# IP TELEPHONY: AN INTRODUCTION

BY ENGR. KHALIL AHMED
Royal Saudi Naval Forces
Riyadh, Saudi Arabia

## ABSTRACT

*IP telephony is a rapidly emerging technology for voice communication that has astonished both the data-communication and telecom industries. Technical developments over last few years have made the use of IP networks for telephony applications a reality. The objective of this paper is to provide the introduction to the technology, and describe the related protocols and the issues behind delivering an "appropriate" quality of service.*

*The paper reviews the IP telephony technology, covering its background, working, and various methods of its implementation. It then explores the pros and cons of this exciting technology and looks at what the future has to offer. It also examines the market trends and looks at what market researchers think about IP telephony. The paper also covers the architecture and protocols underlying IP telephony, and discusses the various signaling standards that make the technology possible. At the end, it covers the issues related to the voice quality and discusses how these may be dealt with.*

## INTRODUCTION

Little more than a decade ago Internet was not available to general public, interactive voice communications were only made by telephone at Public Switched Telephone Network (PSTN) and the data exchange was expensive for long distances. Then we saw appearing some interesting things: Personal Computers to large masses and Internet to general public. People begun to use PCs and Internet to communicate and to exchange data (images, sounds, documents) with new services like email, chat and discussion groups.

Availability of a telephone and access to a high quality public switched telephone network is considered to be essential in modern society. There is however, a Paradigm shift beginning to occur since more and more communications is in digital form and transported via packet networks such as IP, ATM and Frame Relay. Since data traffic is growing much faster than telephone traffic, there has been considerable interest in transporting voice over data networks.

Support for voice communications using the Internet Protocol (IP), usually called "*IP telephony*", has become especially attractive given the low cost, flat-rate pricing of the Internet. In fact, toll quality telephony over IP has now become one of the key steps leading to the convergence of the voice, video and data communications industries.

The first IP telephony product was introduced by Vocaltec in early 1995. In a very short period of time it has caught the world's attention. The technology has now improved to a point where conversations are easily possible and it continues to get better. Technology is improving so fast that it will not be long until phone-to-phone service on the Internet offers the same call quality as conventional phone service.

## IP TELEPHONY

IP telephony (also known as *Internet telephony* or *Voice over IP*) is transport of voice calls over packet-switched IP-based data networks, no matter whether traditional telephony devices, multimedia PCs or dedicated terminals take part in the calls and no matter whether the calls are entirely or only partially transmitted over the IP network. This technology enables standard data packets to transmit multimedia information such as voice, fax or video over the Internet or a corporate intranet (IP-based private network) with suitable quality of service (QoS) and a much superior cost/benefit ratio. It draws on open standards and recommendations generated by international groups such as the Internet Engineering Task Force (IETF) and the International Telecommunication Union (ITU). All suppliers of Internet telephony products meet these standards.

When used only across an IP network, such as intranet or Local Area Network (LAN), it is generally known as

"*Voice over IP*" (VoIP). When the call originates and/or terminates in public switched telephone network, and the transport is Internet or Wide Area Network (WAN), it is generally called "*IP telephony*" or "*Internet telephony*."

IP telephony requires people, who want to talk to each other, to log onto a computer equipped with a microphone and speaker and establish a connection over the Internet. However, a user doesn't have to be online to reap the benefits of IP telephony – A user logged on to a computer may also make a call to a telephone. Even a telephone to telephone call can be made over the Internet. Whenever a telephone is used, the call must be transferred from the Internet to the local telephone system. The companies that provide Internet phone software also provide gateways through which these conversions occur. A fee for using the gateway is incurred by the user; these charges are very low compared to standard long-distance phone call charges. For example, a call to an ordinary telephone in United States from Saudi Arabia over the Internet could be as low as $.04/minute, as compared to $0.80/minute over telephone lines. Today, users can bypass long-distance carriers and their per-minute usage rates and run their voice traffic over the Internet for a flat monthly Internet-access fee.

Rapidly changing technology is making IP telephony a legitimate alternative for voice services over the Internet. And it doesn't stop with just voice over the Internet – other applications include fax over the Internet, call center integration, conference bridging and telecommuter access. These applications are all made possible over a single access line using TCP/IP architecture and the Internet.

## PRINCIPLES OF IP TELEPHONY

To understand IP telephony, it's necessary to be familiar with the fundamental characteristics behind the Internet and how it compares to the Public Switched Telephone Network (PSTN). The most important of these characteristics is the data transport mode, also known as data connection type which is either a circuit switched or packet switched as explained below:

**Circuit Switched Connection:** A device using a circuit switched connection only connects when data is to be sent. The connection is dedicated exclusively to the sending and receiving nodes for the entire duration of the call. Because the two points are connected in both the directions, the connection is called a circuit. The connection is only present when you need it and, since bandwidth remains constant, you only pay for the duration of the connection. While connected on a circuit switched network you have exclusive use of the established connection and data can be sent continuously. This type of data transaction is typically routed through the PSTN. Although the circuit switched network provides a very reliable connection for voice transmissions, it makes very inefficient use of the available bandwidth.

**Packet Switched Connection:** While circuit switched connection is open and constant for the entire duration of the call, packet switched connection opens just long enough to send a small chunk of data, called a packet, from one system to another. A packet switched connection keeps you connected all the time but you only pay for the amount of data transferred. In this case, the data is divided into small packets and each packet contains a source and a destination address. Packets of data are sent from source to destination using the quickest route available. The network bandwidth is shared and multiple simultaneous users are allowed to access multiple locations across a network. This provides for much more efficient use of available bandwidth but can create problems for voice traffic, which is very sensitive to delay.

The PSTN is a circuit-switched network that has been optimized for real-time or synchronous voice communication with a guaranteed Quality of Service (QoS). The PSTN guarantees the QoS by dedicating a full-duplex 64Kbps circuit between the parties of a telephone conversation. Regardless of whether the parties are speaking or silent, they use full 64K dedicated circuit until the call ends. Much of this capacity is wasted during a normal telephone conversation, because while the line is working at full capacity, not all of each user's time is spent transferring data or talking.

On the other hand, IP networks are packet-switched networks that have historically been used for applications where a variable QoS is tolerable. Instead of keeping a circuit open constantly, IP networks send and receive data only as needed, a bit at a time, in data packets. By doing so, IP networks free up network resources, as well as the resources of the computers sending and receiving information. Since these do not dedicate a path between sender and receiver, these cannot guarantee QoS. There are, however, ways that can be used to get reasonable QoS on the IP networks.

### *The Basics of IP Voice Call*

IP telephony technology uses packet-switching to minimize the amount of resources used in a telephone connection. The telephony application digitizes and compresses the analog voice signals. This data is then transmitted as a stream of packets over an IP network. IP network allows each packet to independently find the most efficient path to the intended destination, thereby best using the network resources at any given instant. At the destination, the packets are re-assembled back into their original order. The recipient IP telephony application then decompresses the packets and converts them back into the analog voice signal. The application insures proper reconstruction of the voice signals, compensating for echoes, jitter, and for dropped packets. The actual end-to-end process, however, may involve more steps as described below:

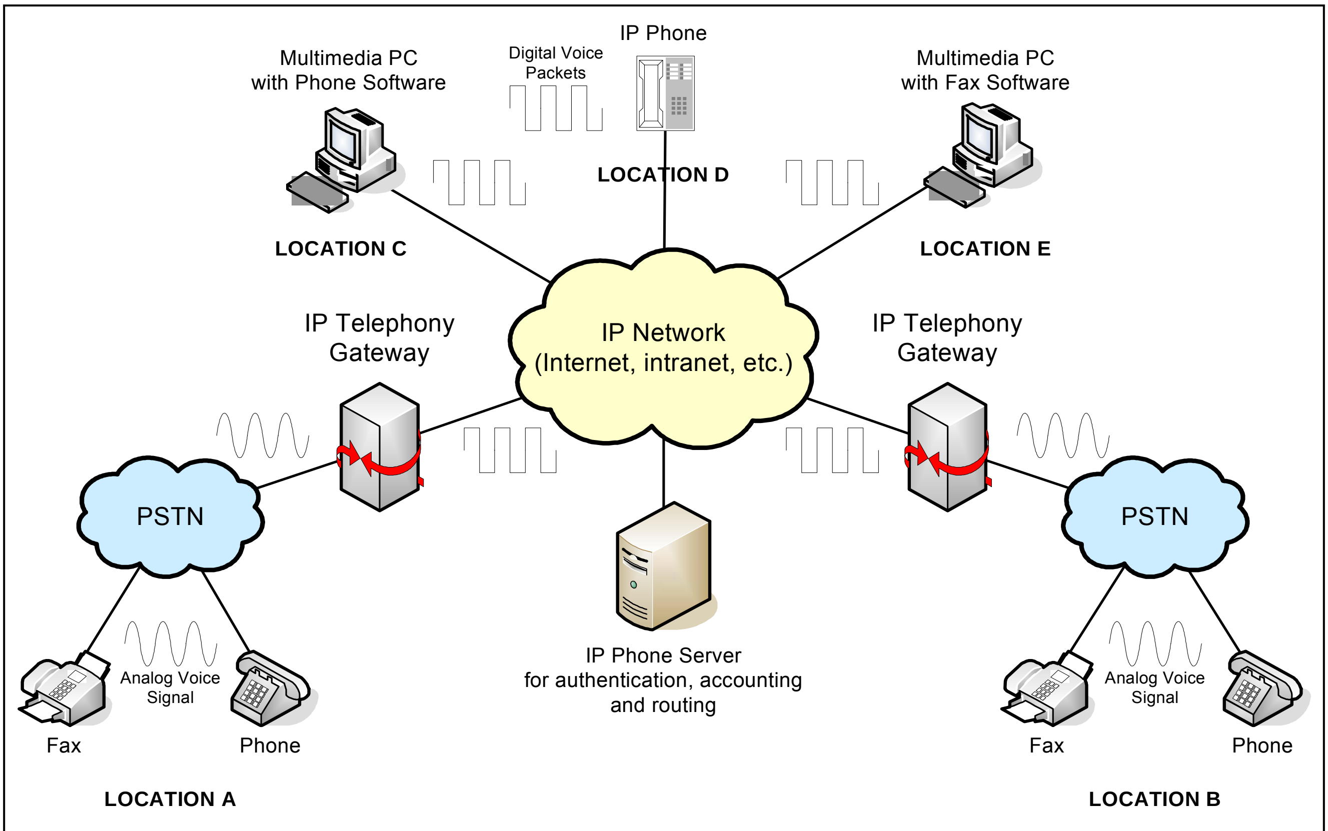



**Figure 1: IP Telephony Infrastructure**

A phone or fax call goes over the local PSTN to the nearest IP telephony gateway (*see* Figure 1). The gateway uses an Analog to Digital Converter (ADC) to encode the voice digitally and to compress/translate the signal into IP packets. It then moves the data onto IP network for transport to a gateway at the receiving end. The gateway at the receiving end decompresses the data and converts it back into analog voice signal using a Digital to Analog Converter (DAC). The gateway then hands over the voice signal to the PSTN network, which in turn passes it on to the phone or fax at the receiving end.

# IP TELEPHONY SCENARIOS

The IP telephony usage scenarios, as shown in Figure 1, are commonly classified by the type of devices terminating an IP call. Because there may be either a PSTN device (e.g. telephone) or a data-oriented terminal (e.g. personal computer) on each side of a call, there are four generic classes as below:

## PC-TO-PC

PC-to-PC communication can be provided for multimedia PCs (i.e. Personal Computers with a microphone, speaker and a sound card) operating over an IP-based network without connecting to the PSTN. PC applications (and IP-enabled telephones) can communicate using point-to-point or multipoint sessions. This set up requires that parties be equipped to talk at the time of the call.

This class is attractive especially for private users who already have an Internet access and a multimedia PC. Necessary software is available from several companies for free or at a very low cost. There is usually no charge for PC-to-PC calls, except for the cost of Internet access. The user doesn't even have to pay for long-distance calls. This pure-IP scenario can also take advantage of integration with other Internet services, such as instant messaging, video conferencing, etc.

## PC-TO-TELEPHONE

The PC-to-Telephone method allows a user to call any ordinary telephone on a PSTN from his computer. In this case, a gateway converting the IP call into a PSTN call has to be used. The gateway is required to be located as near to the called party as possible, to minimize the price for the gateway-to-called party connection. The call is converted to a PSTN call at the gateway and is then sent over PSTN to its destination.

Like PC-to-PC calling, this scenario requires a software client. The software is usually free, but the caller may have to pay a small per-minute charge to a gateway operator. The cost charged by the operator is determined mainly by the cost of the call placed from the gateway to the called party. This solution is commercially available from Net2Phone, PhoneServe and many other companies.

## TELEPHONE-TO-PC

The Telephone-to-PC method allows a user to call from any ordinary telephone on a PSTN to a PC connected to the Internet. In this case, a gateway converting the PSTN call into an IP call has to be used, and the gateway is required to be located as near to the caller as possible. The call is converted to an IP call at the gateway. The voice data then “hops on” the Internet and finds the PC on the other end by using the unique IP address.

A few companies are providing special numbers or calling cards that allow a standard telephone user to initiate a call to a computer user. The caveat is that the computer user must have the vendor's software installed and running on his computer. This solution may require user to pay local call charges, in addition to small per-minute charge to a gateway operator.

## TELEPHONE-TO-TELEPHONE

The Telephone-to-Telephone communication appears like a normal telephone to the caller but may actually consist of various forms of voice over packet network, all interconnected to the PSTN. In this scenario, a caller dials into a gateway using a regular telephone. The call is converted to an IP call at the gateway and the voice data “hops on” the Internet. At the end point the voice data hits another gateway and “hops off” the Internet. The voice data is converted back to PSTN format and sent over the PSTN to its destination.

This class is attractive for those who want to save on long-distance call and do not want to use PC. Since the call has to pass through two gateways – PSTN-to-Internet and Internet-to-PSTN, the cost is charged by both the gateway operators. In addition, the user may have to pay local call charges. This solution is commercially available from many companies, offering discounted rates for long distance IP telephony calls.

# BENEFITS AND DRAWBACKS

## BENEFITS

IP telephony could be applied to almost any voice communications requirement, ranging from a simple inter-office intercom to a complex multi-point teleconferencing environment. Listed below are the main benefits of using IP telephony:

- **Cost Reduction:** The first measure of success for IP telephony is the cost savings for long distance calls. Today flat rate long-distance pricing is available with the Internet and can result in considerable savings for both voice and fax. Large organizations with offices around the world save even more on long-distance calls by using local Internet gateways. IP telephony provides a competitive threat to providers of traditional telephone services that will clearly stimulate improvements in cost and function throughout the industry.
- **Simplification:** IP telephony enables a company to use a single communications medium rather than having to maintain separate systems for voice and data communications. An integrated infrastructure that supports all forms of communication allows more standardization and reduces the total equipment needs.
- **Flexibility:** IP telephony equipment has the flexibility to cater to a wide range of configurations and environments and the ability to blend traditional telephony with IP telephony. Hence, the quality of voice reproduction can be tailored according to the application. Customer calls may need to be of higher quality than internal corporate calls.
- **Network Efficiency:** The sharing of equipment and operations costs across both data and voice users can also improve network efficiency since a packet-switched IP network can handle more calls with the same transmission infrastructure than the PSTN can with its circuit switched TDM approach.
- **Integration:** Universal use of the IP protocols for both data and voice applications holds out the promise of reduced complexity and more flexibility. This provides an opportunity to share facilities such as directory services and security services, and eliminate points of failure.
- **Advanced Applications:** In addition to basic telephony and fax, the longer term benefits are expected to be derived from multimedia and point-of-service applications such as directory services that enable conference calls to be set up from Web-based directories, and wireless unified messaging, which will let users retrieve their voice and e-mail messages via their cellular phones. Combining voice and data features into new applications will also provide the greatest returns over the longer term.

## DRAWBACKS

Listed below are the main issues and drawbacks of using IP telephony:

- **Voice Quality:** The voice quality of an IP telephone call over Internet is usually not quite as good as a conventional phone, but the low price compensates for it. The quality is often compared to that of a speakerphone; a little choppy at times, but generally understandable. Calls on the public switched telephone network usually exhibit 50 to 70 ms delay. That latency increases substantially on the Internet, where it typically ranges to 500 ms.

- **Capacity:** One of the main parameters affecting the quality of service on the Internet is lost packets. It is a persistent problem, particularly with the increasing load of the Internet. This is mainly a function of network congestion. When traffic causes delays or out-of-order packets, some packets are dropped, causing breaks (silence) in the signal. Inadequate network access links, especially local ISP connections to the Internet backbone, are the major cause for bandwidth congestion.
- **Standards:** The major difficulty that IP Telephony technology is facing is the interoperability between IP telephony products. Hence, the users who want to make IP phone call have to have the same kind of software or IP telephony equipment.
- **Regulation:** Traditionally, telephone service has been heavily regulated. However, regulation of Internet telephony is still largely a question mark. Internet telephony has stirred fears from carriers throughout the globe, many of whom have reacted by seeking regulatory protection from the new technology. In most countries, governments or government- authorized entities retain the right for providing telephone service.

# APPLICATIONS

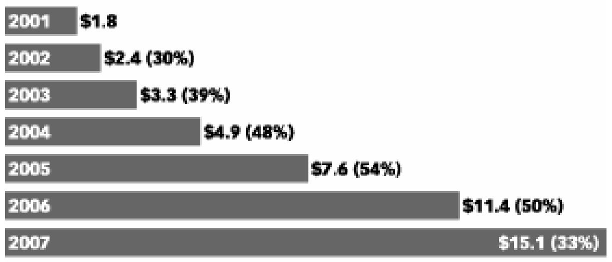
IP telephony enables a whole new generation of applications which are impossible with other telephony architectures. Some examples of the applications that are likely to be useful are as follows:

1. **Advanced Intelligent Network Features:** Use advanced intelligent network (AIN) features such as Caller ID, voice mail, call waiting, pre-and-post paid calling cards, call blocking, and auto call-back in IP telephony.
2. **Voice Calls from Mobile Laptop PCs:** Call office or home, from hotel, airport, etc. using multimedia laptop PCs with wireless connection to Internet. This could be ideal for submitting or retrieving voice messages.
3. **Airlines Reservations:** Use a Java applet to visually display interactive voice response options rather than forcing users to wait through very long recorded instructions and go through multi-level menus requiring the use of a telephone keypad.
4. **Internet-aware Telephones:** Use enhanced ordinary telephone (wired or wireless) as an Internet access device as well as for normal telephony. Directory services, for example, could be accessed over the Internet by submitting a name and receiving a voice (or text) reply.
5. **Voice Annotated Ddocuments:** Send voice messages and voice annotated documents to integrated voice/data mailboxes. Voice annotated documents and multimedia files can easily become standard within office suites in the near future.
6. **Internet Call Center Access:** Access customer service agents online over an Internet call center.
7. **Virtual Call Centers:** Support the integrated voice and data requirements of call center agents working from their homes.
8. **Live Auction Websites:** Create live audio auction websites for excess inventory. Use Java applets on the phone to manage the bidding process and to track who “raised a hand” to bid first, etc.
9. **Presence and Instant Messaging:** Use instant messenger service to determine when geographically distributed colleagues are available for a quick conference call with a customer.
10. **Electronic Business Cards:** Send an enriched electronic virtual business card (vCard) including photo and audio file automatically with every call as caller ID information (or selectively during the middle of call).
11. **Integrated Voice and Data Information:** Integrate voice and data information collected during a call with sales force automation applications.
12. **Personalized Music On-hold:** Play personalized announcements or music from a favorite MP3 recording or Internet radio station while callers are on hold.

# MARKET TRENDS

IP telephony is becoming a key driver in the evolution of voice communications. While it is currently a small fraction of telecommunication market, it is growing quickly. Here is a selection of what market researchers think about IP telephony:

- According to IDC's January 2003 forecast, worldwide sales of IP telephony equipment will see substantial growth between 2004 and 2007, increasing by 48% in 2004 and topping $4.9 billion by year's end. Equipment sales will continue to grow by more than 50% in 2005 and 2006, to reach $11.4 billion in 2006, and $15.1 billion in 2007 (Figure 2).
- Businesses spent an estimated $2.0 billion on IP-capable telephone systems in North America in




Source: International Data Corporation (IDC), January 2003

**Figure 2: Worldwide IP Telephony Equipment Revenues, 2001-2007 (in Billions & % Growth)**

2003, according to Gartner Dataquest, which also predicts that enterprises spending on IP phone systems will more than double during four years, to reach $4.2 billion in 2007.

- According to Gartner, the IP telephone systems accounted for just 1.4% of total business telephony equipment sales in 1999. However, in 2003 IP telephony equipment increased its market share to 56% of all sales, and by 2007, it is expected to account for 97% of all business telephony sales.
- The IP PBX market is expected to grow to $3.9 billion in revenue by 2005, representing nearly 20% of all traditional PBX sales. (Synergy Research, February 2002)
- By 2005, analysts predict 34% of all telephone calls made worldwide will be carried over the Internet, accounting for over 90 billion minutes of telephone communication.
- Wholesale and retail VoIP traffic volume exceeded 6 billion and 15 billion minutes in 2000. VoIP will account for approximately 75% of world voice services by 2007. (Frost & Sullivan, May 2001)
- VoIP will account for approximately 75% of world voice services by 2007. (Frost & Sullivan, March 2002)
- By 2008, wholesale VoIP traffic in the Europe, Middle East and Asia (EMEA) region could reach 57 billion minutes. About 1,771 billion minutes of retail voice traffic originating in EMEA will have a VoIP component in some portion of the route. (Frost and Sullivan, May 2002)
- According to a study conducted by Insight Research Corporation, from a mere $13 billion in 2002, voice over packet (VoIP)-based services will grow to nearly $197 billion by 2007. Also, VoIP services will grow at a compounded rate of over 72 percent over the period 2002-2007, making packet voice services one of the fastest-growing segments in the telecommunications industry (Press Release – "Phone Industry Survival hinges on Explosive Growth of VoIP" – October 28, 2002)
- Enterprises will migrate their voice systems from traditional networks to data networks at a rate that will create a $16.5 billion dollar IP-PBX market worldwide by 2006. (Allied Business Intelligence.)
- 90% of enterprises with multiple locations will start switching to IP systems for voice over next 5 years. (Phillips Group, via Aspect, June 2001)
- The European IP Virtual Private Network (VPN) services market to grow from 1.74 billion at the end of 2002 to 3.58 billion by 2005, when almost 40% of the identified potentials within the European market will have implemented an IP VPN solution. (Frost and Sullivan, January 2003)

# IP TELEPHONY PROTOCOL ARCHITECTURE

Like any conversation between two computing devices, IP telephony requires an agreed upon set of rules, called "*protocol*". Figure 3 shows an architectural diagram of the protocols that are used to govern IP calls. These protocols follow a layered hierarchy which can be compared with the Open Systems Interconnect reference model (OSI 7-layer model), developed by the International Standards Organization (ISO). Although the IP telephony model doesn't exactly match OSI structure, it is useful for discussion.

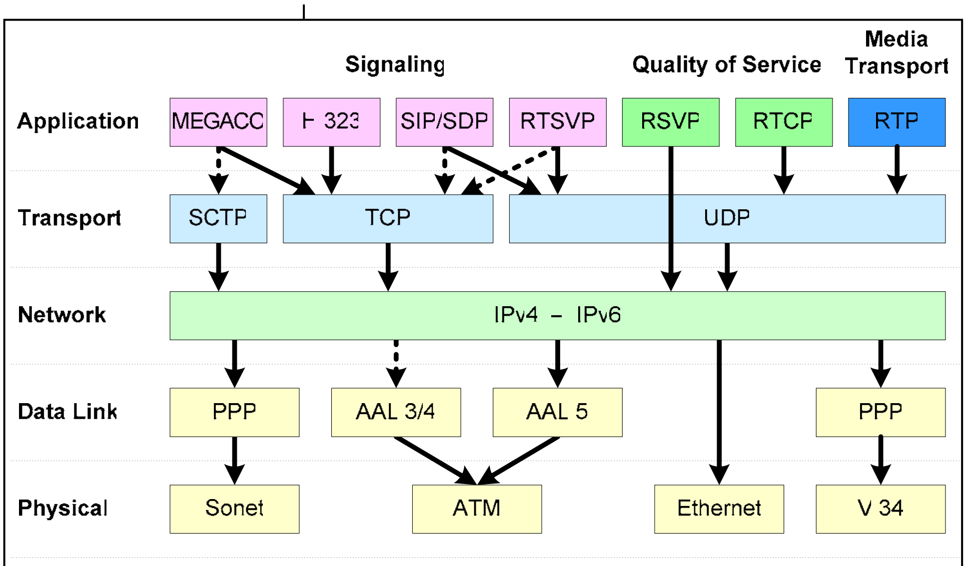



**Figure 3: IP Telephony Protocol Suite in Relation to the ISO Model**

Breaking a system into defined layers can make that system more manageable and flexible. A layer defines a specific data communication function that may be performed by any number of protocols. Each layer has its job, and does not need a detailed understanding of the layers around it. For example, IP datagram (or packet) can be transported across a variety of physical layer systems including serial lines, Ethernet and ATM. The data is passed down the stack when it is being sent to the network and up the stack when it is being received from the network.

The physical and link layer protocols provide the means for the system to deliver data to the other devices on a directly attached network. Unlike higher-level protocols, these protocols must know the details of the underlying network (its packet structure, addressing, etc.) to correctly format the data being transmitted to comply with the network constraints. The physical layer protocol is for the most part irrelevant to IP and need not be the same for first link and final link of a VoIP call.

The major protocols involved in IP telephony, started at the Network layer are discussed in the following sections:

## INTERNET PROTOCOL (IP)

The Internet Protocol (IP) is a network layer protocol and it is the heart of the IP telephony. It provides the basic packet delivery service on which IP telephony networks are built. All protocols, in the layers above and below IP, use the Internet Protocol to deliver data.

IP is responsible for defining the datagram (or packet) and moving data between link layer and transport layer. It is also responsible for routing datagrams to remote hosts and performing fragmentation and re-assembly of the datagrams. IP is a connectionless protocol, that is, it does not establish an end-to-end connection through a network before transmitting data. IP relies on protocols in other layers to establish the connection if they require connection-oriented service.

IP makes no guarantees concerning reliability, flow control, error detection or error correction. This means that the packets could arrive at the destination computer out of sequence, with errors or may not arrive at all. IP just transports the data to a higher layer and relies on protocols in the other layers to provide error detection and error correction.

## TRANSMISSION CONTROL PROTOCOL (TCP)

The Transmission Control Protocol (TCP) is a transport layer protocol and is responsible for delivering data received from IP to the correct application. The application that the data is bound for is identified by a 16-bit number called the port number (for example, the HTTP application is usually associated with port 80). TCP uses 16-bit source-port and destination-port numbers in segment header, to deliver data to the correct applications process.

TCP is a connection-oriented protocol. It establishes a logical end-to-end connection between the two communicating hosts before transmitting data. TCP also provides a reliable data delivery across the network using a mechanism called Positive Acknowledgment with Re-transmission (PAR). Simply stated, a system using PAR sends the data again, unless it hears from the remote system that the data arrived okay.

TCP also handles sequencing and error detection, ensuring that a reliable stream of data is received by the destination application. It views the data it sends as a continuous stream of bytes, not as independent packets. Therefore, TCP takes care to maintain the sequence in which bytes are sent and received.

Although the TCP works smoothly with data – any packet not delivered is simply re-transmitted, it does not work well with real-time voice and video applications. It is because any word received out of sequence within the structure of a sentence will result in a garbled message.

## USER DATAGRAM PROTOCOL (UDP)

User Datagram Protocol (UDP) is also a transport layer protocol and is responsible for the transmission of information between the correct applications on the host computers. The UDP uses 16-bit source-port and destination-port numbers in the message header, to deliver data to the correct applications process.

Like IP, UDP is a connectionless protocol as it does not establish an end-to-end connection between two communicating hosts before transmitting data. It routes data to the correct destination port, but does not attempt to perform any sequencing, or to ensure data reliability. The UDP gives application programs direct access to a datagram delivery service, like the delivery service that IP provides. This allows applications to exchange messages over the network with a minimum of protocol overhead.

For voice applications, UDP ensures that the information is received in correct sequence, reliably and with predictable delay characteristics. UDP also performs certain functions that TCP cannot perform. Hence, it is commonly used for IP telephony.

## REAL-TIME TRANSPORT PROTOCOL (RTP)

Real time applications, such as voice and video, require mechanisms to be in place to ensure that a stream of data can be reconstructed accurately. In IP networks, the connection-oriented TCP protocol guarantees an error-free transmission in the right order but it is not appropriate for real-time applications. For instance, with video, if a packet arrives late, it loses its meaning and may not be inserted correctly in the clip being played. For this reason, UDP is used for voice and video transmission.

Jitter is the variation in delay times experienced by the individual packets making up the data stream. In order to reduce the effects of jitter, data must be buffered at the receiving end of the link so that it can be played out at a constant rate. However, the UDP has no control over the order in which packets arrive at the destination or how long it takes them to get there. Both of these are very important to overall voice and video quality. Real-time transport protocol (RTP) solves the problem by enabling the receiver to put the packets back into the correct order and not wait too long for packets that have either lost their way or are taking too long to arrive. RTP is really just the underlying transport plane, and the other protocols use it to successfully transport voice and video packets.

RTP is designed to provide end-to-end network transport functions for applications transmitting real-time data, such as voice and video, over IP networks. It carries data source and payload type information, and is itself carried inside of UDP. RTP provides the sequence number and time stamp information needed to assemble a real time data stream from packets. However, RTP does not have any mechanisms for ensuring the on-time delivery of traffic signals or for recovering lost packets. Therefore, RTP does not reduce the overall delay of the real time information. Nor does it make any guarantees concerning quality of service.

## RTP CONTROL PROTOCOL (RTCP)

The RTP Control Protocol is the counterpart of RTP that provides control services. The RTCP monitors the QoS and conveys information about the participants in an ongoing session. It provides feedback on total performance and quality so that modifications can be made. Other RTCP functions include carrying a transport-level identifier for an RTP source which is used by receivers to synchronize audio and video.

## RESOURCE RESERVATION PROTOCOL (RSVP)

The Resource Reservation Protocol is a signaling protocol that supports the reservation of resources across an IP network. It manages quality of service by requesting a certain amount of bandwidth and latency in every network hop that supports it. Applications running on IP end systems can use RSVP to indicate to other nodes the nature (bandwidth, jitter, maximum burst, and so forth) of the packet streams they want to receive. The RSVP depends on IPv6.

# IP TELEPHONY SIGNALING PROTOCOLS

Several standards are available for building IP telephony solutions. These include H.323, Session Initiation Protocol (SIP), Media Gateway Control Protocol (MGCP) and Media Gateway Control (Megaco). A high-level comparison of these protocols is included in Table 1.

Of the protocols listed in Table 1, only SIP and H.323 are peer-to-peer protocols. MGCP and Megaco represent the old centralized model and suffer from this model's limitations. Thus, the real choice for a protocol with Web-like benefits comes down to one of the peer-to-peer protocols – H.323 or SIP.

The H.323 and SIP are the two major protocols that are used by VoIP technology to define ways for devices (telephones, computers, etc.) on the data network to communicate with each other. H.323 is a comprehensive and very complex protocol. It provides specifications for real-time, interactive videoconferencing, data sharing and IP telephony. The Session Initiation Protocol (SIP) emerged as an alternative to H.323. It is a much simpler, more streamlined protocol developed specifically for IP telephony. It is smaller and more efficient than H.323 and takes advantage of existing protocols to handle certain parts of the process.

Both of these protocols do essentially the same things; these provide a way for the caller to find the called party (call construction), allow each party to send streams of audio data to the other party, and provide a way for either party to end the call (call tear-down). Both of these protocols also include specifications for audio (and video)

**Table1: Comparison of IP Telephony Protocols**

| | H.323 | SIP | MGCP | Megaco/H.248 |
|---|---|---|---|---|
| Standardization | ITU-T | IETF | IETF | IETF and ITU-T |
| Architectural model | Peer-to-peer | Peer-to-peer | Master/ slave | Master/ slave |
| Media types | Voice, video, limited data | Voice, video, data | Voice | Voice, video |
| Call control | Gatekeeper | Proxy/Redirect Server | MGC | MGC |
| Endpoints | Gateway, terminal | User agent | Media gateway | Media gateway |
| Signaling transport | TCP or UDP | TCP or UDP | UDP | TCP and UDP |
| Network scope | Intranet, Extranet, Internet | Intra-, Extra-, Internet | Intranet only | Intranet only |
| Extensibility | Low | High | Medium | Medium |
| Scalability | Medium | High | Low | Low |
| Ease of deployment | Low | High | Medium | Medium |
| Standardization | ITU-T | IETF | IETF | IETF and ITU-T |

codecs (coder-decoder) to convert audio signal into a compressed digital form for transmission and back into an uncompressed audio signal for replay. These protocols are discussed in detail in the following sections:

# H.323

The H.323 is an umbrella recommendation from the Telecommunication Standardization Sector of the International Telecommunications Union (ITU-T) that specifies the components, protocols and procedures that provide multimedia communication services (real-time voice, video, chat, whiteboard, file sharing, etc.) over packet–based networks, including IP. By conforming to H.323 standards, multimedia products and applications from different vendors can interoperate across IP based networks, including the Internet. H.323 is part of a family of ITU-T recommendations called H.32x that provides multimedia communication services over a variety of networks.

H.323 covers both protected and unprotected connections. Control and data information requires a protected transmission to prevent packets from being lost or not received in the right order. In IP-based networks, TCP protocol guarantees an error-free transmission in the right order but causes delays and has a lower throughput. Therefore, unprotected connections are used for audio and video transmissions, which are more efficient.

The H.323 standard's mandatory components are transmission of audio, connection control according to Q.931, communication with the gatekeeper over the RAS protocol, and use of the H.245 signaling protocol; the rest of the text, including coverage of the ability to transmit video and data, is optional. Although H.323 uses TCP to carry the signaling channels, the real-time media streams are transported on RTP/RTCP (discussed earlier). RTP carries the actual media and RTCP carries status and control information.

Being the first widely available VoIP protocol, H.323 enjoyed a head start as developers implemented it as toll-bypass systems as well as PC-to-phone and video-conferencing applications. The best-known H.323 application was Microsoft NetMeeting.

## *History*

The Version 1 of the H.323 recommendation was accepted in October 1996. It was heavily weighted towards multimedia communications in LAN environment and does not provide guaranteed quality of service. With the development of VoIP, new requirements emerged, such as providing communication between a PC–based phone and a phone on a traditional Switched Circuit Network (SCN). Such requirements forced the need for a standard for IP telephony. Version 2 of H.323, packet-based multimedia communications systems, was defined to accommodate these additional requirements and was accepted in January 1998.

H.323 Version 3, approved on September 30, 1999, and Version 4, approved on November 17, 2000, only makes modest improvements to the Version 2 Recommendation, introducing only a few new powerful features to the base document. The last Version 5 was approved at the end of July 2003. Unlike previous revisions of the recommendation, Version 5 aimed to maintain stability in the protocol by introducing only modest additions to the base protocol, rather than introducing sweeping changes as was the case in prior revisions.

## *H.323 Architecture Components*

The H.323 standard specifies a number of components (entities), which, when networked together, provide the point-to-point and point-to-multipoint multimedia-communication services. Some components are mandatory, while others are optional. The four most important components are listed below:

**Terminal:** An H.323 terminal is an endpoint on a network which provides two-way communications with another terminal, gateway or a Multipoint Control Unit (MCU). An H.323 terminal can either be a personal computer or a stand-alone device such as IP telephone running H.323 and the multimedia applications. It supports audio communications and can optionally support video or data communications.

**Gateway:** An H.323 gateway provides connectivity between an H.323 network and a non–H.323 network. For example, it may route Voice over IP (VoIP) calls from an H323 terminal to the public switched telephone network (PSTN). This connectivity of dissimilar networks is achieved by translating protocols for call setup and release, converting media formats between different networks, and transferring information between the networks connected by the gateway. A gateway is not required, however, for communication between two terminals on an H.323 network.

**Gatekeeper:** A gatekeeper provides basic admission control onto a network by allowing or refusing communications between other H.323 entities within its zone of control. They also provide call-control services for H.323 endpoints, such as address translation (to use name instead of IP address), authentication, accounting and bandwidth management. Gatekeepers in H.323 networks are optional. If they are present in a network, however, terminals and gateways must use their services.

**Multipoint Control Unit (MCU):** An MCU provide services that allow three or more endpoints to take part in a conference call. All terminals participating in the conference establish a connection with the MCU. The MCU manages conference resources, negotiates between terminals for the purpose of determining the audio or video coder/decoder (codec) to use, and may handle the media stream.

## H.323 Protocols

H.323 is a comprehensive and very complicated protocol. As shown in Table 2, it ties together a number of existing recommendations defined by International Telecommunications Union (ITU) and Internet Engineering Task Force (IETF). These protocols, together with some other recommendations, provide specifications for a range of communication including real-time voice, video and data transmission (*see* Figure 4). An overview of these protocols is given below:

**Table 2: H.323 Protocol Suite**

| Video | Audio | Data | Transport |
|---|---|---|---|
| H.261<br>H.263 | G.711<br>G.722<br>G.723.1<br>G.728<br>G.729 | T.122<br>T.124<br>T.125<br>T.126<br>T.127 | H.225<br>H.235<br>H.245<br>H.450.1<br>H.450.2<br>H.450.3<br>RTP<br>X.224.0 |

**G.7xx Audio Codecs:** The recommendations G.711 (audio coding at 64 kbps), G.722 (64, 56, and 48 kbps), G.723.1 (5.3 and 6.3 kbps), G.728 (16 kbps), and G.729 (8 kbps) define the way in which analogue audio signals are encoded into compressed digital form for transmission, and decoded back into an uncompressed audio signal for replay. Because audio is the minimum service provided by the H.323 standard, all H.323 terminals must have at least one audio CODEC support.

**H.26x Video Codecs:** The recommendations H.261 and H.263 define methods for encoding analog video into digital form for transmission and decoding back into analog video code for replay. Because H.323 specifies support of video as optional, the support of video codec is optional as well.

**H.225 Call Signaling:** H.225 call signaling is used to set up connections between H.323 endpoints (terminals and gateways), over which the real-time data can be transported. Call signaling involves the exchange of H.225 protocol messages over a reliable call-signaling channel. For example, H.225 protocol messages are carried over TCP in an IP–based H.323 network. The call-signaling channel is opened between two H.323 endpoints or between an endpoint and the gatekeeper.

**H.225 RAS:** H.225 registration, admission, and status (RAS) is the protocol between endpoints (terminals and gateways) and gatekeepers. The RAS is used to perform registration, admission control, bandwidth changes, status, and disengage procedures between endpoints and gatekeepers. A RAS channel is used to exchange RAS messages.

**H.245 Control Signaling:** H.245 control signaling is used to exchange end-to-end control messages governing the operation of the H.323 endpoint. The messages carried include messages to exchange capabilities of terminals and to open and close logical channels. The H.245 control messages are carried over H.245 control channels.

**Q.931:** Q.931 is a call signaling protocol and is used for establishing H.323 calls. H.225 call control messages are embedded within the user-to-user elements of Q.931 messages to provide additional information not available in Q.931 such as IP address information.

**RTP:** Real-Time Transport Protocol (RTP) provides end-to-end network transport functions for applications transmitting real-time data over IP networks. It provides services such as payload type identification, sequence numbering, time-stamping, and delivery monitoring to real-time applications. RTP can also be used with other transport protocols. RTP is discussed in detail in an earlier section.

**RTCP:** Real-time Transport Control Protocol (RTCP) is the counterpart of RTP that provides control services. The primary function of RTCP is to provide feedback on the quality of the media received using RTP. RTCP is discussed in detail in an earlier section.

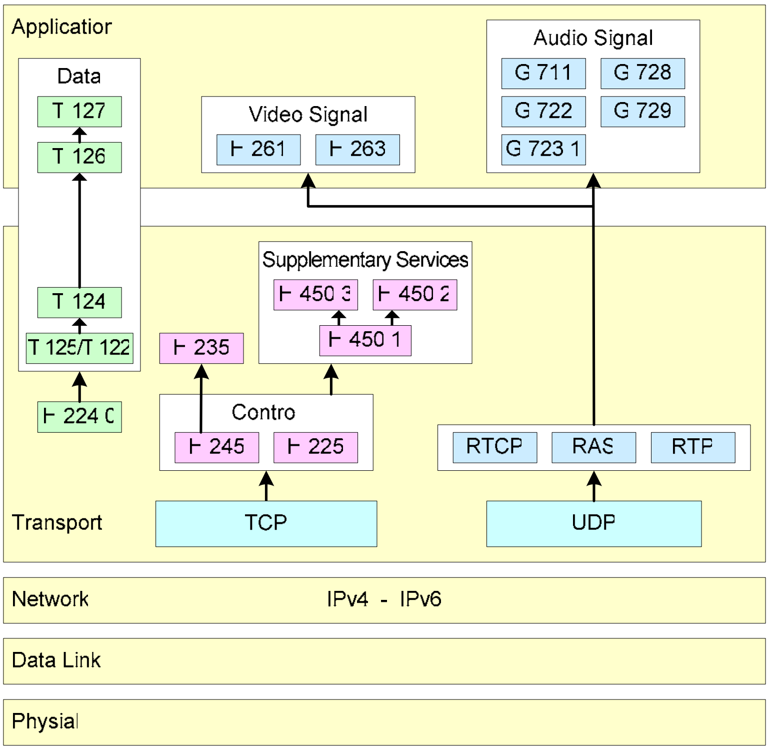



**Figure 4: H.323 Protocol Stack**

# SESSION INITIATION PROTOCOL (SIP)

The Session Initiation Protocol (SIP) emerged as an alternative to H.323, under the auspices of the Internet Engineering Task Force (IETF). SIP is a much more streamlined and powerful protocol, developed specifically for IP telephony. Smaller and more efficient than H.323, SIP takes advantage of existing protocols to handle certain parts of the communication process. For example, Media Gateway Control Protocol (MGCP) is used by SIP to establish a gateway connecting to the PSTN system.

Since H.323 was originally designed for video conferencing over private, high speed LAN, it assumes details like authorization and user identification. H.323 also provides all these extra services which are not usually considered useful for a simple phone call. SIP, however, was designed from the beginning for multimedia sessions and conferences over Internet and wide area network (WAN). Because of these differences in their design objectives, SIP offers numerous compelling advantages in the areas of extensibility, scalability, and ease of deployment over H.323.

SIP is transport layer independent. It can run over any datagram or stream protocol such as UDP, TCP, ATM, etc. It makes use of the Session Description Protocol (SDP) for specification of the session parameters. The audio or video data streams are transported using RTP over UDP. SIP may use any IANA registered codec while H.323 requires ITU-T defined standard only.

SIP enables new services and applications not possible with H.323 and other IP telephony protocols. For example, SIP uses a simple text-based encapsulation (based on the Internet standard MIME) which enables it to transmit data and application programs with the voice call, making it easy to send files, photos, and MP3 encoded information during a call. It also enables developers to push the intelligence to the edge of the networks, implement a distributed architecture, and create advanced features.

Being peer-to-peer protocols, both SIP and H.323 eliminate the need for central servers to control everything.

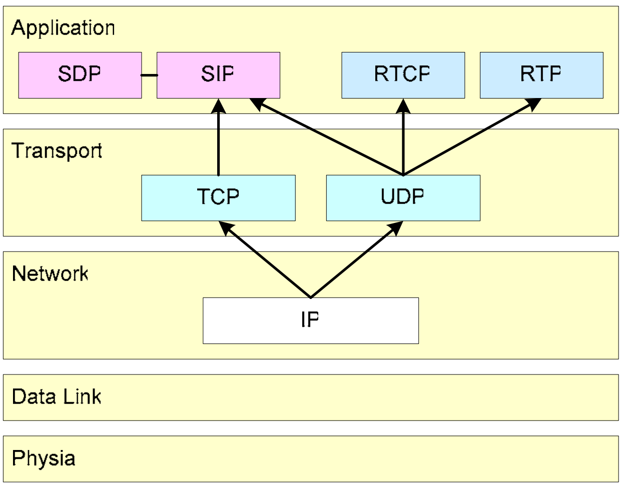



**Figure 5: SIP Protocol Stack**

However, within peer-to-peer protocols, SIP is a much more efficient and less complex protocol, therefore, more scalable than H.323. A high level SIP protocol stack is shown in Figure 5.

## *SIP Concepts*

**Session:** Session is the basic building block in SIP. A SIP session is a multimedia session consisting of a set of multimedia senders and receivers, and the data streams flowing from senders to receivers. All calls and conferences are established by setting up sessions among users.

**Conference:** A conference is a multimedia session, identified by a common session description. A conference can have zero or more members and includes the cases of a multicast conference, a full-mesh conference and a two-party "phone call", as well as combinations of these. Any number of calls can be used to create a conference.

**Call:** A call consists of all participants in a conference invited by a common source. A SIP call is identified by a globally unique call-ID.

## *SIP Architecture Components*

SIP specification defines a number of components that are required to develop a SIP-based network. In many implementations, some of these components are combined into the same software modules.

### SIP User Agents

A SIP user agent (UA) is a program that runs on a SIP device such as IP phones and gateways. It contains a client function and a server function.

The user agent client (UAC) initiates SIP requests such as initiating a call. It is the only entity on a SIP-based network that is permitted to create an original request. The UAC is also known as the calling user agent.

A user agent server (UAS) is one of many server types that receives SIP requests such as an incoming call and sends back responses to those requests. A UAS is also known as the called user agent. Normally, user agents are discussed without any distinction made between their UAC and UAS components.

### SIP servers

The SIP servers are distinguished by their roles played by centralized hosts on a distributed network. There are four types of SIP servers that can be implemented in a user agent (UA). These are as follows:

**Location Server:** A location server is used to obtain information about called party's possible location. A location is the IP address of the domain where a user is located. To locate a user, the name of the user is sent to the location server and the location server returns zero or multiple locations (IP addresses of domains) where a called party may be found.

**Proxy Server:** A proxy server is an intermediary program that acts as both a server and a client for the purpose of making requests on behalf of other clients. Requests are serviced internally by a proxy server or transferred to other servers.

**Redirect Server:** A redirect server is a server that accepts a SIP request, maps the address into zero or more new addresses and returns these addresses to the client. Unlike a Proxy, it cannot accept calls but can generate SIP responses that instruct the UAC to contact another SIP entity.

**Registrar Server:** A registrar server is a server that accepts REGISTER requests. A client uses the REGISTER request to let a proxy or redirect server know the location where the client can be reached. It provides a means whereby users can register their locations with a SIP server dynamically.

## MEDIA GATEWAY CONTROL PROTOCOL (MGCP)

Media Gateway Control Protocol (MGCP), endorsed by the Internet Engineering Task Force (IETF), is a protocol for handling the signaling and session management needed during a multimedia conference. It is used for controlling telephony gateways from external call control elements called media gateway controllers or call agents. A telephony gateway is a network element that provides conversion between the audio signals carried on circuit-switched network (such as PSTN) and data packets carried over packet-switched networks (such as Internet).

MGCP assumes signaling control intelligence outside the gateways, in a media gateway controller (MGC). MGCP makes it possible for the MGC to determine the location of each communication endpoint and its media capabilities so that a level of service can be chosen that will be possible for all participants. MGCP can be used to set up, maintain, and terminate calls between multiple endpoints.

The MGCP specifies a protocol at the Application layer level that uses a master-slave model, where the gateways are expected to execute commands sent by the media gateway controllers. Two Media Gateway Controllers use RTP to talk to one another and successfully transport voice packets.

MGCP is well suited for centralized systems that work with dumb endpoints, such as analog phones. The most celebrated use of MGCP is for high-capacity gateways designed to work with traditional telecom equipment.

## MEGACO / H.248

Megaco/H.248 is the Media Gateway Control Protocol defined jointly by IETF and ITU-T for use in distributed switching environments. The standard is endorsed by the IETF as Megaco and by the ITU-T as Recommendation H.248.

The Megaco/H.248 protocol was developed from the Media Gateway Control Protocol (MGCP) – it refers to an enhanced version of MGCP. Megaco/H.248 provides broadly equivalent functionality and has a very similar structure. The later Megaco/H.248 version supports more ports per gateway, as well as multiple gateways, and support for Time-Division Multiplexing (TDM) and Asynchronous Transfer Mode (ATM) communication.

# VOICE QUALITY CHALLENGES AND SOLUTIONS

Traditional telephony networks are built to provide an optimal service for time-sensitive voice applications; these provide constant but low bandwidth services. However, IP networks were built to support non-real-time applications such as email, file transfer and web information. These networks send and receive data only as needed, in data packets. Because each packet is individually routed across the network, this makes packet-switching networks inherently less efficient in dealing with voice traffic and poses a number of challenges to a quality voice transmission.

## CHALLENGES

It has been found that there are three factors that can profoundly impact the quality of the service. These are:

**Delay (Latency):** Delay in a voice network is the time required for a voice signal to traverse the network. Two problems that results from high end-to-end delay are "*echo*" and "*talker-overlap*". Echo becomes annoying when the delay is more than 50 ms. It is perceived as a significant quality problem. Talker-overlap can cause the receiver to start to talk before the sender is finished. It becomes unacceptable if the one-way delay becomes greater than 450 ms. Delay is inherent in data networks where it has no real impact, but with voice it is necessary that Quality of Service (QoS) features are implemented right across the IP network. Sources of delay in a packet voice call include the collection of voice samples (accumulation delay), encoding/decoding and packetizing time, jitter buffer delays, and network transit delay. IP telephony gateways and terminals also contribute significantly to delay.

**Jitter (Delay Variability):** Jitter is the variation in inter-packet arrival time. In IP networks, packets that belong to the same transmission do not always arrive with the same amount of delay. This variation in delay is referred to as "*jitter*." The jitter causes gaps in the speech pattern; as a result, voice transmissions may sound unnatural.

**Packet Loss:** IP networks do not provide guaranteed delivery of the packets. The packets can be dropped under peak loads and during periods of congestion (caused, for example, by link failures or inadequate capacity). For non-real-time applications, such as email and file transfers, packet loss is not critical – the protocol allows retransmission to recover dropped packages. However, real-time voice data has to arrive within a certain time window to be useful to reconstruct the voice signal.

## SOLUTIONS

In order to deal with these issues and provide a voice service with a reasonable measure of quality, there are many techniques that are employed to deal with network congestion and delay. These techniques include the following:

**Prioritization:** Prioritization is a method of guaranteeing throughput for certain traffic on the network. This can ensure that voice traffic on a data network is given high priority. This prioritization can be based on location, protocol or application type. RSVP is designed to ensure this QoS.

**Fragmentation:** Fragmentation divides the packets into smaller fragments so that their priority can be ensured. This can help reduce the overall delay of voice delivery. However, this can create extra overhead because of the large size of IP headers (20 bytes). So although necessary, fragmentation alone cannot ensure the reliable delivery of real-time voice applications.

**Jitter Buffering:** Jitter buffering is a technique that allows packets to be collected into a buffer and held there long enough for the slower packets to arrive so that they can all be played in proper sequence. Although this can remove packet delay, this creates additional overall delay. The best compromise is to fit jitter buffer in the network's differential delay. This will provide the necessary balance between the packet delay and the overall delay.

**Silence Suppression:** Silence suppression is a technique that is used to suppress the transmission of silence. These techniques take advantage of half-duplex nature of human conversation (one person listens while the other talks) by detecting when there is a gap and then suppresses the transmission of these silences. This can amount to 50-60% of the time of a call, resulting in considerable bandwidth conservation. However, because these silences are necessary for the conversation to sound natural, the receiving device must interpret the lack of packets and re-insert the silent spots, called *comfort-noise*, into the output.

**Echo Cancellation:** Echo cancellation is used to cancel the echo caused by end-to-end delay of a voice transmission. Echo cancellers monitor speech from the far end that passes through its receive-path and use this information to compute an estimate of the echo that is then subtracted from its send-path. ITU protocols G.165 and G.168 define the performance requirements that are currently required for echo cancellers.

**Interpolation of Speech:** Interpolation of speech is an approach used to compensate for packet loss. If certain voice packets are delayed beyond a specific threshold, IP telephony software interpolates by re-playing the last packet, and sending of redundant information. In order to help ensure a quality voice conversation, packet losses greater than 10% are not tolerable. For packet losses under 10%, interpolation can help maintain a continuous flow of voice with minimal distraction to the quality.

**Speech Codecs:** A speech codec transforms analog voice into digital bit-streams, and vice versa. In addition, some speech codecs also use compression techniques to reduce the transmission bandwidth required. Most PSTN networks use ITU-T G.711 recommendation that encodes the speech at 64 kbps. Current ITU-T recommendations include codecs that compress to as low as 5.3 kbps, although quality at this bit rate is well below that of the highest quality G.711. Essentially, compression is a balancing act between voice quality, local computation power, delay, and network bandwidth required. In general, the lower the bit rate, the lower the quality perceived by the listener. However, more modern codec designs are driving up the quality for a given bit rate. The more recent, though computationally intensive codecs, GSMEFR at 12.2 kbps and G.728 at 16 kbps, are comparable to the low-complexity G.726 codec that offers good performance only at 32 kbps or above.

In order for different manufacturers to implement these various techniques and maintain interoperability, the standards like H.323, SGCP, SAP, SIP, RTSP and SDP have been recommended and approved.

## CONCLUSION

Data traffic has traditionally been forced to fit onto the voice network. The Internet has created an opportunity to reverse this strategy – voice, video, fax and other multimedia can now be carried over IP networks.

IP telephony users enjoy free long distance phone calls, coupled with numerous additional features. Both traditional telephone features, such as call-waiting and voice mail are included, as well as non-traditional features, such as group chats and text-based document sharing. Although the quality isn't as good as conventional phone calls, the discounted prices make up for it.

Although the price of IP call is now negligible, the phone companies will be likely to object to the free long distance service offered by the Internet and may raise the price of local phone calls in response. It remains to be seen whether or not IP telephone calls will continue to be such a good bargain to the average user, since the pricing for voice traffic is now undergoing change.

Voice communications will certainly remain a basic form of interaction for all of us. The public switched telephone network simply cannot be replaced, or even dramatically changed, in the short term. The immediate goal for IP telephony developers is to reproduce existing telephone capabilities at a significantly lower "*total cost*" and to offer a technically competitive alternative to the PSTN.

As data traffic continues to increase and surpass that of voice traffic, the convergence and integration of these technologies will not only continue to improve, but will also pave the way for a truly unified and seamless means of communication. The economics of placing all traffic (voice, video and data) over an IP network will pull companies in this direction, simply because IP will act as a unifying agent, regardless of the underlying architecture (e.g., leased lines, frame relay, or ATM) of an organization's network.

The market for IP telephony products is established and is beginning its rapid growth phase. The use of off-the-shelf software and hardware components can allow for a rapid implementation and a great degree of flexibility in the implementation.

The Internet and its underlying IP protocol suite have become the driving force for new technologies. Future extensions will include innovative new solutions including conference bridging, voice and data synchronization, combined real-time and message-based services, text-to-speech conversion and voice response systems.

# REFERENCES


[1] *"VoIP Howto"*, v 1.8 - August 26, 2002, http://www.bertolinux.com/voip/english/VoIP-HOWTO.html

[2] *"An Introduction to VoIP and VOCAL"*, by Luan Dang, Cullen Jennings and David G. Kelly, 10/11/2002

[3] *"VoIP In A Nutshell"*, Website, http://www.novastars.com

[4] *"Protocols.com"*, Website, http://www.protocls.com

[5] *"How IP Telephony Works"*, Jeff Tyson, http://computer.howstuffworks.com/ip-telephony1.htm

[6] *"The Growth of Internet Telephony: Legal and Policy Issues"*, Emir A. Mohammed, http://www.firstmonday.dk/issues/issue4_6/mohammed/#m1

[7] *"Voice over Internet Protocol"*, http://www.iec.org/online/tutorials/int_tele/index.html

[8] *"VoIP"*, http://members.tripod.com/nguyen225/page1.htm

[9] *"Factors in the Success of Voice Quality in Converging Telephony and ip Networks"*, Stefan Pracht

[10] *"Internet Telephony and Voice Compression"*, Kelly Ann Smith and Daniel Brushteyn, http://www.cis.upenn.edu/~kellyann/papers/iphone.html

[11] *"Voice over Internet Protocol (VoIP)"*, Balz Wyss, Microsoft Corporation, March 2003

[12] *"Quality VoIP - An Engineering Challenge"*, R J B Reynolds and A W Rix, BT Technol J Vol 19 No 2 April 2001

[13] *"The Rise of Internet Telephony"*, http://www.nortelnetworks.com

[14] *"Feature Interaction in Internet Telephony"*, Jonathan Lennox, Henning Schulzrinne, Columbia University

[15] *"Next-Gen VoIP Services and Applications Using SIP and Java"*, Pingtel, http://www.techguide.com

[16] *"Voice over IP (VoIP)"*, Jerry Ryan, Telogy Networks, http://www.techguide.com

[17] *"Next Generation Telephony: A Look at Session Initiation Protocol"* Thomas Doumas, May 1999, Hewlett Packard.