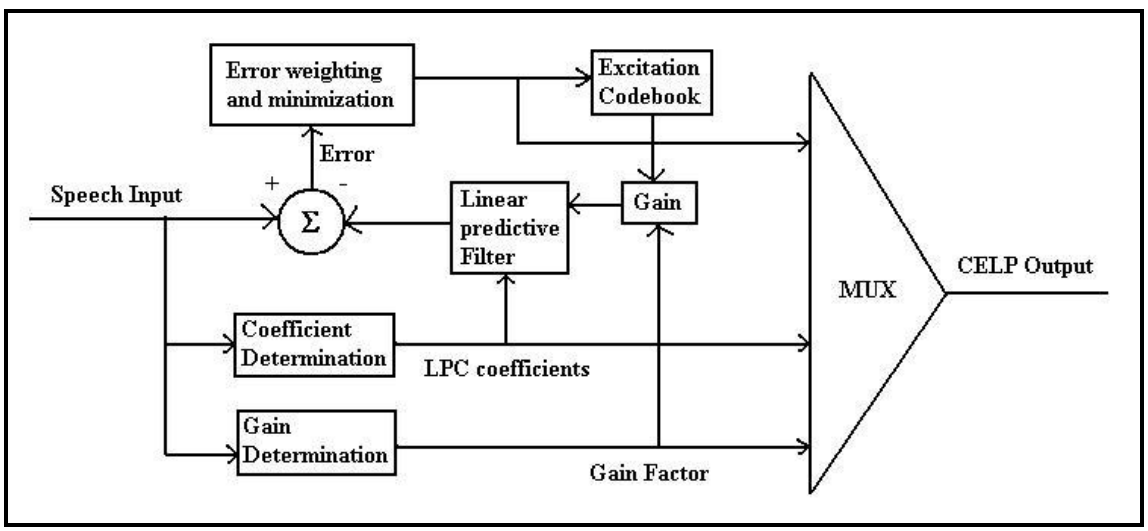
The analysis process determines the coefficients of the vocal tract model ie. A(Z) to minimize the error e(n) with a given set of speech samples X(n). If the decoder knows both A(Z) and e(n), it can re-create the input samples X(n). The information that a LPC encoder/analyzer determines and transmits to the decoder/synthesizer consists of:

- Nature of excitation (voiced or unvoiced)
- Pitch period (for voiced excitation)
- Gain factor
- Predictor coefficients (parameters of vocal tract model)

This brings us to the voce coding standard of ITU, the G.728.

8.2.3 G.728

G.728 coders are based on the principle of Low Delay-Code Excited Linear Predictive coding (LD-CELP). CELP is an example of enhanced excitation LPC whose basic block diagram can be represented as shown below.



CELP uses a codebook for encoding residuals. Instead of encoding a residue (or error) waveform on a sample-by-sample basis and using that as excitation in the decoder, CELP assumes residuals are not random waveforms with independent samples but rather that a block of residue samples can be represented by one of a manageable number of waveform templates. “Excitation encoding” in this case involves selecting a codeword in the codebook that minimizes the overall error in the reconstructed signal. Residual

Decoding" in this context implies using the received codeword as an index into the table of codewords to obtain the residual sequence chosen in the encoder.

The LD-CELP speech coder operates upon 2.5 ms frames of digitized, telephone bandwidth speech or audio signals sampled at 8 kHz. The frames are divided into four 625 ms subframes of five samples each. Typically each sample is represented in companded 8-bit G.711 (m-law/A-law) format. Each group of five 8-bit samples is converted into a 10-bit index into an excitation codebook, resulting in a compression ratio of 4:1. Of course, linear 16-bit speech samples can also be processed by the algorithm, in which case the compression ratio is effectively 8:1.

9 The H.323 Suite of Protocols

H.323 is a standard technology for the transmission of real-time audio, video and data communication over packet-based networks. It specifies the components, protocols and procedures used for providing this multimedia communication. The packet-based networks include the IP-based internet, IPX based Local Area Networks (LANs), Metropolitan Area Networks (MANs) and Wide Area Networks (WANs). H.323 can enable a variety of communication - for example: audio only IP telephony, audio and video telephony, audio and data, video and data, etc. It is not limited to point to point communication, but applies to multipoint communication as well.

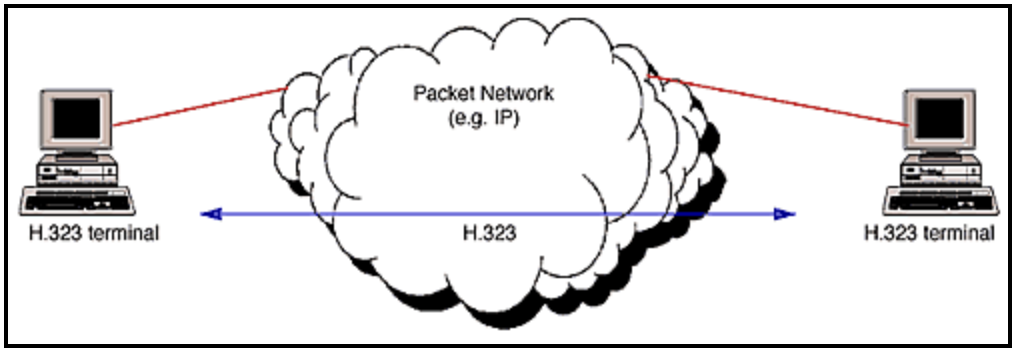


Figure 7 H.323 enables communication between 2 H.323 endpoints over a packet based network such as an IP network.

9.1 H.323 Versions

The H.323 standard is specified by the ITU-T Study Group 16. Version 1 of the H.323 recommendation - *visual telephone systems and equipment for LANs that provide a non-guaranteed quality of service* - was accepted in October 1996. The current version of H.323 under development is version 3.

9.2 H.323 Components

The H.323 standard specifies four kinds of components, which, when networked together, provide the point-to-point and point-to-multipoint multimedia communications services. They are:

- Terminals
- Gateways
- Gatekeepers
- Multipoint Control Units (MCUs)

9.2.1 Terminals

A H.323 terminal is either a personal computer or a stand-alone device running H.323, used for real-time bidirectional multimedia communications. It supports audio communication and can also support video or data streams. The primary goal of H.323 is to enable H.323 to terminals to inter work.

9.2.2 Gateways

A gateway is a logical component of H.323 which is used to connect two dissimilar networks. It can be implemented separately or as part of a gatekeeper or MCU. A gateway provides translation of protocols for call setup and release, conversion of media formats and the transfer of information between H.323 and non-H.323 networks. For example, a gateway in IP Telephony connects the IP network with a PSTN.

A gateway has two sides. On the H.323 side, terminals communicate with it using the H.245 control-signaling and H.225 call-signaling protocol. The gateway translates these protocols to the respective counterparts on the non-H.323 side in a transparent fashion. Another function of the gateway is to perform call setup and call clearing on both sides. The gateway may also provide facilities for audio and video format conversion if they differ on the two sides.

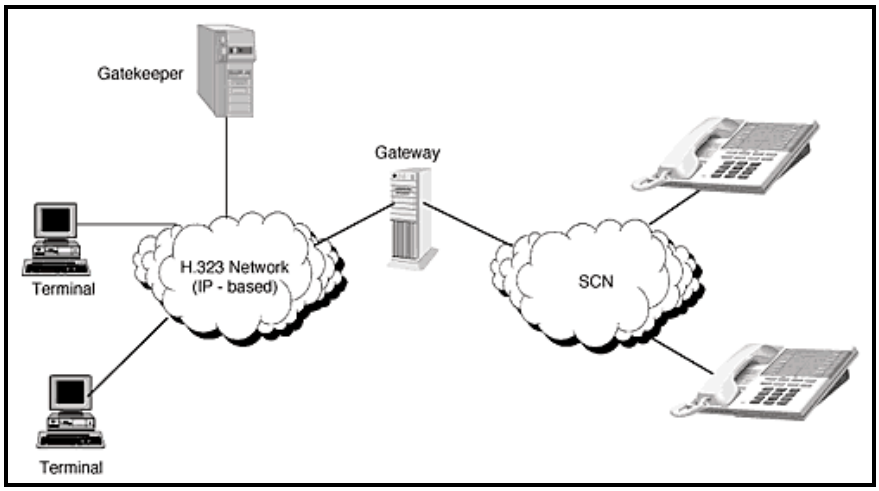


Figure 8 A Gateway connects an IP based based H.323 network (left) with a Switched Circuit Network (right)

9.2.3 Gatekeepers

A gatekeeper can be considered to be the brain of a H.323 network. It is the focal point for all calls within the network. The gatekeeper provides services like address resolution, authorization and authentication of terminals and gateways, bandwidth management, accounting and billing. It may also provide call routing services.

A gatekeeper is an optional component of a H.323 network. However, if it is present in the network, the endpoints of the network must use the services of the gatekeeper. Address Translation is an important function of the gatekeeper. For example, a gatekeeper in a H.323 network which contains a IP Telephony gateway translates the incoming E.164 telephone numbers to the corresponding network address (for example, IP address for IP networks) . Gatekeepers are also involved in bandwidth management and call-signal routing in their zones. The monitoring facilities provided by the gatekeeper can help in balancing the load in the network.

A H.323 endpoint communicates with a gatekeeper using the H.225 RAS protocol. H.225 RAS defines methods for gatekeeper discovery, endpoint registration, endpoint location, and admission control. As the channel carrying RAS messages is unreliable, message exchange may be associated with timeouts and retries.

As in the case of a gateway, the gatekeeper is a logical component of H.323. It can thus be implemented separately or as part of a gateway or MCU.

9.2.4 Multipoint Control Units

Multipoint Control Units or MCUs provide support for conferences of three or more H.323 terminals. All terminals wishing to participate in a conference must establish a connection with the MCU, which manages the conference resources and media streams.

The MCU is also responsible for negotiating the audio and video codecs to be used for the conference.

9.2.5 H.323 Zone

The collection of H.323 terminals, gateways and MCUs managed by a single gatekeeper constitute a H.323 Zone. A zone has only one gatekeeper and must consist of at least one terminal. A zone can be independent of network topology and may be comprised of multiple network segments connected using routers.

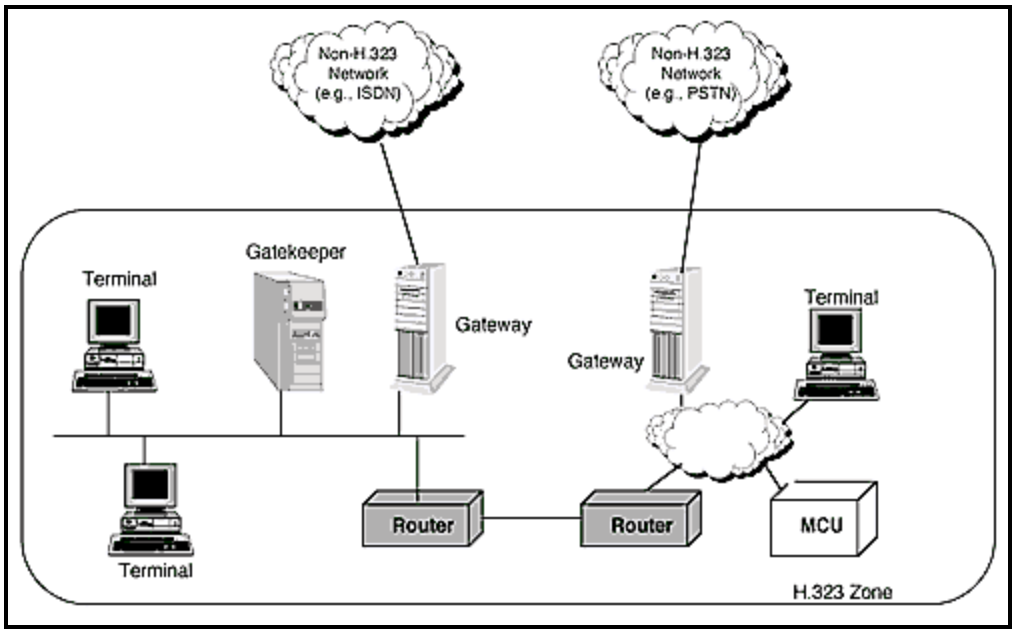


Figure 9 A H.323 Zone

9.3 H.323 Protocols

H.323 is a family of protocols which are independent of the packet network and the transport protocols over which they run. H.323 specifies the following protocols:

- audio CODECs
- video CODECs
- H.225 Registration, Admission and Status (RAS)
- H.225 Call Signaling
- H.245 Control Signaling
- Real-time Transfer Protocol(RTP)
- Real-time control Protocol (RTCP)

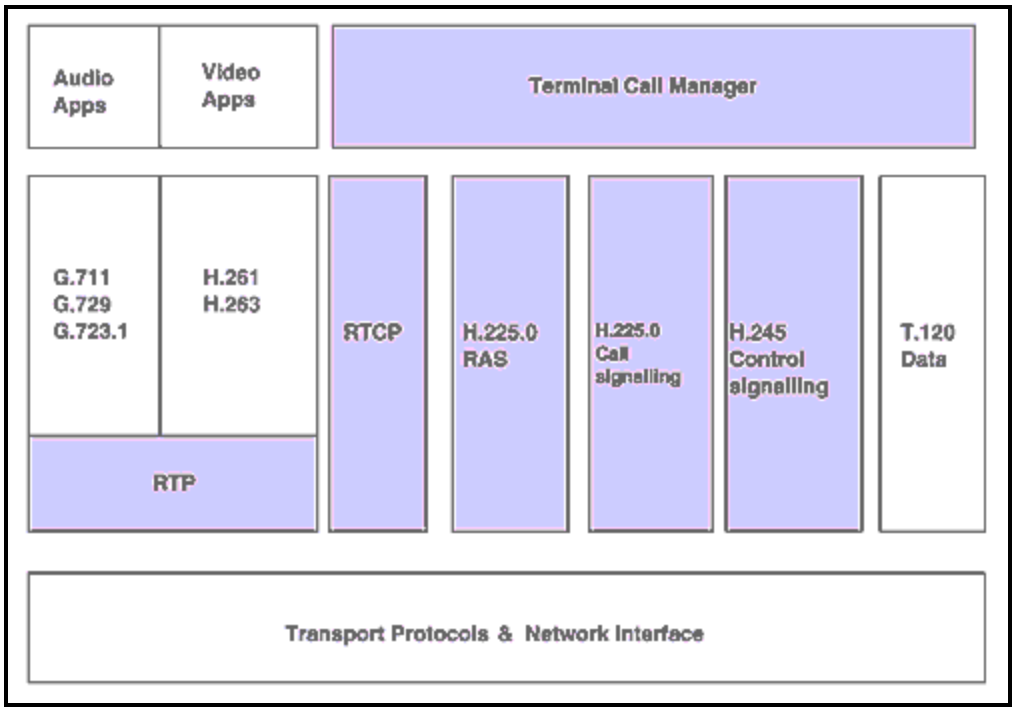


Figure 10 The shaded boxes represent the various protocols under H.323

9.3.1 H.225 Registration, Admission and Status (RAS)

This is the protocol used by an end-point (either a terminal or a gateway) to register itself with a gatekeeper. The protocol is used for registration, admission control, bandwidth changes, status and disengage procedures between the end-points and gatekeepers. The separate RAS signaling channel used for exchange of RAS messages is the first to be opened between the end-point and the gatekeeper.

9.3.2 H.225 Call Signaling

A call should be set up between two terminals, before data can be transferred between them. H.225 call signaling is used to establish a connection between two H.323 end-points or between a H.323 end-point and a gatekeeper. In contrast with H.225 RAS messages, the call signaling messages are carried over a reliable channel. For example, the H.225 protocol messages are carried over TCP in an IP-based H.323 network.

In the absence of a gatekeeper in the network, the H.323 terminals can directly signal the other terminals. This is called *direct signaling*. If a gatekeeper is present in the network, the terminals may communicate with other terminals only through the gatekeeper. This method is called *gatekeeper-routed call signaling*. The method to be adopted is decided by the gatekeeper during the initial H.225 RAS admission message exchange with the terminal.

9.3.3 H.245 Control Signaling

H.245 Control Signaling protocol defines the messages to be exchanged between H.323 endpoints which facilitate capabilities exchange and the opening and closing of logical channels. A logical channel can carry information from one endpoint to another endpoint or to multiple endpoints (in the case of a multipoint conference), in a single direction. The H.245 is the logical channel 0 and it remains permanently open for the duration of the call. The media channels opened through H.245 form the higher numbered logical channels.

The various H.245 end-to-end control messages exchanged between two H.323 terminals can be categorized as follows:

- Capabilities exchange
- Opening and closing of logical channels to carry media streams
- Flow-control messages
- General commands and indications

9.3.4 Real-Time Transport Protocol

The actual audio or video media streams are transported on RTP. RTP provides end-to-end delivery services of real-time audio and video. RTP typically transports data via the User Datagram protocol (UDP). RTP provides payload-type identification, sequence numbering, timestamping and delivering monitoring whereas UDP provides multiplexing and checksum services. Thus together with UDP, RTP provides transport protocol functionality.

RTP by itself does not provide any mechanism to ensure timely delivery or provide quality of service guarantees. An RTP packet may be dropped or delivered out of order by the network traversed. However, the sequence numbers and timestamp information present in the RTP headers enable an application to determine the proper location of the received packet in the media stream. For example, if an audio application receives packet a packet ahead of schedule, it can save the packet in its correct location in a sequential buffer.

RTP and RTCP protocols are defined by RFC 1889.

9.3.5 Real-Time Transport Control Protocol

RTCP is the counterpart of RTP which is used to monitor the quality of service and to convey information about the participants in an on-going session. RTCP messages are sent on a port which is different from the one for actual RTP media transfer. On the basis of RTCP messages, the sender of the media stream can change media transfer parameters so that QoS may be achieved. RTCP messages also help in synchronizing

two different media streams – for example, a video stream and its corresponding audio stream.

RTCP *Sender reports* containing information about the data transmitted and a synchronization timestamp are sent by the source of the media stream. Receivers of the stream in turn send back *Receiver reports* which contain information about the received data, lost packets, jitter experienced and delay. Source Descriptions which contain the name, email address, phone number and identification of a user are also exchanged through RTCP messages.

9.4 Making a H.323 Call

H.323 terminal T1 wants to call the H.323 terminal T2. The call takes place in three stages:

1. **Call Establishment :**
    a. T1 sends a H.225 RAS ARQ registration message to the gatekeeper. The gatekeeper confirms the admission of T1 and allows it to use direct call signaling.
    b. T1 sends a H.225 call signaling setup message to T2 requesting a connection. T2 replies with a H.225 call proceeding message to T1.
    c. T2 can accept the connection only after registering the gatekeeper. On receipt of registration confirmed message from the gatekeeper, T2 sends a H.225 alerting message to T1, followed by a H.225 connect message to T1.
    d. The call is established

2. **Control Signaling**
    a. H.245 control channel is established between T1 and T2.
    b. The respective terminal capabilities of terminals T1 and T2 are exchanged and acknowledged.
    c. T1 opens a media channel with T2 by a sending H.245 openLogicalChannel message. The transport address of the RTCP channel is included in the message.
    d. T2 acknowledges the establishment of the unidirectional logical channel from T1 to T2. The RTP transport address allocated by T2 to be used by T1 for sending the RTP media stream is included in the acknowledgement.
    e. A media channel is set up from T2 to T1 in the same fashion.
    f. Bidirectional communication is established.

3. **Media Exchange**
    a. T1 sends the RTP media stream to T2.
    b. T2 sends RTP media stream to T1
    c. Quality of media delivery parameters are exchanged through RTCP messages between T1 and T2

4. **Call Release**
    a. T2 initiates the call release by sending a H.245 EndSession command to T1
    b. T1 releases the call endpoint and confirms the release by sending an EndSession command to T2.
    c. T2 completes the call release by sending a H.245 release complete message to T1.
    d. T1 and T2 disengage with the gatekeeper.

10 Session Initiation Protocol (SIP)

Session Initiation Protocol is a direct competitor to H.323 in VoIP implementation. SIP is an application-layer protocol for creating, controlling and terminating Internet telephony or multimedia conference sessions between one or more participants. It is prescribed in the Internet Engineering Task Force RFC 3261.

A session with regard to SIP is an exchange of data between groups of participants. A session can involve the transfer of data in different media such as voice, video or test messages. SIP works together with the various protocols devised for the above media transfers and helps the users in setting up the sessions they desire.

SIP is independent of the transport protocol used in media transfer. It does not depend on the type of session being established either. However, SIP does not provide services. SIP is a component that can be used with other protocols to build a complete multimedia architecture. For example, SIP primitives can be used to locate users. Protocols like RTP can then be used for transporting real-time data between the users.

10.1 Functions of SIP

The main function of SIP is in establishing and maintaining multimedia sessions such as Internet telephony calls. The functionality of SIP can be divided into the following five categories:

1. **User Location**: Determination of the end system to be used for communication.
2. **User availability**: Determination of the willingness of the called party to engage in communications
3. **User Capabilities**: Determination of the media and the parameters to be used in the session
4. **Session Setup**: Contacting or “ringing” the destination and the establishment of session parameters at both ends of the call.
5. **Session Management**: Transfer and termination of sessions and modification of session parameters.

10.2 SIP in Action

SIP is based on an HTTP-like request/response transaction model. Each transaction consists of a request that invokes a particular method, or function, on the server and evokes at least one response. The following paragraph highlights a typical SIP transaction.

User 1 at a PC-based soft phone (T1) wishes to communicate with User 2 at a SIP phone (T2) over the internet. User 2 is identified through a SIP Uniform Resource Indicator of the form "*sip:name@domain*". The first step in calling User 2 is to send a *SIP INVITE* message to User 2's terminal T2. As T1 does not know the location of either T2 or T2's SIP server, the *INVITE* message is sent to T1's predetermined proxy server. This proxy locates T2's proxy through a DNS lookup and forwards the *INVITE* message. On receiving the invite message, T2's proxy further forwards it to T2, which in turn sends a reply. Acknowledgements are sent at each stage in the above procedure. The route taken to reach T2 is recorded in the VIA fields in the SIP message header. Hence the path to be taken for T2's replies is fixed.

Once a session is established between T1 and T2, media transfer can take place. The format for media transfer is agreed on the basis of the Session Description Protocol (SDP) packets which formed the payload of the *INVITE* SIP packets. The end-to-end media packets exchanged between T1 and T2 may take a different path from that taken by the SIP signaling messages.

10.3 Registration

Registration is a common operation in SIP. A SIP registration server learns the current location of a terminal when it receives SIP *REGISTER* messages from the terminal at periodic intervals. A user may register from more than one SIP device at the same time. For example, a user may register from both his SIP home phone and SIP office phone. Similarly, more than one user can be registered on a SIP device at the same time. The server records the SIP *URI* – transport address bindings into a location database. The registration functionality is usually incorporated into the SIP proxy server, which is then also called a *REGISTRAR*.

10.4 Structure of SIP

SIP is structured as a layered protocol. This implies that its behavior is described in terms of a set of fairly independent processing stages with only a loose coupling between each stage. The lowest layer of SIP is its syntax and encoding, which is specified using an augmented *Backus-Naur Form (BNF)* grammar. The second layer is the transport layer which defines how a client sends requests and receives responses over the network. The Transaction layer constitutes the third layer. Transactions are a fundamental component of SIP. Any task that a user agent client accomplishes takes place using a

series of transactions. A request sent by a client to a server, along with all responses to that request form a transaction. The transaction layer handles application-layer retransmissions, matching of responses to requests and application-layer timeouts. The layer above the transaction layer is called the *transaction user (TU).* A TU can create transaction instances and passes it to the destination. All SIP entities, except stateless proxies are TUs.

SIP is a text-based protocol and uses the UTF-8 charset. Every SIP message is either a request from a client to a server or a response from a server to a client.

10.5 Session Description Protocol (SDP)

SDP is an IETF protocol intended for describing multimedia sessions. It is a text format description that provides many details of the session like the originator of the session, a URL related to the session, the connection address for the session media, etc. SDP information is often carried as the payload of SIP invite messages.

SDP is specified in RFC 2327.

11 H.323 vs. SIP – What should we use?

11.1 Differences between H.323 and SIP

Both H.323 and SIP are currently competing for the dominance of IP Telephony signaling. Each suite of protocols supports the majority of end-user functions like call setup and tear-down, call holding, call transfer, call forwarding, call waiting and conferencing. The functional differences between H.323 and SIP are minor. For example, H.323 supports message waiting indication which is not supported by SIP, while SIP offers third party control of calls which is not supported by H.323.

More important differences between SIP and H.323 arise in the domain of QoS, scalability and interoperability. It is argued that SIP should scale better as it is a much simpler protocol than H.323. SIP is faster than H.323 – what SIP does in one transaction is achieved by 3-4 H.323 transactions. But version 3 of H.323 allows a FastConnect option , which reduces the number of transactions needed. The text based SIP is easier to code and debug. However, ASN.1 based H.323 messages are more compact in size.

SIP uses URIs to identify users while H.323 uses email aliases. The SIP URIs of the form *sip:user@domain.com* allow calls to be redirected to a webpage, mailto address etc, as the protocol is explicitly stated in the URI. Call priority, which is a legal requirement in some countries, is another feature handled by SIP, but not offered by H.323.

H.323 has a clear advantage over SIP in its use of *logical channels.* H.323 makes a clear distinction between the media types that can be sent and the media types that are currently active and sent over the network. SIP makes no such distinction and therefore, there is no way to open a media connection apart from actually sending the media. This leads to performance degradation and overheads of port allocation.

The Multipoint Control Units (MCUs) present in H.323 networks allow better control over multicast conferences. The floor control offered by SIP is more limited. H.323 also offers better gatekeeper discovery procedures.

At present, H.323 enjoys more industry support. For example, the popular program NetMeeting is a H.323 application. SIP is a more recent protocol and has not been implemented on a large scale. Both SIP and H.323 may be supported on future products.

11.2 What should we use?

Current discussions on a VoIP often turn out to be a debate between the usage of H.323 and SIP. At present, most products support only H.323 versions 1 and 2. SIP is still in the early stages of development and has not received any major vendor support. However, it is the expected that both protocols will be supported in future products. It is also worth noting that both standards appear to be converging by picking up the good features from each other. As of now, using H.323 over SIP in VoIP implementation is a good idea.

12 Media Gateway Control Protocol - MGCP

12.1 What is MGCP?

MGCP or Media Gateway Control Protocol is a protocol for controlling Telephony Gateways (or Media Gateways MG) from external call control elements called media gateway controllers (MGC) or call agents. MGCP was first proposed as an alternative protocol to address the needs of large-scale phone-to-phone deployments. The first proposal came in May 1998 from Bellcore (now Telcordia) and Cisco to address the needs of cable operators. It was called Simple Gateway Control Protocol (SGCP). It was a cost effective alternative and better-suited protocol to implement and deploy than current H323 implementations, in the context of the cable operator's market. Then came the Internet Protocol Device Control (IPDC), presented a month later. This protocol was nearly the same as SGCP but IPDC uses Diameter to carry PDU's between respective entities whereas SGCP depends solely on UDP. Then Bellcore and Level3 played a key role in the reunion of these two protocols into one, MGCP. It was later presented to standard groups (IETF, ETSI TIPHON, ITU-T SG16).

12.2 The Working of MGCP

MGCP is a protocol used by external Media Gateway Controller to control Media Gateways. A Media Gateway is a network element, which convert circuit-switched voice into packet-based traffic (Internet or other packet network). We can have for example:

- **Trunking gateways**: interface between the public telephone network and a VoIP network. Such gateways typically manage a large number of digital circuits.

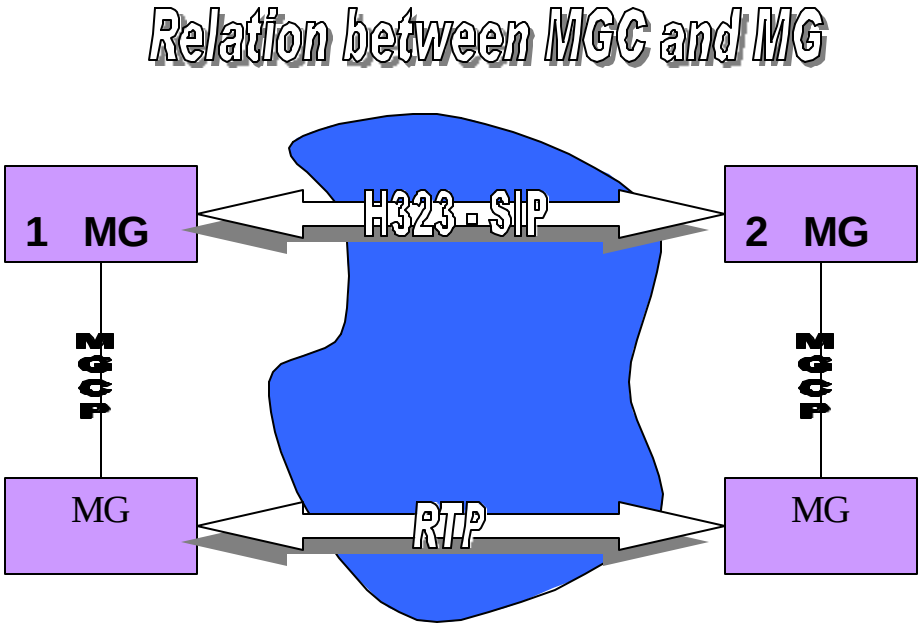
- **Voice over ATM gateways**: same function as the trunking gateways, except that they interface to an ATM network.

- **Residential gateways**: provide a traditional analog (RJ11) interface to a VoIP network. Examples of residential gateways include cable modem/cable set-top boxes, xDSL devices, and broadband wireless devices

- **Access gateways**: provide a traditional analog (RJ11) or digital PBX interface to a VoIP network.

- **Network Access Servers**: attach a "modem" to a telephone circuit and provide data access to the Internet. We expect that, in the future, the same gateways will combine Voice over IP services and Network Access services.

MGCP has call control architecture. That means that the call control (signaling) "intelligence" is separated from the media handling (Media Gateway). These call control elements, or Call Agents, will synchronize with each other to send coherent commands to the gateways under their control. Assuming that the gateways execute the commands sent by the Call Agent, MGCP is a master/slave protocol.



MGCP decomposes the gateways, but these appear to the outside as a single VoIP gateway.

12.3 MGCP Commands

The MGCP implements the media gateway control interface as a set of transactions. The transactions are composed of a command and a response. There are eight types of commands.

Name	Applicant	function
CreateConnection	MGC —> MG	Creates a connection between two endpoints
ModifyConnection	MGC —> MG	Modifies the properties of a connection
DeleteConnection	MGC <—> MG	Modifies the properties of a connection
NotificationRequest	MGC —> MG	Requests the media gateway to send notifications in an endpoint
Notify	MG —> MGC	Informs the media gateway controller when observed events occur
AuditEndpoint	MGC —> MG	Determines the status of an endpoint
AudioConnection	MGC —> MG	Retrieves the parameters related to a connection
RestartInProgress	MG —> MGC	Signals that an endpoint or group of endpoints is take in or is out of service

All commands are composed of a command header, optionally followed by a session description. All responses are composed of a response header, optionally followed by a session description. Headers and session descriptions are encoded as a set of text lines, separated by a carriage return and line feed character (or, optionally, a single line-feed character). The headers are separated from the session description by an empty line.

The command header is composed of:

- A command line, identifying the requested action or verb, the transaction identifier, the endpoint towards which the action is requested, and the MGCP protocol version,
- A set of parameter lines, composed of a parameter name followed by a parameter value.

The command line is composed of:

- Name of the requested verb.
- Transaction identifier correlates commands and responses. Values may be between 1 and 999999999. An MGCP entity cannot reuse a transaction identifier

sooner than 3 minutes after completion of the previous command in which the identifier was used.
- Name of the endpoint that should execute the command.
- Protocol version.

13 Quality of Service in VoIP networks

13.1 Why is QoS required?

The human ear is very sensitive to even minor changes in an audio signal. The encoding of the voice signals is the first stage where voice quality may be compromised. Bigger losses of quality occur while the packetized voice signals are transferred over the inherently unreliable packet-based networks. Delays, jitters, dropped packets and out-of –sequence packets can wreak havoc with the voice quality.

Quality of Service works by classifying traffic into different classes which are handled differently by the network. QoS can be guaranteed using a wide variety of approaches. Some of these methods are more applicable to VoIP networks than the others. The following approaches are usually taken to assure QoS in VoIP networks.

13.2 DiffServ

Every IP packet header has a 8 bit TOS (Type of Service Field). This field has been mainly left unused by the existing IP networks. DiffServ is an attempt at providing QoS by using the first 6 bits of the TOS field, now also referred to as DSCP (Differentiated Services Code Point), to classify the IP packet into one of 64 different classes.

Most IP phones and VoIP gateways set the TOS byte to a non-zero value to denote the priority needed for VoIP. A common value for the byte is 160 for voice transport and 96 for the call setup traffic (as it requires a lesser priority). Using the DCSP field for classification is very efficient as there no overheads involved at the routers. The routers in any case, look at the IP headers.

DiffServ is a good way to implement QoS in VoIP networks. The packet classification should be done as close to the source of the call as possible. The various network elements must be configured to recognize the DiffServ classes and to provide an appropriate level of resources for the packets classified as VoIP.

13.3 MPLS

MPLS or Multi Protocol Label Switching is a technique by which packets are tagged with a label which is used to determine the route to be taken by the packet. By using different labels for different classes of data, MPLS aware Label Switch Routers (LSRs) can set up specific routes through the network. For example, a high priority class of data can be assigned a less congested route through the network. Labels are often assigned to the traffic on the basis of the DiffServ fields.

MPLS is applicable for large network backbones with many routers and is used by national carriers and ISPs. MPLS can be used by service providers to internally implement a VoIP service. This is done by assigning labels to the VoIP traffic such that it is sent over a dedicated Label Switched Path (LSP) through the network, which guarantees the low delay, jitter and loss required by VoIP.

13.4 RSVP

The Resource Reservation Protocol (RSVP), defined in RFC 2205, reserves resources to meet requirements for bandwidth, jitter, and delay on a particular network path through a series of routers. The receiver sends *RSVP RESERVE* messages along the reverse data path to the sender and the routers along the way reserve the necessary resources for the forthcoming application flow.

RSVP is not recommended for the short duration VoIP calls due to its large setup time and the extra bandwidth required for establishing and maintaining the reservations. RSVP works well for longer duration applications like streaming video and also when the number of simultaneous connections is not large.

13.5 IP Phone configuration

Suitable configuration in the IP phones used in VoIP leads to QoS gains. However many tradeoffs are involved in fine tuning the settings to achieve QoS.

The codec used in the IP phone determines the bandwidth required for the VoIP call. Codecs like G.711 consume many times more bandwidth than codecs like G.729. But naturally, the voice quality offered by the low bandwidth codecs is poorer than those offered by the higher bandwidth ones. The silence suppression feature present in some IP phones and softphones can reduce the bandwidth requirements by an order of 50%, by not transmitting any packets when the user is silent. However the speech may sound choppy and clipped. Sending larger VoIP packets reduces the per packet header overhead. However, loosing a larger packet implies a larger loss in voice quality.

The IP phone configuration has to be based on a thorough analysis of the benefits of QoS and the tradeoffs in voice quality involved.

13.6 Other Measures to achieve QoS

Methods to enhance the QoS achieved in VoIP are not limited to DiffServ, MPLS, RSVP and the IP telephone configuration. Using the Link Layer Level Ethernet priority field and RTP header compression also aids QoS. Gatekeepers in H.323 implementations can enforce admission control to prevent the downgrading of service when a large number of users attempt to connect simultaneously. Implementing traffic shapers at the edge of the networks and queuing methods like Weighted Fair Queuing (WFQ) and Weighted Random Early Detection (WRED) also aid in achieving QoS in VoIP.

14 Current Status of VoIP

VoIP is a recent technology, under the process of evolution. However it has achieved immense popularity already. The market for VoIP enabled products is estimated to be over $3 billion in 2002. Many major companies including Cisco, 3Com, VocalTec etc are actively engaged in developing both VoIP hardware and software. VoIP is sure to grow on a larger scale in the future, with more and more people taking advantage of its cost-effectiveness.

VoIP became legal in India only in 2002. VoIP are services are currently being offered by the Internet Service Providers (ISPs) mainly in the form of fixed value calling cards. It is becoming a popular medium of making international phone calls. However, the QoS achieved needs to be improved.

15 Conclusion

This term focussed on VoIP as a technology. It discussed the various services offered by VoIP and the problems associated with it. The protocol suites authored to overcome these problems were described. Achieving QoS guarantees in a VoIP network was also discussed.

VoIP is in a period of constant change. New protocols are being devised and existing ones are being enhanced. A wider ranger of VoIP products and applications are emerging. The promise of VoIP as a cost effective and efficient communication medium is certainly bright.

16 References

- John Q. Walker, Jeffrey T. Hicks, *The Essential Guide to VoIP implementation and management*, NetIQ Corporation

- Olivier Hersent, David Gurle, Jean-Pierre Petit, *IP Telephony*, Pearson Education, 1998
- RFC 2205 – *Resource Reservation Protocol (RSVP)*
- RFC 2327 – *Session Description Protocol (SDP)*
- RFC 3261 – *Session Initiation Protocol (SIP)*
- RFC 1889 – *RTP: A Transport Protocol for Real-Time Applications*

17 Glossary

Keyword	Meaning
Backus-Naur Form	See BNF
BNF	Backus-Naur-Form - a widely accepted format for syntax specification.
Capabilities Exchange	The procedure by which two H.323 endpoints tell each other what capabilities they possess - for example, the kind of media streams they can process.
DiffServ	It is QoS approach based on distinguishing packets into various classes at the edge routers.
Direct Signaling	A H.323 is allowed to directly signal the destination H.323 terminal. See Gatekeeper-routed signaling
DNS	Domain Name Services
E.164	The standard prescribed for international telephone numbers by the ITU
Gatekeeper-routed Signaling	All signaling between any H.323 terminals is routed through the gatekeeper.
H.225 Call Signaling	This protocol is used to establish a call between two H.323 terminals
H.225 RAS	RAS stands for Registration Admission and Status. This protocol is used by a H.323 endpoint to communicate with a H.323 gatekeeper
H.245 Control Signaling	This protocol facilitates capabilities exchange ad the opening and closing of logical channels between H.323 endpoints
H.323	H.323 is a technology to facilitate the transmission of real-time audio, video and data communication over packet-based networks. It is an umbrella specification for a large number of protocols.
H.323 Gatekeepers	It is the focal point of a H.323 network providing services like address resolution, authorization, accounting etc.
H.323 Gateway	It is a logical component of H.323 which is used to two connect two dissimilar networks - for example, a H.323 network with a PSTN
H.323 Terminal	A H.323 terminal is either a personal computer or stand-alone device running H.323 and used for real-time bidirectional multimedia communications.
H.323 Zone	The collection of H.323 terminals, gateways and MCUs managed by a single gatekeeper constitute a zone.
IETF	Internet Engineering Task Force
IP	Internet Protocol - The Network protocol used in the internet. IP addresses specify the address of the destination. Packets are independently routed to a destination across mutlitple networks.
IP Telephony	See VoIP
ITU	International Telecommunications Union - An international organization under the United Nations Organization which promotes international cooperation in the development of telephony

Keyword	Meaning
Logical Channel	A Logical channel defines a media stream in a single direction. The channel need not always be active.
MCU	Multipoint Control Unit
MCU	It is the component of a H.323 network that provides support for conferences between multiple H.323 terminals.
Multipoint Control Unit	See MCU
PSTN	Public Switched Telephone Network - This refers to the "normal" telephone network. A virtual circuit is setup for the duration of the call between the source and the destination.
QoS	Quality of Service specifies the minimum guarantees expected from a particular service. For example, an acceptable voice quality in VoIP
RAS	See H.225 RAS
Registrar	A SIP proxy server which also handles registration messages from SIP terminals.
RFC	Request For Comments - IETF standards and protocols are specified as RFCs
RSVP	Resource Reservation Protocol is used to reserve resources for individual flows along the data path from the sender to the receiver.
RTCP	Real-time Transport Control Protocol. RTCP is a companion protocol of RTP which provides for feedback about the quality of data received.
RTP	Real-time Transport protocol. RTP is usually based over UDP. It supports sequence numbers, timestamps required to transfer real-time data streams over a packet based network
SDP	Session Description Protocol (described by RFC 2327)
SIP	Session Initiation Protocol - A competing standard with H.323 used in VoIP implementation (defined in RFC 3261)
Transaction	A request sent by a client to a server, along with all responses to that request constitute a SIP transaction.
Transaction User	The layer above the Transport Layer in the SIP protocol structure. A TU creates transaction instances and passes it to the destination.
TU	See Transaction User
URI	Uniform Resource Indicator
VoIP	VoIP stands for Voice over Internet Protocol. VoIP is the ability to make telephone calls and send faxes over IP-based data networks.