HERMON LABORATORIES

VoIP Overview

The Voice over IP (VoIP) telephone networks become more and more widely used. These networks deliver voice, facsimile, and/or voice-messaging communications services via the Internet, rather than the public switched telephone network (PSTN). Originally designed and well suited for non real-time data communication the internet technology imposes design and testing challenges for VoIP equipment manufactures to ensure good quality of real-time service (QoS) such as voice.

TCP/IP

Transmission Control Protocol/Internet Protocol (TCP/IP) is a set of interrelated protocols, or what generically is called a protocol stack. IP is a Network Layer (Layer 3) and TCP is Transport Layer (Layer 4) in the OSI reference model. TCP and IP are two protocols out of several protocols comprising the TCP/IP stack.
The initial intent of TCP/IP was to link dissimilar computer networks, independent of computer hardware and operating systems, transmission media and data link technologies.

IP

IP -Internet Protocol basically addresses and sends packets (also called datagrams) in a connectionless mode. Connectionless means that there is no predetermined path through the network established as part of a call setup process. Rather each packet works its way through the network independently and is considered as a separate unit by the network elements. As a result each packet may follow a different path from one host to another across the network, depending on congestion levels, blockings and hardware failures. Although this approach does not seem to be performance-oriented, it uses efficiently the resources of a highly shared network.
IP provides no packet numbering, no guarantee for packet delivery and no inherent error control mechanism.
Each packet contains payload data and a header field which provides signaling and control mechanism.
The total maximum packet size, according to IP version 4, is 65,536 octets (bytes) and the minimum is 576 octets.
The most important fields included in the header are:
Protocol identifies the higher-layer protocol that is running on top of IP and have created the payload data of the IP packet.
Examples of such higher-level protocols are TCP or UDP.
Source and Destination addresses each are 32-bit fields identifying source and destination hosts. The IP addresses are generally represented in dotted decimal notation, which separates the four bytes of the address with periods.
An IP address looks like this: 102.52.94.93. The IP address comprises the network ID, and the host ID within the network.
The subnet mask, also represented in dotted decimal notation, is used to extract these two values from an IP address. The subnet mask is defined on each host. In the binary representation, the subnet mask consists of a sequence of ones followed by a sequence of zeros. The network ID is a result of logical ANDing of the IP address with the subnet mask. The host ID is extracted by NOTing the subnet mask and then ANDing the result with the IP address. For example a 255.255.255.0 subnet mask implies that the network ID is 102.52.94 and the host ID is 93, given that IP address is 102.52.94.93.
Service Type can be used to define Quality of Service parameters such as priority, throughput and reliability. While IP networks, as opposed to ATM networks do not provide guaranteed QoS, they can be designed to support a best-effort Grade of Service (GoS).
Time To Live (TTL) specifies the time in seconds, or more commonly, the number of hops through network nodes the packet can survive. At each hop at a network node the TTL number is decremented until it reaches 0, at which time the packet is discarded. This scheme prevents errored, duplicate or misdirected packets from endlessly travelling around the network and contributing to network congestion.
How does an IP packet addressed to host in another part of the world finds its way to the destination?
The hosts connected in the same LAN, see every packet that is sent by each host or the packets are switched directly to the destination host. Normally, a host will only do something with that packet if it carries the host destination address, or if destination is a broadcast address.

When a packet should pass through different networks (as in case of WAN), a router (the equipment used to connect different networks) compares the packet's destination address with the table of addresses stored in its memory. If it finds a match it sends the packet to the associated address, which can represent another network or next-hop router.

TCP

TCP protocol is one option that can run on top of IP. TCP is connection-oriented, meaning that it establishes a virtual (logical) connection between two systems before they exchange data. TCP takes the data from the sending host, breaks the data into pieces, called segments, adds a header with sequence number and checksum to each segment and passes the segments to IP for transmission to destination.
On the receive side, TCP uses the sequence numbers to rearrange the segments when they arrive out of order, and to eliminate duplicate segments. Also the receiving TCP performs the checksum calculation and compares the result with the received checksum value. If the result is correct the receiver returns positive acknowledgement (ACK) to the sender, together with number of bytes it can accommodate in its buffer beginning from the last received byte.
If the sender does not receive ACK within a timeout interval, the data is retransmitted.
Such mechanism provides data flow control and achieves reliability.
The TCP header includes also source port and destination port fields. The port number identifies which higher level protocol sent (or is to receive) the data. For example, the Telnet protocol uses port number 23. The Simple Mail Transfer Protocol (SMTP) uses port 25, HTTP (HyperText Transport Protocol) uses port 80 and FTP (File Transfer Protocol) uses port 21.

An IP address and a port number taken together uniquely identify a service running on a host, and the pair is known as a socket.
Two processes, communicating over TCP, are said have a logical connection that is uniquely identifiable by the transmit and receive sockets involved.

UDP

User Datagram Protocol (UDP) is another protocol that can run on top of IP.
Unlike TCP, UDP is connectionless and has no support for sequencing, error control, or flow control.
UDP adds four 16-bit header fields (8 bytes) to each datagram sent. These fields are: a length field, a checksum field, and source and destination port numbers.
Although UDP isn't reliable, it is still an appropriate choice for many applications that can benefit from its simplicity and low overhead. It is used in real-time applications like Voice over IP and video where, if data is lost, it's better to do without it than send it again out of sequence.
Since UDP is a Network Layer (Layer 3) it uses Data Link Layer (Layer 2) to interface with the Physical layer.
In Ethernet networks IEEE 802.3 layer 2 is used. For dial-up connections such as analog PSTN and ISDN modem connections PPP (Point to Point Protocol) and SLIP (Serial Line IP) can be used. Also UDP can run on top of ATM and Frame Relay.

Voice over IP

Voice over IP (VoIP) allows the human voice and fax information to travel over an IP data network concurrently with traditional data packets. This data network can be LAN, Internet or WAN connecting several branch offices. The VoIP service can be also provided by connecting via PSTN to an IP network.
The main benefit of VoIP is cost reduction due to use of toll-free internet connections and effectiveness of packet-switched IP networks (as compared to circuit-switched networks).
Also VoIP offers great advantages due to the fact that voice, fax, data and video can be integrated over the same IP-based network and through the same terminal in the form of integrated voice/data/video/multimedia client workstations, running various call control, productivity, unified messaging and data sharing applications.
In most cases VoIP services need to be able to connect to traditional (mainly circuit-switched) public or private network via dial-up (analog/ISDN BRI/ISDN PRI) , E1/T1 PCM, Frame Relay, ATM, DSL lines or connect to wireless network.
H.323 protocol stack addresses this goal by a series of standards for packet-based multimedia networks.
The basic elements of H.323 network include terminals, gateways and optional gatekeepers.
H.323 terminals are LAN-based end-points (such as PC running Microsoft NetMeeting or IP telephones).
The H.323 terminals convert between voice speech and data packets sent and received over LAN interface, by using at least one voice CODEC. Commonly used voice CODECs are ITU-T G.711 (PCM), G.723 (MP-MLQ), G.723.1, G.726, G.729A (CA-ACELP) and GSM. CODEC originally stood for Coder/Decoder, now it is frequently means Compressor/Decompressor.
Terminals can also implement video and data communication capabilities. Common video CODECs are MPEG, H.261, H.263, H.264.
Terminals also need to support signalling functions that are used for call setup and tear down and so forth. The applicable protocols are H.225.0 signalling which is a subset of ISDN's Q.931 signalling or the newer SIP (Session Initiation Protocol). RAS protocol manages registration, admission, status. H.245 negotiates channel usage and capabilities. H.235 deals with security and authentication.

Media streams (voice, video, data) are transported on RTP/RTCP over UDP. RTP carries the actual media and RTCP carries status and control information. The signalling is transported reliably over TCP.

Figure 1. VoIP transmission datagram.

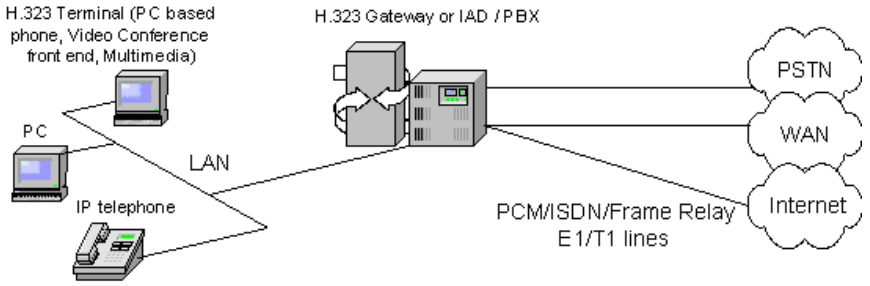
IP Header	UDP Header	RTP Header	Payload Data
20 octets	8 octets	16 octets	20 octets

64-octet Datagram

RTP provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio, video or simulation data, over multicast or unicast network services. RTP does not address resource reservation and does not guarantee quality-of-service for real-time services. The data transport is augmented by a control protocol (RTCP) to allow monitoring of the data delivery in a manner scalable to large multicast networks, and to provide minimal control and identification functionality.

The gateway serves as the interface between the H.323 (e.g. LAN-based) and non-H.323 network (e.g. Public network). As the interface, the gateway needs to translate signalling messages between the two sides as well as compress and decompress the voice.
In the IP world, voice is treated as another application running over the data network, and the PBX can be viewed as an application server residing on that same network.

Figure 2. Voice over IP network.



In order to provide sufficient quality of voice, fax and video real-time and full-duplex communications, the VoIP services should handle the problems such as latency, jitter and packet loss which are inherent to IP (packet-switched) networks.
This can be achieved by increased network bandwidth, voice compression and packet reassembly algorithms implemented in CODECs, prioritization of voice over data transmission, header compression, intelligent routing and balancing of various compression and transmission parameters.
Latency, jitter and packet loss are collectively referred to as Quality of Service (QoS).

Latency, which is the time it takes the packets to go from source to destination, is caused in IP networks by disassembly into packets, buffering and compression at the sending CODEC; network delay which depends on network congestion level; buffering, decompression and processing at the receiving CODEC.
Latency (also called round trip delay) in excess of 200 ms - 250 ms may be noticeable to callers.
ITU-T G.114 specifies maximum 150 ms latency, required to achieve good quality conversation.
Variation of latency in time is called jitter. Buffering is used at the receive end to suppress jitter. The jitter buffer deliberately delays incoming packets in order to present them to the decompression algorithm at fixed time intervals using the time-stamp field in the RTP header of the received packets. The jitter buffer will also fix any out-of-order errors by looking at the sequence number in the RTP frames. On the other hand too deep jitter buffer can introduce intolerable latency.

Table 1. Popular Voice CODEC comparison.

CODEC scheme	Standard (ITU-T)	Description	Compressed rate	MOS (Mean Opinion Score) - voice quality	Imposed delay
PCM	G.711	Pulse Code Modulation. This standard a-law or u-law PCM codec with no compression.	64 kbps (no compression)	4.4 MOS	0.75 ms
ADPCM	G.723	Adaptive Differential Pulse Code Modulation	40/32/24/16 kbps	4.2 MOS	1 ms
LD-CELP	G.728	Low Delay Code Excited Linear Prediction	16 kbps	up to 4.2 MOS	3-5 ms
CS-ACELP	G.729	Conjugate Structure Algebraic Code Exited Linear Prediction	8 kbps	up to 4.2 MOS	10 ms

Packet loss can be caused by several factors, such as network links overload, collisions on a LAN and hardware failure.
The logic embedded in predictive decompression algorithms of CODECs such as CS-ACELP and LD-CELP will take advantage of the delay (about 10 ms) compression/decompression process to make the necessary predictions and perform the interpolations that reduce voice discrepancies. Also various techniques are used for echo cancellation, as echo becomes perceptive when delay exceeds 15 ms.

CODECs can deal with packet loss ratio up to about 5%. Above this ratio or when long continuous losses occur the speech quality degradation becomes noticeable.
In some cases RTP header compression is employed, to reduce the number of packets sent. This can be achieved because some header information is redundant in the lower-level protocols.
Another VoIP technique to reduce the number of packets is based on detection of silence periods during the conversation.

The ITU-T P.800 proposed a method for overall voice quality grading. The P.800 MOS (Mean Opinion Score) grades the voice quality on the scale 1 (worst) to 5 (best) based playing of pre-recorded voice samples over the transmission media to a mixed group of men and women under controlled conditions. A MOS of 4 is considered "toll-quality" voice.
ITU-T P.861 Perceptual Speech Quality Measurement (PSQM) tries to automate this process by defining an algorithm through which a computer can derive scores that have a close correlation to the MOS. The PSQM seems to have some limitations, because it overlooks such important VoIP parameters as jitter and frame loss. The PSQM+ (ITU-T) is an enhanced version of PSQM which adds some post-processing on PSQM values and correlates more to MOS.
British Telecom Perceptual Analysis/Measurement system (PAMS) tries to address PSQM and PSQM+ limitations.
Tests conducted by BT have shown good correlation between automated PAMS scoring and manual MOS results.
The PESQ (Perceptual Evaluation of Speech Quality) defined in ITU-T P.862 combines PAMS time alignment routine with PSQM+ perceptual model.
The PESQ algorithm models the human perception of speech, by comparing a reference speech signal with the "degraded" signal. The reference signal is transmitted by the test instrument to the EUT and the degraded output signal from the EUT is measured.
The most important result is the ITU-T P.862 PESQ MOS (Mean Opinion Score), because it directly expresses the voice quality. The PESQ MOS ranges from 1.0 (worst) up to 4.5 (best).
Additional results such as Delay and Delay Jitter, G.107 rating R factor, various waveforms, level, gain, loudness, SNR, VAD (Voice Activity Detection) parameters - Front End Clipping, Hold Over Time and Drop-outs are also provided.
PESQ can replace an array of lengthy "traditional" TIMS voice frequency measurements.