Analysis of Security Threats of VoIP Systems

Dr. Ritesh Sadiwala [1],
[1,2]Department of Electronics & Communication,
RKDF University, Bhopal, India

Abstract: Voice over Internet Protocol (VoIP) is a new telephony technology. It allows people to make phone calls through data network. It is an IP based voice transmission technology, instead of the traditional analog telephone line, it allows people to make telephone calls through broadband internet connections. VoIP transmits packets via packet-switched network in which voice packets may take the most efficient path. On the other hand, the traditional public switched telephone network (PSTN) is a circuit- switched based network which requires a dedicated line for telecommunications activity. Lower cost and more flexibility are the main advantages of the VoIP which derived the attention of the enterprises for the next generation networks. This paper will describe this new technology i.e., VOIP along with its advantages and disadvantages and the purpose of this paper is to discuss the security threats of VOIP and to propose security methods that can be deployed to prevent from these threats. There are various protocols which are required in VOIP but we will consider two main protocols i.e. Session Initiation Protocol (SIP) and H.323 Protocol. Both are Signaling Protocols of VOIP and here we will discuss about SIP security threats and propose methods to prevent such threats.

Keywords: VOIP, SIP, H.323, Security Threats, VoIP Protocols

I Introduction

Voice over Internet Protocol (VoIP) [1] is one of the most important technologies in the world of communication. VoIP is simply a way to make phone calls through broadband internet Connection. Internet was initially considered to transmit data traffic and it is performing this task really well [2, 3]. To transmit voice conversations over a data network using IP, VoIP technology is used. Such data network may be the Internet or a corporate Intranet or managed networks which are specially used by long distance and local service traditional providers and ISPs (Internet Service Provider). Voice over IP refers to the diffusion of voice traffic over internet-based networks. Voice over Internet Protocol (VoIP) is a rapidly growing technology that enables transport of voice over data networks such as Ethernet local area networks (LANs) or internet [4, 5].

This growth is due to the integration of voice and data traffic over the existing networking infrastructure, low cost, and improved network management offered by the technology. VoIP can be used to call any PSTN telephone or mobile phone anywhere in the world. The goal of VoIP is to replace the operating circuit-switched, public switching telecommunication network to a packet-switched network. VoIP has been successful in deriving the attention of the telecommunication markets of all sizes and introducing the advanced features to the market, while on the other side the integration of the voice and data words caused evident security risks. Apart from advantages like lower cost and more flexibility of VOIP it is also very important to analyze all the security related issues at different levels of VOIP.

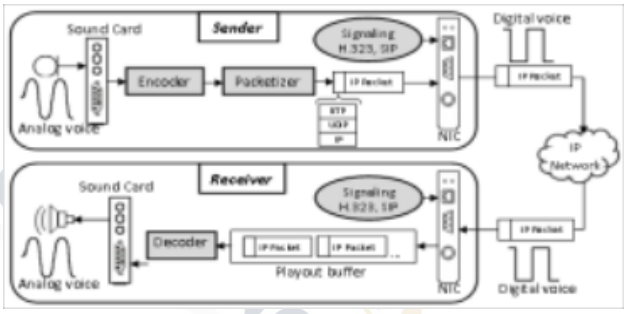

Figure 1: VoIP Components

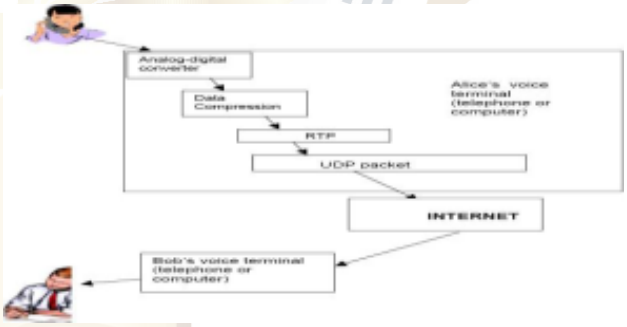

Figure 2: Voice Data flows between two end points

1.1 VoIP Components

VoIP consists of three essential components: CODEC (Coder/Decoder), packetizer and play out buffer [1, 2]. At the sender side, an adequate sample of analogue voice signals are converted to digital signals, compressed and then encoded into a predetermined format using voice codec such as as G.711, G.729, G.723.1a, etc. Next packetization process is performed which fragment encoded voice into equal

size of packets. Furthermore, in each packet, some protocol headers from different layers are attached to the encoded voice. Protocol headers added to voice packets are Real-time Transport Protocol (RTP), User Datagram Protocol (UDP), and Internet Protocol (IP) as well as data link layer header. In addition, to this RTP and Real-Time Control Protocol (RTCP) were designed at the application layer to support real-time applications.

Although TCP transport protocol is commonly used in the internet, UDP protocol is preferred in VoIP and other delay-sensitive real-time applications. TCP protocol is suitable for less delay sensitive data packets and not for delay- sensitive packets due to the acknowledgement (ACK) scheme that TCP applies. This scheme introduces delay as receiver has to notify the sender for each received packet by sending an ACK message. On the other hand, UDP does not apply this scheme and thus, it is more suitable for VoIP applications. The packets are then sent out over IP network to its destination where the reverse process of decoding and depacketizing of the received packets is carried out. During the transmission process, time variations of packets delivery (jitter) may occur. Hence, a playout buffer is used at the receiver end to smoothen the playout by mitigating the incurred jitter. Packets are queued at the playout buffer for a playout time before being played.

However, packets arriving later than the playout time are discarded. The principle components of a VoIP system, which covers the end-to-end transmission of voice, are illustrated in Figure 1. There are signaling protocols of VoIP namely Session Initiation Protocol (SIP) and H.323. These signaling protocols are required at the very beginning to establish VoIP calls and at the end to close the media streams between the clients [3]. H.323 was standardized by ITU-T specifically to smoothly work together with PSTN while SIP was standardized by Internet engineering task force (IETF) to support internet applications such as telephony [6].

In figure 2, VoIP protocol stack is illustrated. Furthermore, in IP networks, IP addresses can be changed from one session to another, especially in dial-up case. Therefore, there is a need for a common meeting point shared among users to enable them finding each other at the establishment stage of communication. This common meeting point is generically known as a call server. This paper comprises of six sections starting with Introduction, next section tells about VOIP, how it works, its advantages and disadvantages. In section III we give different VOIP standards and protocols and discusses two main protocols SIP and H.323 along with its architecture. Section IV tells us about various security threats of VOIP. Section V discusses about SIP security issues and proposes mechanisms to deal with such issues and finally concluding remarks is given in Section VI.

II Voice Over Internet Protocol

The history of VoIP began with conversations by a few computer users over the Internet. Initially, VoIP required a headset to be plugged into the computer, and the participants could only speak with others who had a similar set up. They had to phone each other ahead or sent a text message, in order to alert the user at the other end of the incoming call and the exact time [4]. In its early stages, the VoIP technology was not sufficiently mature. There was a big gap between the marketing structure and the technological reality. It results in an overall agreement that technical shortages stopped any major transition to VoIP. The most of the technical problems have been solved by VoIP technology [5].

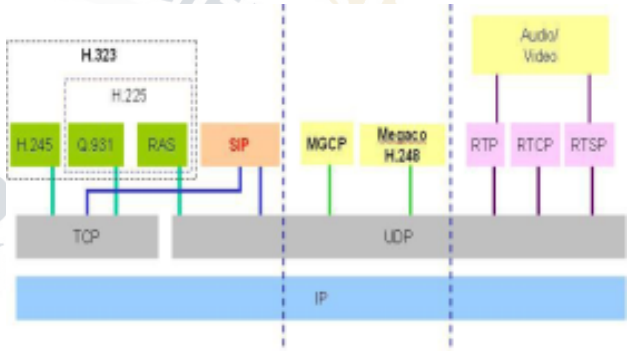

Figure 3: VOIP Signaling Protocols

2.1 What is VoIP?

VoIP (Voice Over Internet Protocol) is an IP network based voice transmission technology, instead of the traditional analog telephone line, it allows people to make telephone calls through broadband internet connections. In other words, just installing network telephone software on the PCs at each end, people can talk through to each other through the IP network. With the development of network technology, network IP telephony grew from PC-PC to IP-PSTN, PSTN-IP, PSTN- PSTN and IP-IP, etc. Here the common characteristic is using the IP network as the transmission medium and this is satisfied by using VOIP as VOIP requires less cost and same existing network to complete a VOIP call [7].

2.2 How does VoIP Work?

VoIP is a technology to transmit analog voice signal through the IP network. Simply speaking, it is accomplished by coding, compressing, packetization, etc, processes. After the voice data are transmitted to the destination through the network, in order to be received at the receiving end, it will be re-assembled by the opposite processes. Here is how the VoIP transmission is completed.

Step 1: Voice to Digital Data Transformation
Voice data is analog data, no matter in real time application or unreal time application, to transfer voice data in the IP packet, the first thing to do is to transform the voice data from analog signal into the digital bit stream, that is digitalizing an analog voice signal. In Digitalization the source and destination must use the same coding algorithm, so that the digitalized bit stream can be reverted to understandable analog voice data.
Step 2: Digital Data to IP Transformation
After digitalizing the voice data into bit stream, the next step is compressing and coding the voice packet into specific frames, this is done by using complex algorithms. Such as if a coder uses 15ms frame, then the first 60ms packet will be divided into 4 frames and coded in order. After coding, the 4 frames will be compressed into one IP packet and sent to the network processor. The network processor will add control header and payload in the voice packet, and send the voice packet to the destination through Internet.
Step 3: Transmission
In this session, the entire network will receive the IP packet from the sender and transmit it to the destination within a specific time, the time can be in different values within a specific range, it reflects the jitter in the network transmission process. Each node in the network checks the address information in the IP data, and uses this information to send the data to the next node. During the transmission, packets can be lost, damaged, or have errors. In the ordinary data transmission, the lost/damaged data can be retransmitted, but since VoIP is real time application, therefore a complicated error detection or correction method is needed.
Step 4: IP Packet to Digital Data Transformation
The destination VoIP equipment starts to process the IP packet after receiving it. A buffer is used to accommodate many voice IP packets. User can change the size of the buffer, small buffer generates small latency, but can not adjust big jitter. Address informaion and other control information will be removed, only the original data can be reserved, the reserved original data will be sent to the decoder, the decoder will decode and decompresse the voice data into new voice data.
Step 5: Digital Voice to Analog Voice Transformation
Here just the reverse process of step 1 is done and analog voice is received at the destination end. This process is depicted in Fig.2.

2.3 Advantages and Disadvanatages of VOIP

VoIP is a new service which comes in order to improve the legacy voice communication by supporting it with data communication as well. VoIP allows data, images and videos to be transmitted simultaneously Here, below benefits of VOIP are given as under.

- **Cost Savings:** Reduce the communication cost for the users which means that the communication can be done over private or internet data network line, instead of commercial telecommunications line.
- **Extendibility:** VoIP can be extended easily to any number of users and without any geographical boundary limitation.
- **Reuse-ability:** The available resources can be reused. Available network can be used for VoIP implementation. Data and voice service are combined easily with Rich media of services.
- **Easy Implementation:** Speech communications can be designed by computer networks companies within any organization. Collaboration and integration with other applications: This is because some protocols can collaborate with other applications easily, so it can take benefits from its properties.
- **Mobility of the Service:** The users can use the services from anywhere like voice mail, call features and so on.
- **User Control Interface:** Most of VoIP have user controls interface or graphical user interface (GUI) like in web, which make it easy to use.
- **Phone Portability:** The users do not need to change the communication details where ever they go on remove.

III VoIP Standards and Protocols

3.1 VoIP Standards

With the growth of VoIP, new requirements are brought forwarded, such as providing communication between a PC based soft phone and a phone on PSTN. Such requirements strengthen the need for a standard for IP telephony. Same as other technologies, there are various standards proposed to be accepted by the industry [8]. Two major standard bodies which govern the multimedia transmission over IP network are:

- International Telecommunications Union (ITU).
- Internet Engineering Task Force (IETF)[6].

3.2 VoIP Protocols

There are a number of protocols involved in VoIP service. In this section, we only focus on the most common protocols which are being used today, the protocols are RTP (Real Time Transport Protocol), H.323, SIP (Session Initiation Protocol) and Multimedia Gateway Control Protocol (MGCP). The relationship between VoIP protocols and other network protocols is displayed in Fig.3.

3.2.1 Real-time Transfer Protocol

Real-Time Transport Protocol (RTP) is an internet standard protocol, used to transfer real time data, such as audio and video. It can be used for IP telephony. RTP includes two parts: data and control. The control part is called Real Time Control Protocol (RTCP). VoIP uses protocols such as real-time protocol (RTP) and H.323 to deliver packets over the internet. Each VoIP packet has an internet protocol (IP)/UDP/RTP header with a total size header, 40 bytes. G.711 and G.729 are he two widely used voice encoding standards that are used with VoIP products [9].

- **Real Time Protocol (RTP):** It carries real time data. It provides support for real-time applications, includes timing reconstruction, loss detection, security and content identification. [7]
- **Real Time Control Protocol (RTCP):** It carries control information, the information is used to manage the QoS. It provides supports for applications such as real-time conference. The supports include source identification, multicast-to-unicast translator, and different media streams synchronization [7].

The RTP data structure is shown below: The real time data

IP Header	UDP Header	RTP Header	RTP Payload

Figure 4: RTP Data Structure

is in the RTP payload. RTP Header contains information of the payload, such as the soruce address, size, encoding type, etc. From Figure 4, we can see that RTP works on top of UDP. To transfer RTP packet on the network, we need to use User Datagram Protocol (UDP) to create a UDP header. To transfer UDP packet over IP network, we also need to create an IP header. To guarantee QoS, RTP use Synchronization Source (SSRC), Sequence Number and Timestamp to implement real time transmission. To protect conversations from being eavesdropped, secure RTP is designed which provides encryption, authentication and integrity check of the multimedia stream [10, 11].

3.2.2 How does RTP work?

Internet is a shared network, packets sent on the network may have delay, for multimedia application, transmission delay is important, thus RTP provides time stamping and sequence numbering to guarantee that the data are transferred within acceptable time limits.In RTP time stamping is important.

Here the sender sets time stamps on the packets according to the first octet on the packet. After receiving the data packet, the receiver reassembles the data according to the time stamp in the correct order. Time stamp is also used for synchronization, for ex., to synchronize audio and video data in MPEG format. Apart from this sequence number is also used to detect data loss in packets of video data. In RTP payload is used to indicate what mechanism

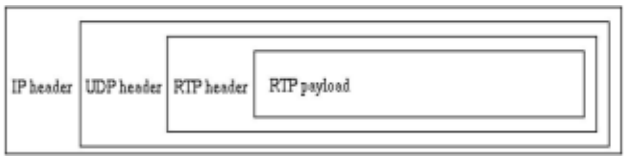

Figure 5: RTP Data in IP Packet

is used to coding/compressing the data, the receiving side uses this identifier to choose correct mechanism for decoding/decompressing the data. At one time, RTP can send only one type of payload. Another function of payload is source identification, it enables the receiving side to know where the data come from. The following figure depicts the RTP data in IP packet.

To set up RTP session, the application defines a pair of destinations: network address and a pair of ports. In multimedia session, each medium use a separate session, thus the RTCP can report the transmission quality separately. Such as transmission of audio and video, audio and video data use different RTP session, thus the receiver can choose whether or not to receive one medium.

3.2.3 Real Time Control Protocol (RTCP)

RTCP is a control protocol, it works together with RTP. RTCP is sent periodically by participant to get feedback of transmission quality. There are five types of RTCP packets:

1. **RR:** Receive Report. This is created by the receiver, it is used to report the transmission quality to the sender.
2. **SR:** Sender Report. This is created by the sender. It is used to synchronize packets, and calculate packet counters, and the number of bytes sent.
3. **SDES:** Source Description Items. It contains information to describe the source.
4. **BYE:** used to indicate that participation is finished.
5. **APP:** application specified functions.

By using the control information listed above, RTCP can provide services like QoS monitor and congestion control which is the most important function in RTCP. RTCP sends a feedback of the transmission quality to the sender, the sender then uses this information to adjust the transmission speed. Network administrator can also evaluate network performance with this information.

3.2.4 RTP Features

RTP has many interesting features. They are

- To provide end-to-end delivery service for real time data, such as audio and video.
- RTP uses time stamps and sequence numbers to implement reliable delivery, flow control and congestion control.
- RTP is only a protocol framework, it is open to new multimedia software.
- RTP and RTCP provide functionalities to deliver real time data. RTP and RTCP aren't responsible for synchronization, or something like it which is the higher level task.

3.3 H.323 Protocol

Before multimedia data can flow from a device to another device, various protocols are used to define how to transfer the stream. The protocol aimed at this functionality is called call- signaling protocol. The two major protocol standards for VoIP signaling are: H.323 protocol (ITU) and Session Initiation Protocol (SIP) (IETF). However, each standard uses different methods for call signaling and call control. More importantly, they are not interoperable [8]. From the figure 3, people can easily see that VoIP signaling protocols H.323 and SIP work in the Session layer, the responsibility of Session layer protocols is to establish or cut off communications between processes.H.323 is a standard, it specifies the components, protocols and procedures to provide multimedia communication services over packet based network.H.323 is based on RTP, RTCP and other protocols.

H.323 is a part of family of ITU-T recommendations called H.32x which provides multimedia communication services [8].The Network Architecture of H.323 protocol is given in Fig.4 and H.323 protocol Stack Architecture is represented in Fig.5. H.323 is a protocol stack, the protocols and standards work together to enable the conference on packet-based network. Each protocol in H.323 performs a specific function, such as H.261, H.263 and H.264 are video codec's, they are software algorithm used to compress/encode and decompress/decode video signals. H.323 architecture is given in Fig.6. To implement communication over network, there are four important components in H.323.

The four components are: Terminals, Gateways, Gatekeepers, and Multipoint Control Units. (i)Terminals:-Used for real time two-way multimedia communications, an H.323 terminal plays a key role in IP telephony services. It can be a PC or a standalone device, such as an IP telephone set. H.323 terminals may also be used in multipoint conferences. (*ii*) **Gateways:** An H.323 gateway provides connectivity between H.323 network and non-H.323 network, such as an

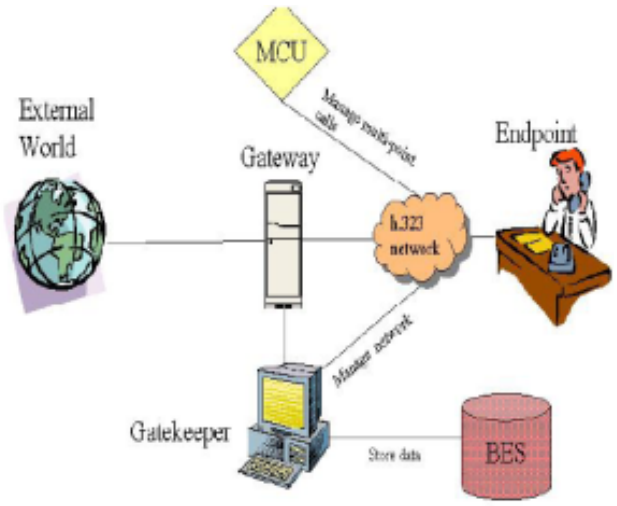

Figure 6: H.323 Network Architecture

IP network and a circuit-switched network (PSTN). To connect different networks, it is necessary to translate protocols and transfer information between different networks, such as translation between different formats (H.225 to H.221), between communication procedures (H.245 to H.242), the gateway also translates between audio and video codes and establish calls or cut off calls. An example of H.323/PSTN gateway is shown in Fig 7.

(*iii*) **Gatekeepers:** A gatekeeper can be thought of as the most important component in H.323 network. Gatekeeper provides important services, such as addressing, authorization and authentication of terminals and gateways, bandwidth management, accounting, billing and charging, gatekeepers may also provide call routing services. Gatekeeper performs two important functions: translation of address from alias to IP address and bandwidth management. For example, if the network administrator defines the upper limit of how many terminals can join a conference, then when the up limitation is reached, it will refuse more terminals to join the conference.

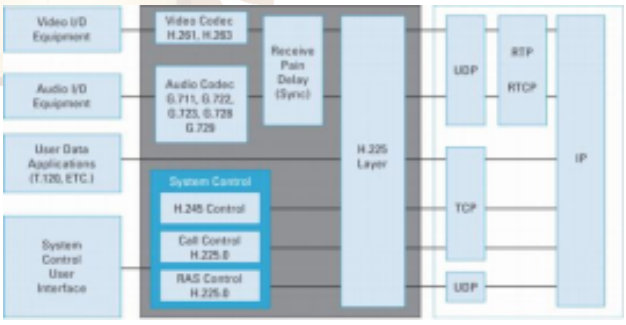

Figure 7: Protocol Stack Architecture

The benefit of this function is to limit the bandwidth which is allocated to the VoIP, thus the left bandwidth can be used to transfer e-mail, fax, file, etc. The required func-

tions of gatekeeper are address Translation by referencing a table, translate address between alias and transport address (IP address), this table is updated with the Registration Message. Second one is Bandwidth Control which is based on bandwidth management. Also Admission Control to authorize the access to the LAN and Zone Management where gatekeeper provides the above functions to the terminals, gateways, or MCUs which has registered in the zone are its additional functions [10].

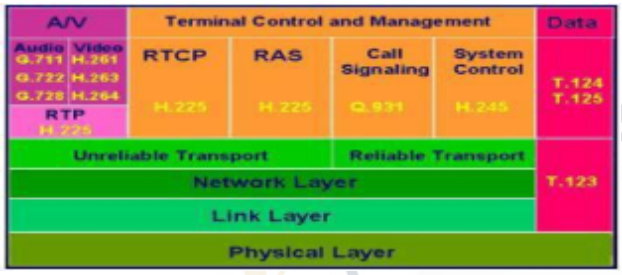

Figure 8: H.323 Architecture

(*iv*) **Multipoint Control Units:** The H.323 MCU is used to setup conferences for three or more H.323 terminals. All terminals participating in the conference need to setup a connection with the MCU. The MCU can be a stand-alone device or integrate into another H.323 component, such as gatekeeper. MCU includes of Multipoint Controller (MC) and Multipoint Processor (MP). The MC uses H.245 to negotiate between all terminals to determine the audio and video process capability. MC also controls the conference resource to determine which stream (audio or video) should be multicast. MC doesn't process stream directly, MP process stream, it mixes switches and process audio, video and data bits. MC and MP can exist in a separate device or integrated into H.323 components.Fig 8 which shows how call sets up in H.323 protocol.

3.4 Session Initiation Protocol (SIP)

Another signaling protocol is Session Initiation Protocol (SIP) which is used to create, manage and terminate sessions in an IP based network [9]. SIP has been used in VoIP in the recent past, it is a standard put forwarded by Internet Engineering Task Force (IETF).

SIP is still growing and being modified to include other relevant features, but the job of SIP is limited to only set up sessions. Unlike H.323, SIP is not a complete protocol for multimedia communication. Instead, SIP works together with other protocols to provide functionalities similar to H.323. The relationship between SIP and other protocols is shown in Fig.9. SIP is a session layer protocol, it has two basic functions: signaling and session control. Signaling is used to translate signals between different networks and Session control is used to control the attributes of the end to end call.

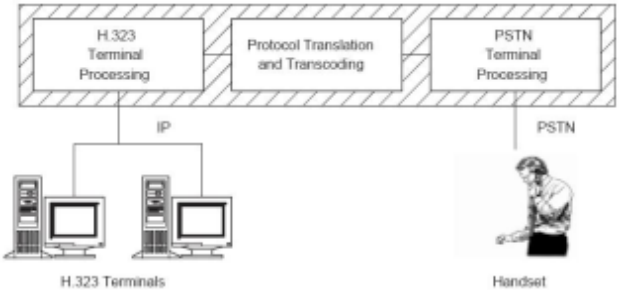

Figure 9: H.323/PSTN Gateway

SIP has the ability to provide address resolution, name mapping and call redirection, it can find the location of the end node, determine the capability of the end node, conferences can only be established between end nodes which have enough capabilities, provide different ring back signals such as if the end node is busy, SIP provides busy tone to the caller, establishes session between two nodes if the call can be completed, etc. Session Initiation Protocol (SIP) is the Internet Engineering Task Force's (IETF) standard for multimedia conferencing over IP [11].

SIP is an application layer control protocol, it can be used to setup, maintain and cut off calls between two or more terminals. SIP is designed for providing signaling and session management service over packet based network. Signaling service enables calls to be transmitted across networks, session management is used to control the attribute of the end-to-end call. The services provided by the SIP protocol are to determine the target location, address resolution, name mapping and call rerouting. SIP uses Session Description Protocol (SDP) to determine the "lowest requirement". SIP establishes a session between the two terminals and handles the transfer and termination of sessions.

3.4.1 SIP Overview

SIP consists up of two types of entities: User agent (UA) and Networkservers.

User Agents (UA): SIP is a peer to peer protocol, the two peers in a session are User Agents. The user agent consists of two functionalities: User Agent Client (UAC) and User Agent Server (UAS). The UAC is used to initiate calls, the UAS responds to call requests, by exchanging request and response, User Agent can initiate and cut off sessions between each other. User Agent Client is a client application which is used to initialize a SIP request while User Agent Server is a server application, when User Agent Server gets a request, it contacts the user and returns a response to the User Agent in the name of the user. The UAC and the UAS can be located on the same device such as an IP telephone. SIP calls can be made to another UA directly, or through either the redirect server or the SIP proxy server [8].

Network Server: There are four types of SIP network servers, they are registration server, location server, proxy

server and redirect server.
(*a*) **Redirect Server:** Redirect server accepts SIP request from a client, maps the SIP address of the called party and returns the address to the client. Redirect Server doesn't forward request to other servers [12].
(*b*) **Registrar Server:** a registrar server is a server which accept register request from a client, and update the location database, the location database is used to store contact information [12].
(*c*) **Proxy Server:** It handles SIP requests for the source UA. A proxy server can perform as a server or a client to make a request in the name of clients. Requests are serviced either locally or passed on to another server [12]. (*d*) **Location Server:** is used to store terminals location, and provide a terminals location to the proxy server or redirect server. The SIP network architecture is shown in Fig 10.

SIP has two types of messages request and response. Request sent from client to server and Response sent from server to client. There are two types of responses and six types of classes. Response Types are Provisional and Final. Provisional response come under 1xx class and is used by the server to indicate progress, but this response doesn't terminate the SIP transaction and Final response can be in any of the 2xx, 3xx, 4xx, 5xx, 6xx class and is used to terminate the SIP transaction. Along with this there are six classes where the message can be categorized.

They are 1xx which indicates that the request is received and is continuing to process the request.2xx means the action is successfully received, understood and accepted.3xx means redirection i.e., more actions are needed to complete the request.4xx means that client has error. The request received has error.5xx means that server has error. The server cannot fulfill a valid request and 6xx means global failure i.e., the request cannot be fulfilled at any server [12].

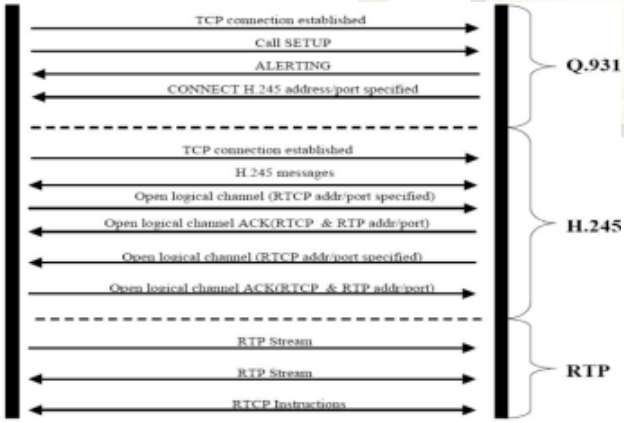

Figure 10: H.323 Call Setup Procedure

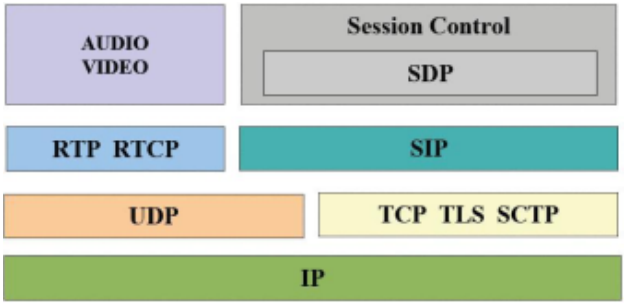

Figure 11: SIP and Other Protocols

3.4.2 How does SIP Work

SIP call setup procedure is given in Fig.11.Suppose User A wants to call User B, User A agent sends an invite message to its proxy server, the proxy server finds that there is another proxy server which provide service for User B, so it forwards the invite message to the next proxy server and sends an message back to first user agent indicate that the invite message has been received, the next server also sends back a message to the first proxy server to indicate the invite message is received and forward this invite message to User B agent, when the telephone of User B rings, user agent of User B sends back a provisional message (ringing), this provisional message is forwarded to first user agent by the proxy servers, when User B decides to answer the call, User B agent sends an OK response to first user agent, together with other information, such as codec, etc, First user agent sends back an ACK as confirmation, now the voice data can be transferred by RTP, when one of the parties want to hand up, the user agent of this party sends an BYE message to the other side, and the other side sends back an OK message, then the call is disconnected. In the example given below in Fig.11 Alice is taken as first user and Bob is taken as second user.

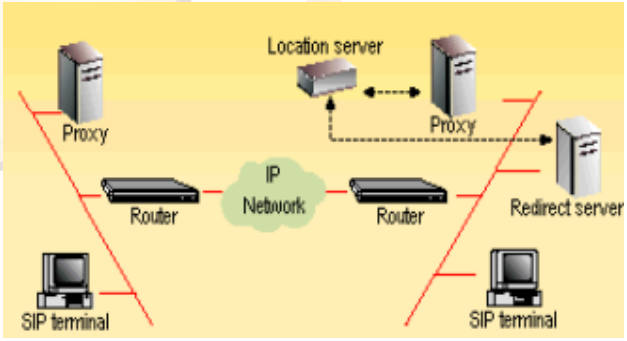

Figure 12: SIP Network Architecture

3.4.3 Media Gateway Control Protocol (MGCP)

Media Gateway Control Protocol (MGCP) is another protocol which is used to control the media gateways. MGCP

is created from other two protocols, Internet Protocol Device Control (IPDC) and Simple Gateway Control Protocol (SGCP). MGCP which is the extended H.323 gatekeeper model. MGCP handles the traffic between media gateway and the controller. It is the controller which performs conversion from packet switched network to circuit switched network.

This is a master-slave protocol, the master has absolute control and the slaves just follow the commands. The master is the media gateway controller or soft switch, the slave is the IP phone or VoIP gateway. This protocol is a contrast to peer-to-peer protocol which means that the client cannot establish connection with another client. MGCP is designed to reduce the workload of the IP telephones so that the IP telephones can be un-expensive and less complex.

Figure 13: SIP Call Setup Procedure

3.4.4 MGCP Overview

In MGCP, there are eight ways for exchanging data between the media gateway controller (call agent) and the media gateway.

(A) Notification Request from call agent to gateway: This is used to request the media gateway to notice the special telephone events, such as off-hook, on-hook, fax tones, modem tones, flash hook, continue tone, etc. The nice aspect of this request is that it integrates the events with actions. Such as when a call agent requests the gateway to notice digits, it can also request the gateway to store digits.

(B) Notification from gateway to call agent: This is used by the gateway to send back the events which are requested by the call agent, the media gateway can send one or several events in one notification command. But the events sent back by the media gateway are in the order the call agent sent to it.

(C) Create Connection: This is sent from call agent to gateway to create a connection between two end points. In addition to the necessary parameters for gateway to establish the connection, there are other parameters, such as QoS, security, bandwidth, type of service, etc. The call agent can request the gateway to execute several actions at the same time, such as, ask the gateway to prepare connection ask the gateway to start ring, ask the gateway to notify call agent when the phone goes off-hook. All these events can be done with one Create Connection command.

(D) Modify Connection from call agent to gateway: This is done by the call agent to modify an established connection. Call agent can use this command to change parameters, like activation, deactivation, change codec, packetization period, etc.

(E) Delete Connection from call agent to gateway: This is used by call agent to delete a connection.

(F) Audit Endpoint from call agent to gateway is used to check if an end point is up.

(G) Audit Connection from call agent to gateway: Call agent uses this command to get all the parameters of the connection.

(H) Restart in Progress from gateway to call agent: Media gateway uses this command to report to the call agent that more than one end points have problem.[13].

The MGCP protocol architecture is given in Fig 12. The key component in MGCP is Media Gateway (MG), it is responsible for switch information between a packet based network to a circuit based network, it also handles RTP media steams across the IP network. There are several types of gateway in VoIP, they are trunking gateway, residential gateway, access gateway, network access server, etc.

(*a*) **Trunking Gateway:** It is the interface between the telephone network and the VoIP network. (*b*) **Residential gateway:** It provides an analog interface to VoIP network. (*c*) **Access Gateway:** It provides analog or digital PBX interface for VoIP network. (*d*) **Network Access Server:** It can be linked to a modem to a telephone circuit and provide internet access at the same time. Media Gateway Controller

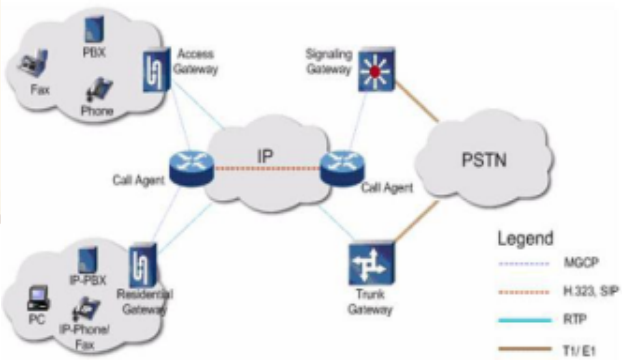

Figure 14: MGCP Architecture

deal with media gateway registration, management and control, it performs signaling transformation between different networks, such as from PSTN network to IP network. Thus these are some of the protocols which are used in VOIP, next we discuss about the security threats in VoIP.

IV Security Threats of VoIP

VoIP maximizes the usability of network, reduces cost and time, and provides new service opportunities. VoIP extends services to remote locations with lower cost. VoIP brings new multimedia service opportunities, such as PC based call, web- based multimedia conference [14].While VoIP brings many benefits to us, it also put forward security problems in front of us. The following section gives description of security issues in VoIP. There are many different methods which can be used to attack VoIP. Some attacks try to steal information while others attempt to shut down the network. The attacks to VoIP aim at confidentiality, integrity and availability.

Confidentiality: Confidentiality means the privacy of information, sensitive information, such as username, password, financial information, security information, etc, should be protected. Usually, an attack to VoIP has the aim of destroying service, stealing service or destroying privacy. In the traditional telephony system, there is physical protection for the information confidentiality, since it is difficult to reach the physical equipment, such as physical telephone line, telephone switch. But in VOIP voice data are transferred over Internet that means everybody with a computer and a modem has the ability to reach the voice data.

Thus protection of confidentiality in VoIP is more difficult. Attacker can make use of user authentication and authorization tools to intrude system, share privilege with legal user, steal sensitive information, or gain unauthorized access to network resources. Integrity: Integrity of information means the information cannot be modified by unauthorized user. For example, the bank account numbers can only be changed by the user himself, or other security administrator. In VoIP scenario, damage data integrity on the server may result in the attack like denial of service.

Availability: Availability means the service, information or resource are always available when it is needed by authorized user. Attacks to availability may result in bad service quality or denial of service. In addition to this some other threats are given below

4.1 Malformed Message Threat

Malformed Message Threat is one of the most representative cases using the vulnerabilities of text-based protocol. The attackers are able to cause malfunctions of proxy server by manipulating SIP headers. For instance, overflow-space, overflow-null, specific header deletion and using non-ASCII code are involved in these malformed message threats. By intercepting the messages transferred between server and client, the attacker can get the public key, and then get messages which are sent by the client, decrypt the message with the key. After decryption of the message, the attacker can modify the message and forward the message to the server, or without modifying the message. For the server and the client, they don't know there is an intruder between them. This attack is known as man-in-the-middle. Similarly, the attacker can open entries on the network by accident or purposily, this enables back door attack. Another by pretending to be a service provider, the attacker can track the user to connect to it and get sensitive information of the user. This attack is known as masquerading.

4.2 SIP Flooding Threat

IP phones generate requests or responses to send to a specific UA, called by victim. As a result, a single UA is overwhelmed by receiving excessive SIP messages within a short duration of time, so that the UA cannot provide normal services. INVITE flooding is one of the most typical threat. Basically, flooding attack is also the issue of IP layer. In case of INVITE flooding, however, it could be more annoying threat for the VoIP user because the one should see many call requests at the same time and hear ringing of calls.

4.3 Spoofing Threat

Spoofing [15] can be done when an attacker searches to be someone else in order gain access to restricted resources or steal information. This type of threat can take a variety of different forms, for instance, an attacker can change the protocols which are used as the Internet Protocol (IP). Also, an attacker may send fraudulent emails and set up fake websites in order to capture user's login names, passwords and account information. A phishing attack is any fake email or websites. Another type of spoofing involves setting up a fake wireless access point and tricking victims into connecting to them through the unauthorized connection. There are two kinds of spoofing threats which are possible, first one is IP spoofing threat and another is URI spoofing threat.

IP spoofing threat is a way for IP source addresses in order to feign a trusted user. In URI spoofing threat the attacker who hijacked SIP messages between two UAs forges their URI field, so the attacker can hide himself from trace backs. If spoofed BYE requests (BYE DoS attack) are sent to a victim, then the call would be terminated by this attacker. Spam over internet telephony (SPIT) is unwanted, automatically dialed, prerecorded phone calls using Voice over Internet Protocol (VoIP). It is similar to E-mail spam. By IP spoofing or session hijacking, an attacker can access network in the name of a legal user. By using sniffer to get data from network, attacker can obtain information like username, password, and with these information to perform further attack network vulnerable to eavesdropping.

4.4 Denial-of-service (DoS) or VoIP Service Disruption

Many systems does not have authentication, so an attacker can log onto a computer which is on the VoIP network, and then the attacker send ARP flood to corrupt ARP caches.

ARP flood attack to the switch makes them flooding bad request to key component (such as server, gateway) in VoIP, the component may be crashed, and cannot provide service to legal user, this attack is knows as Denial of Service (DoS).Denial-of-service (DoS) threats can affect any IP-based network service, and are the most challenging threat in VoIP applications. One type of attack in which packets can simply be flooded into or at the target network from multiple external sources is called a distributed denial-of service (DDoS) attack.

4.5 Call Hijacking and Interception

Call interception and eavesdropping are other major concern on VoIP networks which cause theft of information and services on VoIP networks. The existence of this threat in VoIP applications is because of the deficiency or absence of authentication measures. This threat demonstrates the need for security services that enable entities to authenticate the originators of requests and to verify that the contents of the message and control streams have not been altered in transit.

4.6 H.323-Specific Attacks

H.323 is signaling protocol in VoIP communications which is encoded according to ASN.1 PER encoding rules. The implementation of H.323 massage parser, rather than the encoding rules themselves cause vulnerabilities in H.323 suits.

4.7 Signaling Initiation Protocol (SIP)-Specific Attacks

SIP is an unstructured text-based protocol which suffers vulnerabilities according to its encoding format, because it is not possible to check all permutations of SIP messages throughout development for security vulnerabilities. Since SIP protocol links other protocols and services together, it may cause other typical vulnerabilities in services such as SSL, hypertext transfer protocol (HTTP), simple mail transfer protocol (SMTP) to occur in VoIP environment. In next section we will concentrate on the SIP security threats and give measures to prevent it.

V SIP Threats and Its Measures

The SIP protocol resides in an application layer. It is a text based client-server protocol within the UDP or TCP Transport that exchanges plain-text messages.SIP does not use any encryption mechanism, so it is very easy to access the sensitive information contained in the SIP protocol like information of sensitive IP address, address of the contact, information of Port address, SIP compliance capabilities, Username, Media steam attribute, Type of MIME Content There is a list of several factors which makes the SIP insecure.

(*i*) **Maturity:** SIP is a relatively new standard.
(*ii*) **Complexity** SIP is not a complex protocol, but all the necessary extensions make the SIP a complicated protocol.
(*iii*) **Encoding:** SIP is text message protocol, and is easily visible to any sniffer.
(*iv*) **Extensibility:** SIP supports extensions, these features are new but often weak from a security perspective.

SIP sessions are used by network elements for modifying, terminating a session, and resource discovering. Therefore SIP security such as authentication, Confidentiality and authorization is an essential element. Different attacks like Denial of the services (DoS), Man in the middle, and ping attacks can cause security threats. To offer further integrity SIP used a built mechanism for protection against different kind of attacks, and it relies on different protocol like IPSec, Transport Layer Security (TLS) and Secure Mime (S/MIME) [17].

5.1 SIP Threats

A SIP based system is vulnerable to common IP and VoIP attacks. There are several security issues concern to SIP based VoIP system. The lists of attacks that are unique to SIP are as follows:

Registration Hijacking: This threat occurs when an intruder in the network impersonates a valid UA into a registrar and replaces his address as a legitimate user. Then all of the incoming calls send to the attacker legitimate address. The Registration process normally uses UDP protocol that provides a weak security mechanism. Most of the registrar just requires a simple username and password. It can easily be defeated by generating dictionary-style attacks. In dictionary- style attacks, an attacker needs just to know the username and then he steps through a list of built-base passwords like enterprise name, office branch name or organization name. Some organizations use a shared mechanically generated weak password such as an extension with additional word, so this way an attacker may learn one of enterprise's passwords and then he may be able to learn all of its passwords [18]. The Registration Hijacking Threat of SIP is shown in Fig.13.

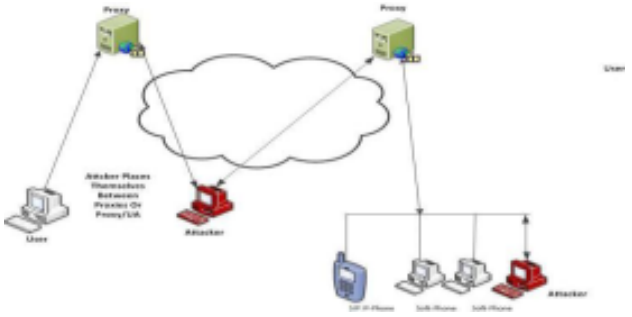

Figure 15: Registration Hijacking

Proxy Impersonation: During communication with a rogue proxy these threat occur when an attacker/intruder tricks one of enterprise's SIP servers (UA), if this attack occurs successfully then the attacker can access all SIP messages and control on all SIP calls[18]. Proxy Impersonation in SIP is shown in Fig.14.

Figure 16: Proxy Impersonation of SIP

Message Tempering: In this threat an attacker/intruder intercepts the packets, and modifies these packets when the SIP component exchanged their messages. It can also occur through proxy impersonation, registration hijacking, or an attack on any trusty SIP component. As we know, SIP messages do not have any built-in security so they will not provide integrity. An attacker can use the same type of attacks against any insecure system for registration hijacking and proxy impersonation [18].Message Tempering is shown in Fig.15.

Figure 17: Message Tempering in SIP

Session Tear Down: In this threat an attacker/intruder observes the signals of a call. After receiving the signals he sends spoofed SIP "BYE" messages to the participating UAs, which tear down the session. In most of the SIP sessions UA does not require a strong authentication, they open the gate for an attacker to send a properly crafted "BYE" message. An attacker does not need to observe the call signaling because UAs do not check the available packet value and if an attacker knows the active address of UA (like media-gateway, Interactive Voice Response (IVR), Automated Attendant (AA), or Virtual Machine systems (VM), or trading floor phones, etc...), then he can send a "BYE" message to tear down the call. The UDP ports become open for a legitimate call, so flooding the firewall with a "BYE" messages by an attacker, may cause to tear down the session [18]. This threat is shown in Fig.16.

Figure 18: Session Tear Down in SIP

Denial of Services (DoS): This threat can occur through any of means whether through additional DoS specific attacks because strong authentication is rarely used in SIP processing components or SIP messages. These are malformed packets, which are manipulating the SIP states, and cause flooding. SIP implementations are highly vulnerable by these kinds of attacks. It can cause a high level of damage if it targets the network voice resources like media-gateway, Interactive Voice Response (IVR), Automated Attendant (AA), or Virtual Machine systems (VM). It generates a large number of toll emergency calls such as (911), and information call such as (411)[18].Denial of Services(DOS) is shown in Fig.17.

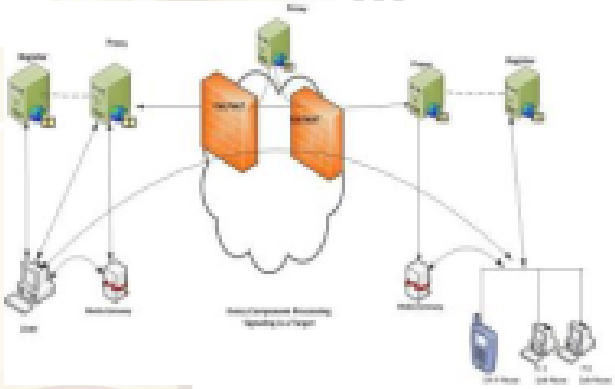

Figure 19: Denial of Services (DoS) Attack in SIP

There are basically six aspects to securing a SIP network: Authentication, Authorization, Confidentiality, Integrity, Privacy and Non-repudiation. **Privacy:** Privacy is defined by the VOIP alliance as "The privilege to have their communication systems and content free from unauthorized access, interruption, delay or modification" [19]. In other words, users should know what kind of information is delivering, and it should be encrypted and finally it should successfully reach the intended party. Privacy issues between the parties present many threats to the applications such as message tampering, and message eavesdropping [20].

Privacy requires an implementation of a set of secure interfaces, which provide authentication, authorization and integrity.

Confidentiality: Confidentiality can be achieved by using different encryptions techniques, which provide user authentication. For ex: a hash record key with a shared secret is used between the parties to prevent malicious users from call monitoring. Such measures should be taken to get confidentiality [20]

Integrity: To protect the source of data we use Integrity that provides user authentication. It is used for origin integrity, and without integrity control, any non-trusted system has the ability to modify the different contents without any notice.

Authorization: Authorization requires querying a database containing the basic account information for a subscriber. This account information provides the public as well as private identities for the subscription, and all the services the subscriber is authorized to access.

Authentication: Authentication requires the use of passwords and the exchange of credentials. Whenever a subscriber registers his or her location with the network, the registrar should always challenge the initial registration.

Non-repudiation: It prevents subscribers from accessing services and later denying that they used those services. If the operator implements the right tools and audit systems, you should have total visibility to every network transaction that takes place. It also includes any downloads that the subscriber may have made. Above aspects of SIP security was common and was used generally in the past. However, some more security measures at the protocol level are proposed. They are using HTTP digest authentication, using S/MIME for integrity protection, using RTP to encrypt data for confidentiality and using IPSec to provide signaling protection.

HTTP Digest Authentication: SIP uses HTTP Digest Authentication method to authenticate data, such as password. HTTP Digest authentication offers one-way message authentication and replay protection, but it doesn't protect message integrity and confidentiality. By transmitting an MD5 or SHA-1 digest of the secret password and a random challenge string, HTTP Digest can protect password. Although HTTP digest authentication has the advantage that the identity of the user is encrypted, and transmitted in cipher text, but if the password is short or weak, by intercepting the hash value, the password can be decrypted easily. Another problem is that there is no encryption mechanism to ensure the confidentiality and the integrity of the SIP message. Some SIP messages (such as ACK) doesn't require response. Authentication for these messages is based on the previous request that means an attacker can send a modified message to perform a DoS attack.

S/MIME: MIME bodies are inserted into SIP messages. MIME defines mechanisms for integrity protection and encryption of the MIME contents. SIP can use S/MIME to enable mechanisms like public key distribution, authentication and integrity protection, confidentiality of SIP signaling data. S/MIME relies heavily on the certification of the end user. Moreover self certification is vulnerable to man-in-the-middle attack, so either the certificates from known public certification authorities (CAs) or private CAs should be used, so the S/MIME mechanism is seriously limited.

IPSec: SIP uses IPSec to protect message exchanged between user agents. IPSec assumes a trusted relationship between peers, and it can only be used in hop-to-hop mode.

Firewall/Network Address Translation (NAT): Firewalls are usually used to protect trusted network from un-trusted network. Firewalls usually work on IP and TCP/UDP layer, it determines what types of traffic is allowed and which system are allowed to communicate. Firewall doesn't monitor the application layer. Since SIP needs to open ports dynamically, this enhances the complexity of firewall, as the firewall must open and close ports dynamically. Thus, NAT is used to preserve IP address. Also for a secure session in VOIP we should take following measures

- Use and maintain anti-virus and anti-spyware programs.
- Do not open unknown attachments of mails which have unknown or fake IDs.
- Verify the authenticity and security of downloaded files and new software.
- Configure your web browser(s) properly by enabling/disabling the necessary cookies.
- Active firewall session in your network and always place your back-up securely.
- Create strong passwords and change them regularly and do not disclose such information publicly.

In addition to this some mechanisms which can be used to avoid such threats are:

- To prevent message alteration established secured communication channel between communicating parties. To prevent media alteration and degradation use SRTP protocol.
- Another technique for preventing message tampering in SIP is to send SIP message digitally signed to receiver. As a result, any modification in a SIP message can be detected and discarded by the SIP server. Generally, digital signatures can protect SIP messages from any sort of tampering attack. For example send e-mails to anyone by using your digital signatures.
- Use secured devices for communication and switching of voice as well as data.
- Use Strong authentication and password at device level.

- Change defaults passwords and enable SIP authentication. Use the devices which support SRTP cipher technique.
- Use VLAN with 802.1x in internet to split data and voice traffic.
- Disable Telnet in the phone configuration, allow only to administrators.

To avoid message tampering and voice pharming attack use encrypted transmitted data using encryption mechanisms like IPsec, TLS and S/MIME. IPsec provide encryption of SIP message at network layer. IPsec supports both end to end and hops to hops encryption. IPSec support Internet Key Exchange (IKE) protocol for key management.

VI Conclusions

This paper is based on the security threats in VOIP. In the early days of VoIP, there was no big concern about security related issues. People were mostly concerned with its cost, functionality and reliability. Now that VoIP is gaining wide acceptance and becoming one of the mainstream communication technologies, security has become a major issue. In this paper we have described what is VOIP, how it works, its advantages, its standards and different protocols used in it in detail. Then we concentrate on two main protocols SIP and H.323 and discussed the various security threats that SIP protocol is concerned with and propose various mechanisms to prevent VOIP threats. We have also given measures that should be used and implemented on regular basis in VOIP networks in order to get prevention from such threats. Security measures in VOIP are in its beginning stage and a lot of research has to do in this area. As the need and speed of the internet and data traffic will increase in future more new threats will come into picture as now the attackers/hackers are not only threatening on the network level but also at the protocol level. So a level based approach has to be used both at the network level and at the protocol level to monitor them and to take immediate preventive measures against them.

REFERENCES

[1] S. Ganguly and S. Bhatnagar, *Basics of VoIP*. Wiley, 2008. [Online]. Available: https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=8045312

[2] T. Chakraborty, I. S. Misra, and R. Prasad, *Overview of VoIP Technology*. Cham: Springer International Publishing, 2019, pp. 1–24. [Online]. Available: https://doi.org/10.1007/978-3-319-95594-0_1

[3] W. A. Flanagan, *Network Management for VoIP and UC*. Wiley, 2011. [Online]. Available: https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=8044504

[4] W. Flanagan, *VoIP Signaling and Call Processing*. Wiley, 2011. [Online]. Available: https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=8043824

[5] William A. Flanagan, *VoIP and Unified Communications Define the Future*. Wiley, 2011. [Online]. Available: https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=8043231

[6] C.-Y. Wu, K.-P. Wu, J. Shih, and H.-M. Lee, "Voips: Voip secure encryption voip solution," in *Security-Enriched Urban Computing and Smart Grid*, R.-S. Chang, T.-h. Kim, and S.-L. Peng, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 84–93.

[7] T. Daengsi, N. Khitmoh, and P. Wuttidittachotti, "Voip quality measurement: subjective voip quality estimation model for g.711 and g.729 based on native thai users," *Multimedia Systems*, vol. 22, no. 5, pp. 575–586, Oct 2016. [Online]. Available: https://doi.org/10.1007/s00530-015-0468-3

[8] M. Hruby, M. Olsovsky, and M. Kotocova, *Solving VoIP QoS and Scalability Issues in Backbone Networks*. Dordrecht: Springer Netherlands, 2013, pp. 537–549. [Online]. Available: https://doi.org/10.1007/978-94-007-6190-2_41

[9] P. Włodarski, "Quality of service for aggregated voip streams," in *Software Engineering and Algorithms in Intelligent Systems*, R. Silhavy, Ed. Cham: Springer International Publishing, 2019, pp. 431–437.

[10] E. Imen, A. A. Imen, and M. Debyeche, "Framework for voip speech database generation and a comparaison of different features extraction methodes for speaker identification on voip," in *2015 3rd International Conference on Control, Engineering Information Technology (CEIT)*, May 2015, pp. 1–5.

[11] T. Sinam, I. T. Singh, P. Lamabam, N. N. Devi, and S. Nandi, "A technique for classification of voip flows in udp media streams using voip signalling traffic," in *2014 IEEE International Advance Computing Conference (IACC)*, Feb 2014, pp. 354–359.