2.1 Introduction

In this section can provide the necessary background on the structure of VoIP applications and on their component, and the transmission protocols generally used in VoIP.

2.2 Voice over IP (VoIP)

VoIP means that calls are transmitted over an IP network such as the Internet instead of Public Switched Telephone Networks. Since access to the Internet is available at more and more places in the world, it is possible to use VoIP in a higher degree. VoIP converts standard telephone voice signals into compressed data packets that can be sent over IP. Before transmitted over packet switched networks, the speech signal has to be digitized at the sender; the reverse process is performed at the receiver. The digitalization process is composed of sampling, quantization and encoding [6].

The spread of Internet and its underlying communication protocols IP gave rise to the notion of everything over IP. One of the applications that is experiencing high growth and popularity is Voice over IP (VoIP). Voice over IP is gaining success because of multiple reasons:

- **Lower Equipment Cost:**

PSTN PBXs cost millions of dollars and are very slow to change to allow the addition of new features. IP network components are less expensive and enjoy higher interoperability allowing equipment's to be sourced from multiple vendors who are very competitive reducing the cost of these equipment's. Also, the cycle of rolling out new features is counted in months not years as in the case of switching centers.

- **Integration Of Voice And Data:**

The use of one network to carry both voice and data allows savings of management and operational manpower, operational costs and the efficient use of communication links between different sites. Also the integration of voice and data allows the creation of new sets of applications that make use of both. For example: click to talk, voice mail, video conferencing.

- **The Widespread Availability Of IP:**

IP networks are widely available geographically across continents and within most countries. This allows most people to have access to a PC linked to the Internet. Also the availability of gateways to/from PSTN allows calls to use VoIP even for a portion of call, the initiating end, the terminating end, or an intermediate link. For example, a transoceanic link can use VoIP to maximize utilization of the expensive bandwidth.

2.2.1 VoIP Network Components

Voice over IP networks in general is composed of four different types of components: End stations, Servers, IP network and Interface to PSTN if needed.

The end stations initiate and maintain the signaling required to establish calls over the IP network, and convert voice to data packets and vice versa. Servers enable call establishment and support additional features. For example, a SIP location server allows users to forward calls to a different location.

Links to PSTN (Gateways) allow the interface between VoIP network and PSTN networks, if needed. Gateways perform two main tasks:

a) Convert signaling used to establish, tear down and maintain a call between PSTN and VoIP protocols

b) Convert voice samples at the PSTN side (8000 samples/sec) to VoIP packets and vice versa.

Gateways perform a multiple of other functions such as the exchange of billing information and SS7 network interface. SS7 network is a management and control network used to manage and control PSTN networks.

The Signaling Transport (sigtran) working group of the IETF defined protocols to provide all functionality needed to support SS7 signaling over IP networks, including:

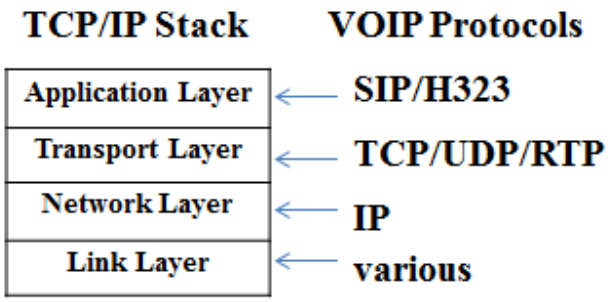
- Flow control and in-sequence delivery of signaling messages
- Identification of the originating and terminating signaling points
- Error detection, retransmission and other error correcting procedures
- Recovery from outages of components in the transit path
- Controls to avoid congestion on the Internet
- Detection of the status of peer entities
- Extensions to support security and future requirements

IP networks provide the infrastructure linking all the components together and the necessary routing of calls and data packets between one end of the network to another. An IP Network could be a LAN or a combination of LAN and WAN links connecting two stations at the far ends of the globe.

2.2.2 VoIP Signaling Protocols

Most VoIP signaling protocols run over TCP/IP networks, which provide a full reliable transfer of data packets between clients or between clients and servers. The transfer of real-time packets (RTP protocol) is carried over UDP, which does not provide a loss-less packets transfer between the two ends of the link, because re-sending lost packets is unnecessary since they usually arrive too late to be used in the voice stream. VoIP uses signaling protocols such as Session Initiation Protocol (SIP) or H.323 for establishing, modifying and tearing down unicast or multicast sessions consisting of one or several media streams [7].

Different standards are emerging to specify VoIP protocols. The following are the main standards used in this area: SIP, H.323, and MGCP. A brief introduction is included hereafter for the two most popular protocols (SIP and H.323). Figure (2-1) gives a high-level view of the SIPand H.323 protocols and their interaction with the TCP/IP stack. Traditional VoIP protocols, such as Session Initial Protocol (SIP) and H.323 (lTU recommendation), work in a centralized manner [8].



Figure(2-1): VoIP protocols over TCP/IP stack

The Session Initiation Protocol (SIP) is an ASCII-based, peer-to-peer application layer protocol that defines initiation, modification and termination of interactive, multimedia communication sessions between users [9].

SIP is developed by the Internet Engineering Task Force (IETF) and is derived from Hyper-text Transfer Protocol (HTTP) and Simple Mail Transfer Protocol (SMTP). SIP is defined as a client-server protocol, in which requests are issued by the calling client and responded to by the called server, which may in itself be a client for other aspects of the same call. SIP is not dependent on TCP for reliability but rather handles its own acknowledgment and handshaking. This makes it possible to create an optimal solution that is highly adjusted to the properties of VoIP.

The ITU-T recommended H.323 protocol show in figure (2-2) below suite has evolved out of a video telephony Standard [10] H.323 is known for quite complex

signaling, high connection setup latencies, and implementation difficulties. However, H.323 is widely implemented and is the primary common denominator for all VoIP.

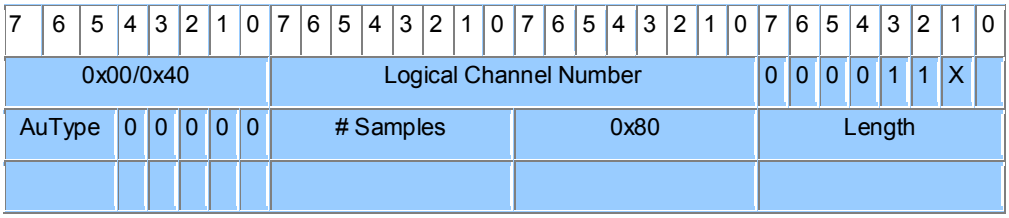


Figure (2-2): H.323 Header

Logical Channel Number : The number of the H.245 logical channel 1.

AuType : The audio codec to be used.

Samples : The number of samples - 1 per Audio packet as defined in ITU-T Rec. H.245.

SIP and H.323 provide similar functionality: Call control, call setup and teardown, basic call features such as call waiting, call hold, call transfer, call forwarding, call return, call identification, or Call Park, and capabilities exchange. Each protocol exhibits strengths in different applications. H.323 defines sophisticated multimedia conferencing which can support applications such as white boarding, data collaboration, or video conferencing.

SIP supports flexible and intuitive feature creation with SIP using SIP-CGI (SIP Common Gateway Interface) and CPL (Call Processing Language). Third party call control is currently only available in SIP. Work is in progress to add this functionality to H.323.

2.2.3 VoIP Challenges

The use and adoption of VoIP is faced by a multitude of challenges resulting from two main factors. First, the Internet was not designed to transfer real time data; it is a best effort network. Network equipment drop packets and may have queues that cause jitter in packets transfer delays. Also routing is more time consuming when compared to switching. Network delays and loss of packets affect the quality of service of VoIP [11].

These factors are discussed in more detail later. Multiple efforts are undergoing in various directions to reduce or eliminate those QoS variables, among which are Reservation Protocols (RSVP), design separate high priority queues for real time traffic and the use of a mix between routing and switching (MPLS) to speed up packets through routing points.

Second, PSTN has grown and added multiple features that are different to emulate. The main challenge that exists presently is the support of 911 emergency services. Emergency workers responding to a 911 call can determine the exact location of the originator of the call, because of the tight relation between telephone number and geographic locations. Calls originating from a VoIP client are very difficult to correlate to a geographic location due to the lack of geographic structure in IP addresses, and the dynamism of IP networks. Several efforts are undergoing to solve this problem but it still possesses a great challenge [12].

In addition, there is a growing concern about the privacy and security of VoIP conversations. As discussed, with the ability of capturing voice packets using a network sniffer, eavesdropping is easier in VoIP networks than it is in PSTN. Using wireless networks combined with VoIP further complicates these VoIP challenges.

2.4 VoIP over Transmission Protocols

Generally, there are many protocols available at the transport layer when transmitting information through an IP network. These are TCP (Transmission Control Protocol), UDP (User Data-gram Protocol) and (Real-Time Transmission Protocol).

2.4.1 VOIP over TCP

TCP stands for Transmission Control Protocol and is one of the main protocols used for data transmission over the Internet and LANs. It works together with the IP protocol to make the well-known TCP/IP protocol suite. Since other protocols like IP do not provide reliability over a network, TCP ensures that data transmission is reliable. It ensures that, during a transmission, there is no packet loss; there is an acceptable delay between the packets. TCP also bundles data into TCP packets. The data packets however do not contain an address for the source and destination machines, since IP packet stake care of the addressing and routing [13].

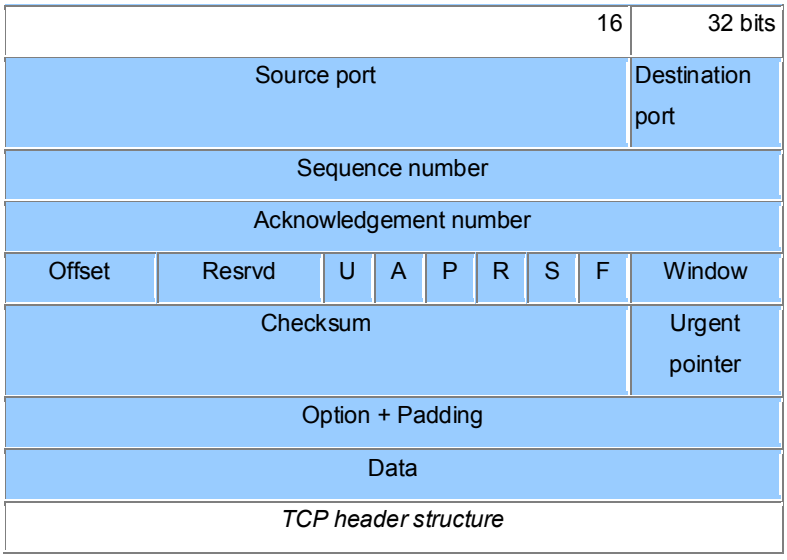


Figure (2-3) : TCP header

Source port : Source port number.

Destination port : Destination port number.

Sequence number : The sequence number of the first data octet in this segment (except when SYN is present). If SYN is present, the sequence number is the initial sequence number (ISN) and the first data octet is ISN+1.

Acknowledgment number : If the ACK control bit is set, this field contains the value of the next sequence number which the sender of the segment is expecting to receive. Once a connection is established, this value is always sent.

Data offset : 4 bits. The number of 32-bit words in the TCP header, which indicates where the data begins. The TCP header (even one including options) has a length which is an integral number of 32 bits.

Reserved : 6 bits. Reserved for future use. Must be zero.

Control bits : 6 bits. The control bits may be (from right to left):

U (URG) Urgent pointer field significant.
A (ACK) Acknowledgment field significant.
P (PSH) Push function.
R (RST) Reset the connection.
S (SYN) Synchronize sequence numbers.
F (FIN) No more data from sender.

Window : 16 bits. The number of data octets which the sender of this segment is willing to accept, beginning with the octet indicated in the acknowledgment field.

Checksum : 16 bits. The checksum field is the 16 bit one's complement of the one's complement sum of all 16-bit words in the header and text. If a segment contains an odd number of header and text octets to be checksummed, the last octet is padded on the right with zeros to form a 16-bit word for checksum purposes. The pad is not transmitted as part of the segment. While computing the checksum, the checksum field itself is replaced with zeros.

Urgent Pointer : 16 bits. This field communicates the current value of the urgent pointer as a positive offset from the sequence number in this segment. The urgent pointer points to the sequence number of the octet following the urgent data. This field can only be interpreted in segments for which the URG control bit has been set.

Options : Options may be transmitted at the end of the TCP header and always have a length which is a multiple of 8 bits. All options are included in the checksum. An option may begin on any octet boundary.

There are two possible formats for an option:

- A single octet of option type.
- An octet of option type, an octet of option length, and the actual option data octets.

The option length includes the option type and option length, as well as the option data octets.
The list of options may be shorter than that designated by the data offset field because the contents of the header beyond the End-of-Option option must be header padding i.e., zero.

A TCP must implement all options.

Data : TCP data or higher layer protocol.

The connection is formed via a handshake between two hosts with connection requests and acknowledgments. Once the connection is formed, the data being transmitted is broken into segments. Before the segment is transmitted, a header is attached which contains a sequence number. The receiver will respond to the arriving packet with an acknowledgement if no errors are found. If no acknowledgement arrives at the original sender after a certain timeout period, the sender will re-transmit the packet [14].

2.4.2 VOIP over UDP

The UDP is a simple protocol that passes data along from the application layer to IP to be transmitted. It performs none of the error checks that TCP does, and is therefore unreliable.

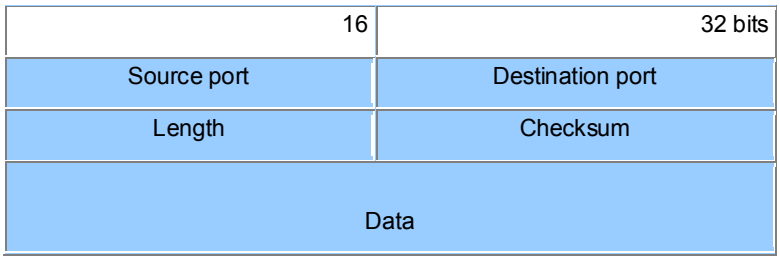


Figure (2-4) : UDP header

Source port : Source port is an optional field. When used, it indicates the port of the sending process and may be assumed to be the port to which a reply should be addressed in the absence of any other information. If not used, a value of zero is inserted.

Destination port : Destination port has a meaning within the context of a particular Internet destination address.

Length : The length in octets of this user datagram, including this header and the data. The minimum value of the length is eight.

Checksum : The 16-bit one's complement of the one's complement sum of a pseudo header of information from the IP header, the UDP header and the data, padded with zero octets at the end (if necessary) to make a multiple of two octets.

Data : UDP data field.

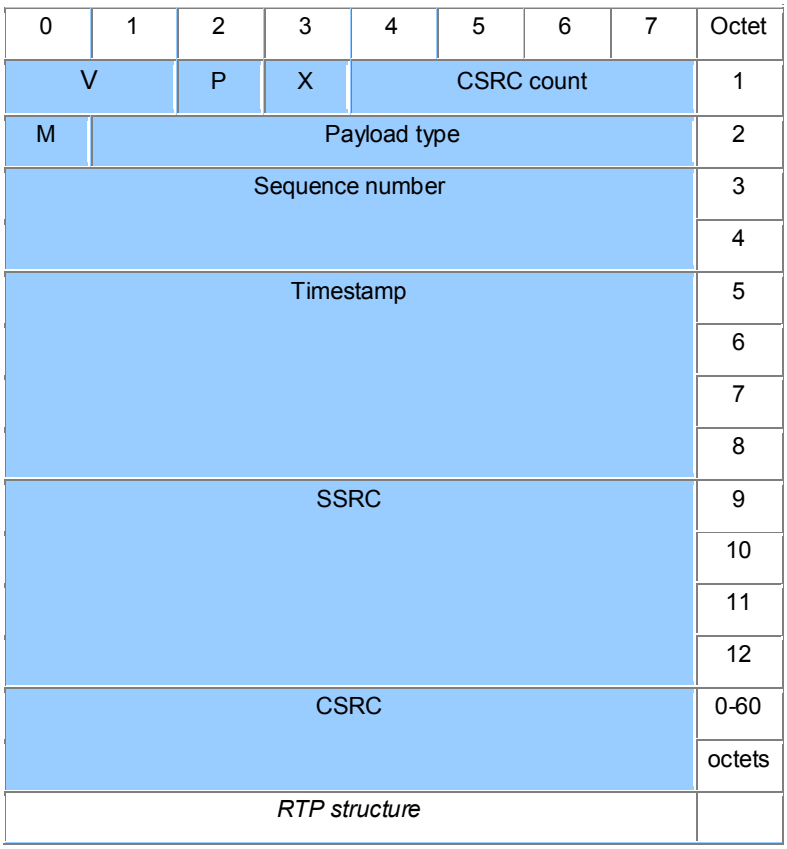
A UDP header merely consists of an optional source port, a destination port, the length of the datagram, and a checksum [15]. As previously mentioned, the main reason for using UDP over TCP in VoIP applications is the reduced delay. In general, the sporadic loss of packets in a conversation will not be as disruptive as excessively long delay times. In fact, a packet loss of about 5% is said to be tolerable depending on how the losses are distributed [16]. Investigate how well a UDP-based VoIP network performs in contrast to its TCP counterpart.

2.4.3 VOIP over RTP

The RTP is an application layer protocol that attaches itself to UDP to provide added benefits for real-time applications [17] Applications typically run RTP on top of UDP[18]. The RTP packets contain the audio and video elementary streams associated with the selected program and information about the standard used for the compression[19].

(RTP) was developed for the transportation of real-time multimedia, such as VoIP service. Traditionally, VoIP applications use the protocol stack of RTP/UDP/IP to

convey voice data. As well known, UDP is a connectionless transport protocol widely used in multimedia services [20].



Figure(2-5): RTP Header

V :Version. Identifies the RTP version.

P:Padding. When set, the packet contains one or more additional padding octets at the end which are not part of the payload.

X: Extension bit. When set, the fixed header is followed by exactly one header extension, with a defined format.

CSRC count :Contains the number of CSRC identifiers that follow the fixed header.

M: Marker. The interpretation of the marker is defined by a profile. It is intended to allow significant events such as frame boundaries to be marked in the packet stream.

Payload type: Identifies the format of the RTP payload and determines its interpretation by the application. A profile specifies a default static mapping of payload type codes to payload formats. Additional payload type codes may be defined dynamically through non-RTP means.

Sequence number: Increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence.

Time stamp: Reflects the sampling instant of the first octet in the RTP data packet. The sampling instant must be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations. The resolution of the clock must be sufficient for the desired synchronization accuracy and for measuring packet arrival jitter (one tick per video frame is typically not sufficient).

SSRC: Identifies the synchronization source. This identifier is chosen randomly, with the intent that no two synchronization sources within the same RTP session will have the same SSRC identifier.

CSRC: Contributing source identifiers list. Identifies the contributing sources for the payload contained in this packet

An RTP header includes a sequence number to help preserve the order of the transmitted packets. It also includes a timestamp, which is meant to provide information to the destination application so that it may compensate for problems such as delay or jitter if they arise. The optional companion protocol, RTCP

(specified in RFC 3550), is used as a means of exchanging information on session quality, which can include the number of lost packets or the average delay time . RTP is the protocol of choice for streaming media over the Internet and is widely used in VoIP applications [21]. RTP is typically run on top of UDP to make use of its multiplexing and checksum functions. TCP and UDP are two most commonly used transport protocols on the Internet. TCP provides a connection-oriented and reliable flow between two hosts, while UDP provides a connectionless but unreliable datagram service over the network. UDP was chosen as the target transport protocol for RTP because of two reasons. First, RTP is primarily designed for multicast, the connection-oriented TCP does not scale well and therefore is not suitable. Second, for real-time data, reliability is not as important as timely delivery. Even more, reliable transmission provided by retransmission as in TCP is not desirable. For example, in network congestion, some packets might get lost and the application would result in lower but acceptable quality. If the protocol insists a reliable transmission, the retransmitted packets could possibly increase the delay, jam the network, and eventually starve the receiving application [22].

2.5 Previous Studies

Sreekanth Asodi. [17] Said that SCTP and UDP complete with each other under the considered quality metrics for voice transmission. The good performance of UDP in VOIP applications makes it a preferred transport layer protocol to carry voice packet from source to destination. However, it`s likely that SCTP many preform batter with some modification/extensions in the as observed, it performance is comparable to UDP in most of the cases.

Camarillo et al. [23] implements SIP over SCTP, UDP and TCP protocols under different network conditions and observe that UDP is good only for the light traffic. Under heavy traffic load, TCP and SCTP are better than UDP. However,

SCTP has some advantage over TCP owing to its features as multi-streaming and multi-homing. In general, SCTP performance increases with worsening of network conditions [17].

Pallavi Gangurde.[24] compare the performance of TCP, UDP and SCTP with traffic analysis. And kept the packet size of 1000 bytes and run the simulation with constant bit rates over all transport protocols and the channel capacity 0.2 Mb. On comparing the results it seen that TCP is performing the best with least number of packet loss as compared to that of SCTP and that of UDP. SCTP is best effort because its multi homing and multi association but its packet delivery acknowledgement is time consuming and researches are required to be done in future so that SCTP would be more advantages over TCP. Also more or less satisfactory performance is observed in competing traffic with UDP and SCTP, although UDP has an edge being free from all sorts of transport overheads. But in the case of packet loss, where SCTP suffers a bit of delay variations UDP suffers from the effect of application layer retransmission. With increasing effect of packet loss the performance of SCTP undergoes a severe degradation.
UDP on the other hand keeps a consistent behavior as the packet drop has no effect on its application. Same rate of packet loss in SCTP causes packets drops at transport layer and delays increase in a consistent manner. So it is very easily observable that SCTP has no comparison with UDP. Serious considerations are required to be made in the future regarding protocol redesign that can be best suitable to carry VoIP signaling messages.

2.6 Quality of Service (QoS)

Quality of Voice as IP was designed for carrying data, so it does not provide real time guarantees but only provides best effort service. QoS represents the set of techniques necessary to manage network bandwidth, delay, jitter, and packet loss.

If the network bandwidth is not enough, even high-priority traffic may not get through. Traffic engineering, which enables QoS, is about making sure that the network can deliver the expected traffic loads [26]. For voice communications over IP to become acceptable to the users, the delay needs to be less than a threshold value and the IETF (Internet Engineering Task Force) is working on this aspect. To ensure good quality of voice, but can use either Echo Cancellation, Packet Prioritization (giving higher priority to voice packets) or Forward Error Correction [27].

Quality of Service (QoS) refers to the concept of being able to control and measure data transmission rates, or throughput, and error rates. Specifically, QoS refers to the ability of a network to provide better, more predictable service to selected network traffic over various underlying technologies, including IP-routed networks. Multimedia applications producing high data rates hence have the potential to cause network congestion. This can disrupt both the users' quality of experience, and potentially the operation and stability of the network for other customers [28].

Traditionally networks did not require strict measures for QoS because the data wasn't multimedia and the end-user could not notice or be materially affected by latencies. But, as the use of network spread far beyond simple data transfer to intense multimedia applications, the need to address Quality of Service (QoS) issues becomes extremely important. Both the enterprise and consumer markets are now beginning to demand data intensive, time-sensitive movement of things like audio and video around a network.

2.7 VoIP QoS Factors

There are many factors or parameters affecting voice quality in a VoIP network. These parameters are complicatedly related to other and affect voice quality

[29].Voice applications have different characteristics and requirements from those of traditional data applications. Because they are innately real-time, voice applications tolerate minimal delay in delivery of their packets. Additionally, they are intolerant of packet loss, out-of-order packets, and jitter. To effectively transport voice traffic over IP, mechanisms are required that ensure reliable conveyance of packets with low and controlled latency. Thus the primary goal in the context of VoIP QoS, then would be to provide dedicated bandwidth, controlled jitter and latency (required by some real-time and interactive traffic), and improved loss characteristics.

2.7.1 Throughput

Throughput is the average amount of data that is successfully delivered over a communication link. The maximum throughput is less than or equal to the amount of digital bandwidth. It is measured by bits/bytes per seconds [23]. Throughput for VoIP traffic varies based on the type of codec being used and the amount of compression applied. This represents the number of packets received with in a given time interval. Data transmission from Node A to Node B, throughput refers to the total amount of data, received at Node B [30]. As an example, for simulation Node 0's throughput was measured in kbps as follows:

$$\text{Throughput node A} = \frac{\text{byte received node B}}{\Delta t * 10^3}$$

The event type "r" in the first column of the trace was the primary tool used to track throughput.

2.7.2 Packet Loss

Lost packets is the failure of one or many packets of data travelling across the network to reach their destination. Packet loss is one of the important error types in digital communications [31]. Approaches used to compensate for packet loss include interpolation of speech by replaying the last packet and sending redundant information. Packet losses greater than 10 percent are generally intolerable, unless the encoding scheme provides extraordinary robustness. Packet loss a measure of the amount that lost between the source and the destination in the network. Two measures were taken:

1. Instantaneous packet loss: describes how many packets are lost at each time interval.
2. Cumulative loss: describes the total number of packets loss of all time.

2.7.3 End To End Delay

The time delay incurred in speech by the Internet Protocol (IP) telephony system. One-way delay is the amount of time measured from the moment the speaker utters a sound until the listener hears it. Round trip time is the sum of the two one-way delay figures that compose the user's call. The lower the delay, the more natural interactive conversation becomes; accordingly, the additional delay incurred by the VoIP system is less noticeable. delay is measured by the time taken by voice packets to travel between two end points. delay occurs when packets of data take longer than expected to reach their destination and causes some problems in voice quality[32].When coders/decoders (codecs) in VoIP terminals compress voice signals they introduce three types of delay: Processing, or algorithmic, delay – the time required for the codec to encode a single voice frame, Look ahead delay – the time required for a codec to examine part of the next frame while encoding the

current frame (Most compression schemes require look ahead) and Frame delay the time required for the sending system to transmit one frame.

In general, End to End delay is the time taken for a packet to reach from source to destination.[33] it can be seen that greater levels of compression introduce more delay and require lower network latency to maintain good voice quality. Most VoIP sessions require one-way latency of not more than about 200 milliseconds. This delay budget is reduced by any delays introduced by codecs in the end systems. When round-trip delays exceed approximately 300 ms., natural human conversation becomes difficult.

The equation below is used to calculate the delay on the packet:

$$D = T_f + l + \sum_{h \in Path} (T_h + Q_h + P_h) + D_{play}$$

Where:

D : end-to-end delay
T_f : formation delay
l : look a head
T_h : transmission delay
Q_h : queuing delay
P_h : propagation delay
D_{play} : playout buffer delay

The end-to-end delay is measured as follows:

1. Obtain the time when a packet is created/transmitted

2. Obtain the time when the same packet reached its destination

3. Take the difference of the two intervals.

The procedure is performed for all transmitted/received packets within a time interval and the resulting differences are averaged.

Delay introduces two other difficulties echo and talker overlap. Echo is caused by the signal reflections of the speaker's voice from the far end telephone equipment

back into the speaker's ear. Echo becomes a significant problem when the round trip delay becomes greater than 50 milliseconds. Since echo is perceived as a significant quality problem, Voice over Packet systems must address the need for echo control and implement some means of echo cancellation. Secondly, Talker overlap (or the problem of one talker stepping on the other talker's speech) becomes significant if the one-way delay becomes greater than 250 msec. The end-to-end delay budget is therefore the major constraint and driving requirement for reducing delay through a packet network.

To support VoIP traffic consistently and reliably, a network must therefore be able to provide three things:

- Packet-forwarding latency that does not exceed the maximum tolerable level for a VoIP conversation.
- Packet-forwarding jitter, which is the variation in latency over time that does not exceed the maximum tolerable level to sustain a VoIP session.
- Guaranteed network bandwidth and capacity for VoIP sessions during periods of network congestion.

In other words, a network needs to provide performance – low latency and low jitter – and protection – to maintain stringent measures of quality of service.

From the above discussion of QoS, it can be seen that Assurance of Quality of Service is critical for proper operation of a VoIP network. The evolution of IP based applications place more stress and require more sophistication in equipment designed to support these applications over real world networks, while delivering services at similar reliability levels to those experienced over non-integrated traditional networks. The ability to have a flexible mechanism that enables a user to tailor the QoS policy to his specific needs is a critical component of an overall integrated network. The time taken by a packet to reach from a source to destination, delay can be occurred from different sources like delay at source,

delay at receiver, delay in network. Delay at source and receiver is due to coding like changing analog to digital and digital to analog and packetiztion, while network delay is due to transmission and queuing [34].

2.7.4 Jitter

In as much as IP networks cannot guarantee the delivery time of data packets (or their order), the data will arrive at very inconsistent rates. The variation in inter-packet arrival rate is jitter, which is introduced by variable transmission delays over the network.

Jitter represents the time variation between the arrivals of packets. Packets should arrive at regulate intervals ,but sometimes difference appear between the moment a packet is expected to arrive and the moment it actually arrives Problems caused by jitter may lead to gaps in the conversation or overlaps that may disturb the participants of the conversation [35].

Removing jitter to allow an equable stream requires collecting packets and storing them long enough to permit the slowest packets to arrive in time to be played in the correct sequence. The jitter buffer is used to remove the packet delay variation that each packet encounters transiting the network. Each jitter buffer adds to the overall delay. There are numerous ways of calculating this quantity.

$$J_K = J_{(K-1)} + (1/16 * (| \Delta_k| - J_{(K-1)}))$$

Where :

J_K: Filtered Jitter.

$J_{(K-1)}$: Previous Filtered Jitter.

Δ_k: Jitter, different between two consecution delay.

Finding the difference between the delay of the first and last packet of each time interval, this find the difference in delay between each adjacent packet of in an interval and take an average.

2.8 Summary

In this chapter, the research community's efforts on VoIP implementation over TCP,UDP and RTPs with an emphasis on QoS, was explored, And the technology component and challenge of VoIP. The literature shows the analysis of VoIP QoS for transmission protocol and signaling protocol.