Analytical Study of the Aspects of VoIP Communications

G. E. Lakshantha

Faculty of Information Technology
University of Moratuwa

eranda1985@yahoo.com

Abstract: *The cost savings and new features associated with Voice over IP (VoIP) are driving its adoption by people. Saving money is one good reason to use VoIP, but there are others. VoIP offers the same features that can be obtained with traditional phone service plus many others that can both increase productivity and create a more flexible work environment for the workforce. VoIP allows for real-time collaboration so that people can use it to set up instant conference calls or check the availability of their coworkers. It can also help organizations to improve customer service by allowing to stay connected on the go. VoIP is an emerging technology where there are still some areas under research. In this paper, I primarily focus on how VoIP works, architecture and VoIP infrastructure, VoIP security, performance & quality issues, and applications of VoIP.*

1. Introduction

In the 1990s, a number of individuals in research environments, both in educational and corporate institutions, took a serious interest in carrying voice and video over IP networks[1]. This technology is commonly referred to today as VoIP and is, in simple terms, the process of breaking up audio into small chunks, transmitting those chunks over an IP network, and reassembling those chunks at the far end so that two people can communicate using audio.

VoIP is a method for taking analog audio signals and turning them to digital audio data that can be transmitted over the internet. VoIP becomes extremely useful since it converts a standard internet connection into a way to place free phone calls. This can be achieved by using free VoIP software. There are three different flavors of VoIP services commonly used in today. Those are ATA, IP phones, and Computer-to-Computer.

ATA uses a special device known as ATA (analog telephone adaptor) which is used to connect a standard phone to the internet connection. ATA is an analog to digital converter. It takes the analog signal from the traditional phone and converts it into digital data for transmission over the Internet. Providers like Vonage and AT&T CallVantage are bundling ATAs free with their service [2]. IP phones are specialized phones which have both RJ-11 and RJ-45 connectors. IP phones connect directly to the router and have all the hardware and software necessary right onboard to handle the IP call. Computer-to-computer is certainly the easiest way to use VoIP. Skype and Yahoo Messenger are good examples for this type.

One of the most important things to mention is that, VoIP technology is not limited to voice communication. It is applicable to voice, video, and data conferencing. Therefore VoIP has the ability to integrate a stand-alone telephone or a videophone with a personal computer to transmit audio.

This paper will provide a broader view on the key aspects of VoIP communications. In section 2 it will give a description of the overview of voice communications, protocols and how VoIP works. Section 3, provides detailed information of VoIP infrastructure within enterprise. VoIP quality issues are described in section 4. In here, the speech evaluation techniques are explored in more detail. From Section 5, security considerations of VoIP communications are described. Section 6 describes the crimes and forensics under VoIP context. Applications of VoIP are discussed in Section 7. A comprehensive study of IP Multimedia Subsystems is illustrated in section 8. Section 9 describes the future directions of VoIP. Section 10 describes the summary of the findings related to this study.

2. Overview – VoIP Communications

The prime usage of VoIP technology is it that allows people to make voice calls through the public internet [3]. VoIP works by taking analog signals and converting them to digital data that can be sent over a network such as Internet or private network. It requires having a broadband network connection. The following diagram depicts an overview of how VoIP works.

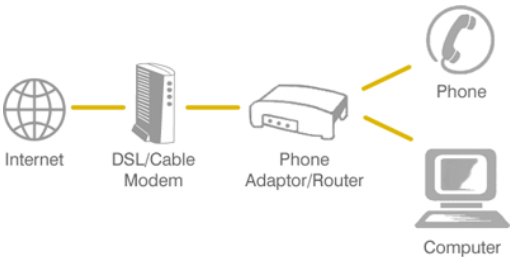


Figure 1 – Overall VoIP Architecture
Source: [4]

Depending on the type of VoIP service, a VoIP call can be made from a computer, a special VoIP phone, or a traditional phone with or without an adapter. In addition, new wireless hot spots in public locations such as airports, parks, and cafes allow to connect to the Internet, and may enable to use VoIP service wirelessly. If the VoIP service provider assigns a regular telephone number, then people can receive calls from regular telephones that don't need special equipment.

There are two main techniques to transfer voice. The current phone system relies on a largely inefficient method for connecting calls known as circuit switching. This technique, which has been used for over 100 years, means that when a call is made between two people a connection is maintained in both directions between callers for the duration of the call. The other technique is known as Packet Switching where the analog call is converted into a digital signal so that it can be sent over an IP network. While circuit switching maintains a constant and open connection, packet switching opens connections just long enough to send bits of data called packets from one computer to another. This allows the network to send the call in packets along the least congested and cheapest lines available.

2.1 How VoIP works

This section provides a detailed explanation of how VoIP works. In order to transmit VoIP traffic, the computer must first sample the sound. The sampling is done at a rate 8 kHz or higher and stored in the computer. When recording the sound samples, the computer might compress those sounds so that they require less space and will certainly record only a limited frequency range. There are a number of ways to compress audio, the algorithm for which is referred to as a compressor - de-compressor, or simply CODEC. Many CODECs exist for a variety of applications and, for VoIP, the CODECs are optimized for compressing voice, which significantly reduce the bandwidth used compared to an uncompressed audio stream. Speech CODECs are optimized to improve spoken words at the expense of sounds outside the frequency range of human speech. Recorded music and other sounds do not generally sound very good when passed through a speech CODEC, but that is perfectly suit for VoIP.

Once the sound is recorded by the computer and compressed into very small samples, the samples are collected together into larger chunks and placed into data packets for transmission over the IP network. This process is referred to as packetization. Generally, a single IP packet will contain 10 or more milliseconds of audio, with 20 or 30 milliseconds being most common.

Through this section, it has been focused on computers that communicate with each other. However, VoIP is certainly not limited to desktop computers. VoIP is implemented in a variety of hardware devices, including IP phones, analog terminal adapters (ATAs), and gateways. In short, a large number of devices can enable VoIP communication, some of which allow one to use traditional telephone devices to interface with the IP networks; one does not have to throw out existing equipment to migrate to VoIP.

2.2 VoIP protocols

There are a number of protocols that may be employed in order to provide for VoIP communication services. Generally there are two classes of protocols for VoIP. Those are,

- Speech transmission protocols
- Signaling protocols

Speech transmission protocols are used in order to transmit audio and video packets between computers. The most commonly used protocols are as below.

- RTP (defined by IETF in RFC 3550 [5]).
- UDP (User Datagram Protocol)
- TCP (Transmission Control Protocol)

The improved version of RTP is called Secure RTP where it provides encryption, authentication, and integrity of the audio and video packets transmitted between communicating devices. Before audio or video media can flow between two computers, various protocols must be employed to find the remote device and to negotiate the means by which media will flow between the two devices. The protocols that are central to this process are referred to as call-signaling protocols. The most commonly used VoIP signaling protocols are as follows: [20]

- Megaco H.248 Gateway Control Protocol
- MGCP Media Gateway Control Protocol
- RVP over IP Remote Voice Protocol Over IP Specification
- SAPv2 Session Announcement Protocol

- SDP Session Description Protocol
- H.323
- SGCP Simple Gateway Control Protocol
- SIP Session Initiation Protocol
- Skinny Skinny Client Control Protocol (SCCP)

Among these protocols, H.323 and SIP gained more popularity. H.323 and SIP both have their origins in 1995 as researchers looked to solve the problem of how two computers can initiate communication in order to exchange audio and video media streams [1]. H.323 created by ITU first emerged in early 1996 whereas IETF worked on SIP (1996).

Fundamentally, H.323 and SIP allow users to do the same thing. That is to establish multimedia communication. However, H.323 and SIP differ significantly in design, with H.323 being a binary protocol and SIP being an ASCII based protocol. Today, H.323 commands the bulk of the VoIP deployments in the service provider market for voice transit, especially for transporting voice calls internationally. H.323 is also widely used in room-based video conferencing systems and is the #1 protocol for IP-based video systems. SIP has, most recently, become more popular for use in instant messaging systems.

Despite the protocols mentioned above, there are some non-standard protocols introduced by various companies that have been very successful in the market. Skype is one such company that has been extremely successful using a proprietary protocol.

2.3 VoIP Codecs

In the VoIP world, codecs are used to encode voice for transmission across IP networks. Codecs generally provide a compression capability to save network bandwidth. Following table depicts some of the codecs along with their corresponding algorithms and bit rates.

Codec	**Algorithm**	**Bit Rate (Kbps)**
ITU G.711	PCM (Pulse Code Modulation)	64
ITU G.722	SBADPCM (Sub-Band Adaptive Differential Pulse Code Modulation)	48, 56 and 64
ITU G.723	Multi-rate Coder	5.3 and 6.4
ITU G.726	ADPCM (Adaptive Differential Pulse Code Modulation)	16, 24, 32, and 40
ITU G.727	Variable-Rate ADPCM	16-40
ITU G.728	LD-CELP (Low-Delay Code Excited Linear Prediction)	16
ITU G.729	CS-ACELP (Conjugate Structure Algebraic-Code Excited Linear Prediction)	8

Table 1 – VoIP codecs
Source: [6]

3. VoIP Infrastructure

Traditionally, enterprises have maintained two separate networks, voice network based on PBX and PSTN and IP network for data applications such as email, web, VPN etc [7]. Voice over IP (VoIP) technology allows voice and data networks to be merged into a single network because voice can be treated as just another application running over the IP (data) network. The main advantage of using a single network for voice and data is that the IT department in the enterprise now only needs to manage one kind of network. Indeed, many enterprises have chosen to deploy a single IP network in their new office buildings where there are no existing analog (PSTN) phone extensions [8].

For an example consider an enterprise network with desktop and notebook clients connected to LAN and WLAN segments. The WLAN segment can be operated by an 802.11 access point (AP). These segments can be connected to the corporate network via an Ethernet switch. Other components on the corporate network would be enterprise servers, a hybrid PBX, legacy phones and VoIP phones. A VoIP phone can be a desktop phone device with Ethernet connectivity or a PC with a VoIP application running on it.

3.1 VoIP Equipment

Any use of VoIP will require having a reliable power source and a broadband Internet connection. Here's a quick look at the various types of VoIP-specific equipment that might also be a part of the infrastructure:

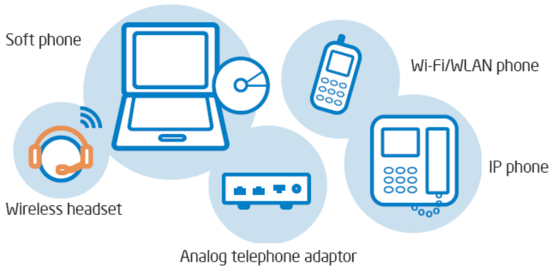


Figure 2 – VoIP equipments
Source: [9]

ATA-An ATA (analog telephone adaptor) is a device that converts analog signals to digital data. It allows to connect a standard phone to the Internet connection for use with VoIP. ATAs are sometimes referred to as VoIP gateways.

Softphone-A soft phone is actually a software application that can be installed on the computer to create a VoIP user interface. In order to use a soft phone, a headset and/or microphone is needed.

IP phone-An IP phone, or hard phone, is a self-contained piece of equipment (that looks like a regular phone) that can communicate directly via the Internet connection. Phones don't have to be routed through the computer or an ATA, and don't require any software.

Wi-Fi/WLAN Phone-Like IP phones, Wi-Fi/WLAN phones don't require a computer or ATA to use VoIP. They link directly to the IP Internet connection. Unlike IP phones, they're wireless and connect to the Internet via a wireless base station.

4. VoIP quality and performance

Unlike media streaming, VoIP communication is interactive, so that participants are both speakers and listeners at the same time. High-quality VoIP services are required as for the Internet communications to be an alternative towards Public Switched Telephone Network (PSTN) [10]. In this respect, delays higher than 100-150 msec can greatly impair the interactivity of conversations, and therefore delayed packets are usually dropped by the receiver codec [11]. As mentioned in earlier section, VoIP codecs use buffering at the receiver side to compensate for slightly delayed packets, and use forward error correction (FEC) or packet loss concealment (PLC) mechanisms to overcome the effect of packet loss or excessive delay [11].

4.1 Factors that affect the speech quality

IP networks do not guarantee reliable packet delivery and, regarding real-time voice applications such as VoIP, some factors have to be taken into account to provide quality of service as good as in switched telephony [12]. These factors should be monitored in order to manage a VoIP system performance.

Codecs: these are voice signal encoder/decoder algorithms that perform speech compression in order to better utilize the available network capacity (bandwidth saving). Certain aspects must be taken into account when comparing different codecs: bit rate, delay, packet loss robustness and voice quality.

Delay: ITU-T Rec. G.114 recommends 150ms as the maximum allowable end-to-end delivery value for VoIP systems [12]. As end-to-end delay reaches values above this, conversation becomes less interactive, and users become more likely to accidentally talk over each other.

Packet loss: Basically, there are three main reasons for voice packet loss in a VoIP communication: transmission errors, discarded packets at the network routers and at the de-jitter buffer [12]. The loss is calculated using expected values so as to allow more significance for the number of packets received [13]. To reduce packet loss impairment on speech quality, some codecs implement packet repair techniques, known as PLC (Packet Loss Concealment).

Jitter: Jitter is the statistical variance of the packet inter arrival time [13]. Jitter problems arise because of delays occurring in the operational system and at the packet queue in the network nodes. The problem involving delay variation can be solved by the de-jitter buffer, which stores voice packets for a certain period of time, delaying their delivery time sufficiently to enable the majority of packets to be played within the scheduled time [12]. Packets that arrive too late are considered lost by the buffer. Those that are received in advance wait for their turn to be sent to the decoder.

4.2 Speech evaluation techniques

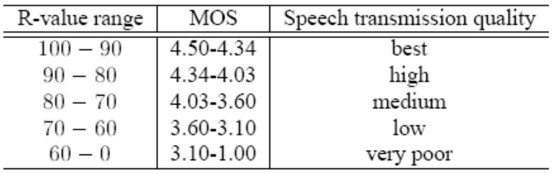
Speech quality assessment techniques has been developed by researchers to measure how near a given processed speech is to natural speech. This implies that speech quality evaluation is a comparison procedure, even though sometimes the comparison is not explicit.

MOS (Mean Opinion Score) tests, defined on ITU-T Rec. P.800 [10], require that some people listen to voice samples and evaluate their voice quality in five grades, from 5 (excellent) to 1 (bad). By this procedure, listeners do not compare the samples, but they use their previous auditory experience to rate voice sample quality. The arithmetic mean of all collected opinion scores is the MOS.[12] Objective methods have been developed in order to automatically estimate MOS. There two main objective methods**.**

- PESQ (Perceptual Evaluation of Speech Quality)
- E-model (ITU-T Rec. G.107)

PESQ, defined by ITU-T Rec. P.862 evaluates the speech quality of a transmitted (and thus degraded) voice signal by comparing it with the correspondent reference signal in the time–frequency domain, providing an objective quality measure that may be mapped to a MOS score.[12]

The E-model is a computational methodology that predicts the subjective quality that will be experienced by an average listener combining the impairment caused by transmission parameters (such as loss and delay) into a single rating. The rating can then be used to predict subjective user reactions, such as the Mean Opinion Score [14]According to ITU-T Recommendation G.107, every rating value corresponds to a speech transmission category, as shown in Table 1. A rating below 60 indicates unacceptable quality, while values above 70 correspond to PSTN quality and values above 90 corresponding to very good quality [14].



R-value range	MOS	Speech transmission quality
100 − 90	4.50-4.34	best
90 − 80	4.34-4.03	high
80 − 70	4.03-3.60	medium
70 − 60	3.60-3.10	low
60 − 0	3.10-1.00	very poor

Table 2 – Ratings and corresponding quality
Source: [14]

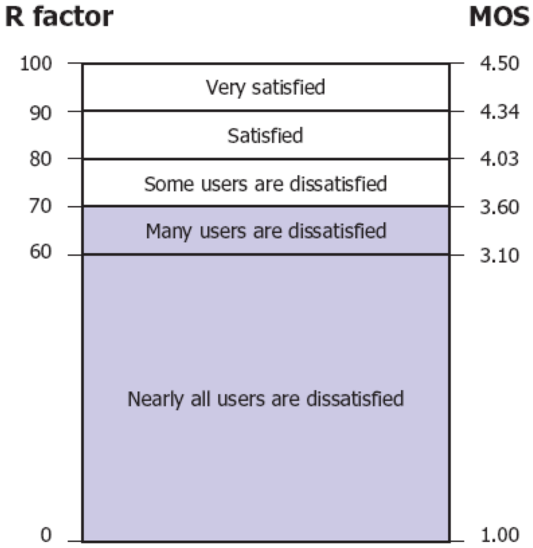


Figure 3 – Speech quality categories of user's satisfaction
Source: [12]

4.3 Quality management in VoIP

As many VoIP services become available and grow in acceptance among users, the need to manage the quality of these services arises, in order to guarantee an understandable speech quality level. A typical network management system establishes a set of activities for monitoring, analysis, and control of network resources, working together in a control loop to provide services to customers with a certain level of quality.[12]

Monitoring tasks are concerned with collecting data about activities and events. This data may be logged, displayed in some form for a human manager, or may be used directly as input to analysis and/or control functions [12, 15]. *Analysis* tasks try to make sense of the activity data in an attempt to ascertain the 'health' of the overall system and of its parts. Ultimately, the goal of analysis is to influence the control management function in general [12,15]. *Control* tasks permit assertive actions to be applied to various parts of the system. These actions may be human-invoked or may arise as natural occurrences in an automated feedback path [12]. For example, based on input from the analysis function, the system configuration can be altered in some way to attempt to better meet the overall speech quality.

For instance, this study concerned only with speech quality monitoring of the voice flow. Signaling monitoring is also desirable to reach a good network management control, but it is a subject for further studies. SNMP is usually employed for management purposes on IP networks [12].
A typical management architecture for VoIP networks comprises six interrelated entities, as depicted in Figure 5:

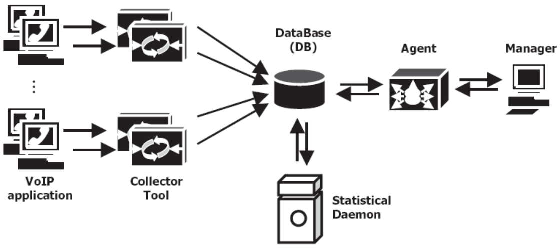


Figure 4 – VoIP management architecture
Source: [12]

VoIP Application: This can be any softphone able to generate, for each established call, a trace file containing signaling and voice flow events.

Collector Tool: This extracts in real-time the speech quality information from the trace file of VoIP applications, such as those described in earlier (codec type, delay, network and de-jitter buffer packet loss, among others), to compute the MOS value. It updates the database about speech quality data after the end of each monitored call.

Database (DB): This stores the speech quality reports sent by the collector tool and makes it available to the statistical daemon. Moreover, it stores the statistical calculation made by the statistical daemon and makes it available to the agent.

Statistical Daemon: This performs statistical calculation using the speech quality information stored in the database by the collector tool and delivers these results to be stored in the database.

These tasks (reading and writing the DB) are performed at scheduled times.

5. Security Issues for VoIP

With the introduction of VOIP, the need for security is compounded because now we must protect two invaluable assets, our data and our voice. Government agencies are required by law to protect a great deal of information, even if it is unclassified. Both privacy-sensitive and financial data must be protected, as well as other government information that is categorized as sensitive but unclassified. Protecting the security of conversations is thus required.

The current Internet architecture does not provide the same physical wire security as the phone lines. Effect of security threats and possible security weaknesses in VoIP features and implementation should be considered, so user authentication and authorization, along with software, should be carefully implemented and monitored [16]. The key to securing VOIP is to use the security mechanisms like those deployed in data networks (firewalls, encryption, etc.) to emulate the security level currently enjoyed by PSTN network users [17]. This section describes the attacks and defenses relevant to VOIP and explores ways to provide appropriate levels of security for VOIP networks at reasonable cost.

5.1 Firewalls

Firewalls are a staple of security in today's IP networks. Whether protecting a LAN, WAN, or just protecting a single computer, a firewall is usually the first line of defense against would be attackers. Firewalls work by blocking traffic deemed to be invasive, intrusive, or just plain malicious from flowing through them. If networks are castles, firewalls are the drawbridges. Traffic not meeting the requirements of the firewall is dropped. Processing of traffic is determined by a set of rules programmed into the firewall by the network administrator. Firewalls simplify security management by consolidating security measures at the firewall gateway, instead of requiring all the endpoints to maintain up to date security policies. This takes an enormous burden off the VOIP network infrastructure [17].

Most VOIP traffic travels across UDP ports. Firewalls typically process such traffic using a technique called packet filtering. Packet filtering investigates the headers of each packet attempting to cross the firewall and uses the IP addresses, port numbers, and protocol type (collectively known as the 5-tuple) contained therein to determine the packets' legitimacy [17]. In VOIP and other media streaming protocols, this information can also be used to distinguish between the start of a connection and an established connection. There are two types of packet filtering firewalls, stateless and stateful. Stateless firewalls retain no memory of traffic that has occurred earlier in the session. Stateful firewalls do remember previous traffic and can also investigate the application data in a packet. Thus, stateful firewalls can handle application traffic that may not be destined for a static port.

5.2 VOIP specific Firewall Needs

In addition to the standard firewall practices, firewalls are often deployed in VOIP networks with the added responsibility of brokering the data flow between the voice and data segments of the network. Thus, it is recommended that all PC-based phones be placed behind a stateful firewall to broker VOIP media traffic. Without such a mechanism, a UDP DoS attack could compromise the network by exploiting the open ports. Some specific needs can be identified where firewalls are necessary.

- PC-Based IP phones require access to the segment to place calls, leave messages
- IP Phones and call managers accessing voice mail
- Users accessing the proxy server
- The proxy server accessing network resources
- Traffic from IP Phones to the call processing manager or proxy server must pass through the firewall because such contacts use the data segment as an intermediary

5.3 Network Address Translation

Network Address Translation is a powerful tool that can be used to hide internal network addresses and enable several endpoints within a LAN to share the same (external) IP address. NATs also indirectly contribute to security for a LAN, making internal IP addresses less accessible from the public Internet. Thus, all attacks against the network must be focused at the NAT router itself. Like firewalls, this provides security because only one point of access must be protected, and the router will generally be far more secure than a PC directly connected to the Internet (less likelihood of open ports, malicious programs, etc.). The abstraction of the LAN from the Internet through a NAT also simplifies network management.

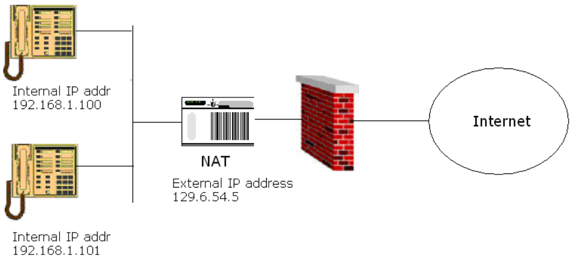


Figure 5 – IP Telephones Behind NAT and Firewall
Source: [17]

5.4 IPSec

Firewalls, gateways, and other such devices can help keep intruders from compromising a network, but firewalls are no defense against an internal hacker. Another layer of defense is necessary at the protocol level to protect the data itself. In VOIP, as in data networks, this can be accomplished by encrypting the packets at the IP level using IPsec.

5.5 Encryption at the End points

Security may lead to bottlenecks. One solution to the bottlenecking at the routers due to the encryption issues is to handle encryption/decryption solely at the endpoints in the VOIP network. One consideration with this method is that the endpoints must be computationally powerful enough to handle the encryption mechanism. But typically endpoints are less powerful than gateways, which can leverage hardware acceleration across multiple clients.

5.6 Secure Real Time Protocol (SRTP)

The Secure Real-time Protocol is an extension of the Real-time Transport Protocol (RTP) offering not only confidentiality, but also message authentication, and replay protection for the RTP traffic as well as RTCP (Real-time Transport Control Protocol).

5.7 Security issues of SIP

SIP is a real time signaling protocol for IP voice that was developed by the IETF. This protocol is used for a bidirectional communication session initiation in which its messages are exchanged between two or more nodes. The basic call control tasks such as communication session setup and tear down, or signalling for call initiation, dial tone and termination are considered as SIP's responsibilities.

SIP consists of different components including User Agent (UA), redirect server, registrar server, location server, and proxy server. UA software includes client and server components. The outgoing calls are made at the client side, while incoming calls are received at the server site. After some processing or translation is done, traffic forwarding is performed by a proxy server. Request aunthentication is done by the registrar server, and the redirect server is responsible for resolving information for UA clients. UA clients send requests to UA servers for call initiating [24].

A major source of vulnerabilities lies in VoIP specific protocols. SIP is text encoded which makes it easier to analyze with standard parsing tools such as Perl. SIP traffic is plain text in its basic form, so the voice traffic is vulnerable to packet sniffers (looking for caller IDs or passwords and allows an attacker to forge packets for manipulating of device and call state). For example, this kind of manipulation results in terminating call prematurely, redirecting calls, and making payment fraud easy to perform. It is also relatively easy to intercept unencrypted VoIP calls. Hackers can download free software available on the Internet and intercept. To protect caller IDs, account information, SIP traffic needs to be encrypted. Although some efforts have been made for developing encrypted signalling but so far no widespread solution has been found to adopt.

There are specific attacks based on the vulnerabilities in the VoIP signalling protocols, including SIP. One of these attacks is known as BYE attack [24]. The goal of a BYE attack is to tear down an existing communication session prematurely which can be considered as a DoS attack.

Some Intrusion Detection Systems (IDSs) are able to detect this kind of attack. By creating a rule an IDS can detect flowing of RTP packets after a BYE message. In the IDS, an alarm is activated if RTP flow is detected after receiving a BYE message.

The SIP protocol is not only used in VoIP call setup, but also in Instant Messaging. An attacker can also manipulate the header of an instant message and send a forged message to the receiver. This attack is called Fake Instant Messaging.

There is also another signalling-based attack which is called Call Hijacking. In this attack, the attacker takes advantage of REINVITE message which is used for call migrating. In the case of Call Hijacking attacks, one of the UA clients will experience continuous silence since the other part is not sending voice packets to it any more. This kind of attack can also be considered as a DoS attack. This attack results in breach of confidentiality. For detecting this kind of attack a similar approach for a BYE attack can be used [24].

5.8 Botnet Threats

A DoS attack is achieved by preventing the legitimate use of a service. Distributed DoS (DDoS) attacks take advantage of multiple systems in the network to launch a DoS attack. To launch a DDoS attack, an attacker first looks for multiple vulnerable agent machines. This scanning process is normally performed automatically through the scanning of remote systems, looking for potential vulnerabilities. When the vulnerability is found, it is exploited to break into the machine. Then the attacker sends remote control programs called 'bot' to these selected machines.

After planting the remote control programs on the bot-infected machines (called zombies), those machines wait for commands coming from the attacker (bot-herder) [24]. So an attacker can direct a large number of compromised systems against a target. The network of these bot-infected computers is called botnet. These compromised systems normally selected among vulnerable computers which are high bandwidth and always on. The real identity of an attacker is always hidden by attackers through IP spoofing in which the source address field of attack packets is spoofed [24].

In a DoS attack, a server can be targeted with a flood of information requests which may bring the system down. Botnets are normally controlled through Internet Relay Chat (IRC) channels where bot-infected machines listen for instructions coming from the bot-herder [24].

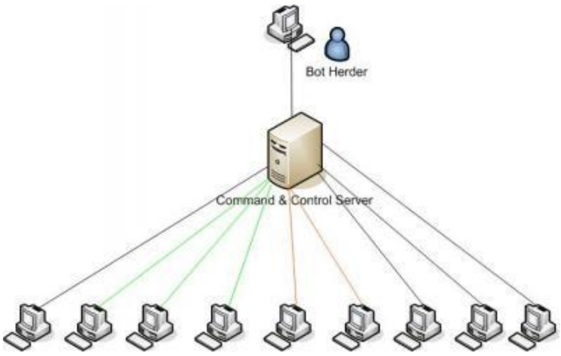


Figure 6 – A basic botnet
Source: (www.naspa.com)

To catch the attacker, investigators monitor IRC channels and also can block traffics to IRC channels which are used by hacked machines. VoIP is inherently vulnerable to network attacks including DoS and DDoS [24]. VoIP applications like Vonage and Skype could give better opportunities to attackers for controlling their bot-infected machines.

5.8 Spam over VoIP

Just as email systems, VoIP is also susceptible to spam (also called SPIT, for Spam over Internet Telephony). SPIT can be considered as another threat of botnets, which could potentially disable the VoIP system. If a VoIP user gets a lot of calls from audio spammer every day that makes him to be reluctant to employ VoIP technology. SPIT is even worse than spam, if we get our Emails with even a few minutes delay it is not a big problem while SPIT is very noticeable to end users as it hits the gateways and degrades the quality of voice. SPIT can target any IP based phone systems [24].

To address the SPIT issue, several companies are developing solutions for this kind of threat. Techniques for combating SPIT include filtering, black/white lists, and caller's reputation. All of them are becoming less effective as SPIT methods become clever.

6. VoIP Forensics

With the tremendous growth in popularity and bandwidth of the Internet, VoIP technology has emerged that allows phone calls to be routed over Internet infrastructure rather than the traditional Public Switched Telephone Network (PSTN) infrastructure. The issues faced by law enforcement authorities concerning VoIP are very different from that of traditional telephony. Wiretapping is not applicable to VoIP calls and packet capturing is negated by encryption [25]. Therefore it's important to explore the methods that electronic evidence may be collected from systems where VoIP conversations play an important role in suspected criminal activity or communications.

6.1 VoIP and Crime

The popularity of VoIP is increasing as the cost savings and ease of use is realised by a wide range of people and corporations. The technology is attractive to criminals, especially the non-carrier VoIP, as it often does not require verification of any details to commence using the service. The security of placing such calls may also be appealing to criminals, as many implementations use strong encryption to secure both the voice payload as well as control messages. Skype uses 256 bit AES encryption (Skype Privacy FAQ 2006) while Google Talk does not encrypt its payload [25].

It is essential that computer forensic research evaluate the use of VoIP technology and formulate methods to allow law enforcement agencies to overcome some of the aspects that are advantageous to criminals. Wire-tapping is not applicable to VoIP communications and therefore other methods of recovering evidence and information are required.

6.2 Evidence Collection

Several techniques were used in the process of recovering VoIP evidence. A particular scenario is that a suspect’s computer has been used to make VoIP calls and has not been powered off after the conversation.

The first step in the process is to acquire the memory image from the target system. The subsequent steps involve finding packets within the image, identifying the type of packets, extracting the payload from RTP packets and recreating audio files by reconstituting consecutive packet payloads into one file. Each step in this process requires a practical solution and consequently both existing and new solutions have been either used or developed.

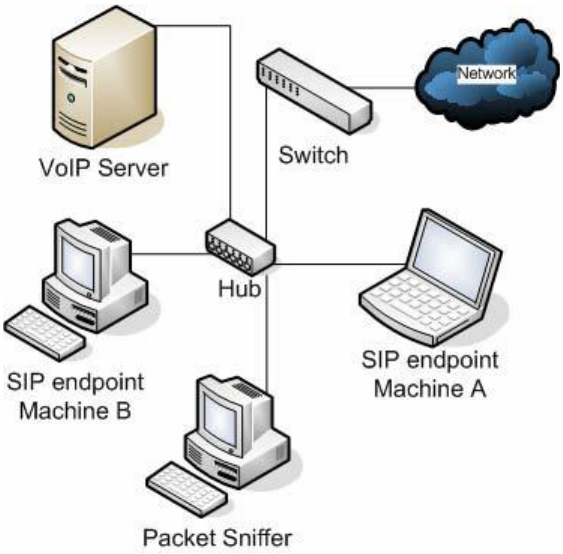


Figure 7 – The setup for evidence collection
Source: [25]

The technique of imaging memory uses an existing tool created for acquiring memory images under Microsoft Windows XP. The algorithm for finding packets in a binary image is a newly created method formed to address this specific problem. Details of the algorithm, and the process used to recover evidence from the memory image of a target machine are currently not open for disclosure.

7. Applications of VoIP

VoIP could be applied to any voice communications requirement, ranging from a simple inter office environment to complex multi-point teleconferencing environments passing through applications like Net2Phone and voice mails. A small office could gain access to corporate voice, data, and facsimile services using the company's Intranet. Voice calls from a mobile PC via the Internet can be achieved to cut unnecessary expenses. All of these benefits are possible through the use of special VoIP software. Even though there are number of softwares available today, (like Messenger and Skype), I primarily focus on Cisco Unified Communications for the purpose of this study.

Cisco Unified Communications comes with a set of applications that gives the companies the ability to efficiently access data on demand. Those applications are described as below.

Cisco Unified Communications Manager

Cisco Unified Communications Manager is the software-based call-processing component of the Cisco Business Communications Solution. Cisco Unified Communications Manager software extends telephony features and capabilities to packet telephony network devices such as IP phones, media processing devices, voice over IP (VoIP) gateways, and multimedia applications.[18]

Cisco Unified Communications Manager Express

Cisco Unified Communications Manager Express enables quick, efficient, feature rich call processing in medium-sized companies and branch offices. Cisco Unified Communications Manager Express can now play the role of survivability for a branch office, allowing IP phones to use the centralized Cisco Unified Communications Manager as their primary call processing and Cisco Unified Communications Manager Express for survivability. [18]

Cisco Unified Personal Communicator

This is a Cisco Unified Communications client software which transparently integrates a user’s most frequently used communication applications and services into a single, unified client. From an easy-to-use interface on the PC or Mac, it gives users quick access to powerful communication tools voice, video, instant messaging (IM), and Web conferencing. [19]

Cisco IP Communicator

Mobile employees can use the Cisco IP Communicator as an alternative to a physical IP phone handset. This software-based application delivers enhanced telephony support through a VPN connection from a personal computer. [18]

8. IP Multimedia Subsystems

As the mobile phones and cellular networks are becoming extremely popular, VoIP providers tend to target mobile devices for offering robust services to customers. Voice over 3G is an emerging trend in mobile networks which delivers voice/video telephony (based on IP) between 3G handheld devices. 3G networks are ideal for IP telephony because of its high bandwidth and improved QoS. But the competition in mobile market drives service providers for rolling out new multimedia services that span across WiFi, 2.5G, 3G, CDMA, GPRS, and wireline networks. This demands a standardized platform which allows adopting a service created by third parties and create a service that integrates with other services.

Therefore, IP multimedia subsystems emerged as the Next Generation Networking architectural solution for implementing IP based telephony and multimedia services. IMS was initially defined by the 3rd Generation Partnership Project (3GPP) [21]. In fact, IMS is a new mobile network infrastructure that enables the convergence of data, speech and mobile network technology over an IP-based infrastructure. It fills the gap between the existing traditional telecommunications technology and internet technology and act as a way to offer Internet services everywhere using cellular technology. IMS was specifically architected to enable and enhance real time, multimedia mobile services such as rich voice services, video telephony, messaging, conferencing, and push services. IMS enables these user-to-user communication services via a number of key mechanisms including session negotiation and management, Quality of Service (QoS) and mobility management [22].

8.1 IMS Architecture

IMS architecture supports a wide range of services that are enabled based on SIP protocols. IMS architecture delivers multimedia services that can be accessed by a user from various devices via an IP network or traditional telephony system. The underlying network architecture can be divided into four layers such as Device Layer, Transport Layer, and Control Layer and the Service Layer [23].

Device Layer - The IMS devices such as computers, mobile phones, PDAs, and digital phones are able to connect to the IMS infrastructure via the network. Devices like analog telephones can be connected to an IP network via PSTN gateways.

Transport Layer - This layer is responsible for initiating and terminating SIP sessions and providing conversion of data into IP packet format. Devices connect to the IP network in the transport layer via a variety of transmission media, including WiFi, SIP, GPRS, and WCDMA.

Control Layer – SIP server which is known as CSCF is the core element in this layer. It processes the SIP signal in the appropriate application server. Another element in the control layer is the Home Subscriber Server (HSS) database that stores the unique service profile for each end user.

Service Layer – This layer resides on top of the IMS architecture. The services are all run by application servers. The application servers are not only responsible for hosting and executing the services, but also provide the interface against the control layers using the SIP protocol.

8.2 IMS Service Examples

The mobile industry is in a transition phase from traditional voice system to a variety of new multimedia services and applications. These include, two-way radio sessions (Push-to-Talk), sharing a view, sharing files, shared whiteboards and multiplayer game experiences. It also provides the ability to combine existing services effectively, for example when playing a multiplayer game during a call. Following are some examples in detail.

Real Time Video Sharing

A real time video sharing service is a peer-to-peer, multimedia streaming service that can be offered entirely as a packet switched service or as a combinational service, combining the capabilities of the circuit switched and IMS packet switched domains. The media is delivered and consumed almost real-time, with only a marginal delay, thus providing the experience of being there and sharing the moment. The spirit is always live, even when sharing a stored video clip, since there is the possibility that users can have an ongoing voice conversation at the same time.

Interactive Gaming

When considering the end-user demand for basic gaming with mobile terminals, it’s potentially a good revenue for service providers to establish interactive gaming sessions between players.

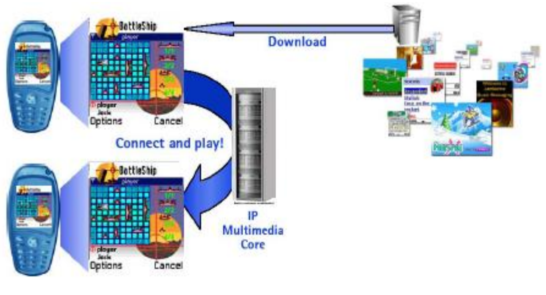


Figure 8 – Interactive Applications
Source: [22]

Shared Folders

Content sharing enables users to share files between terminals. A typical use case includes sharing of files or contents such as images, documents, notes, contacts or calendar information, even simultaneously while in a voice-call session.

Instant messaging

Instant Messaging is a communication service that allows end-users to send and receive messages instantly. Messages can contain any MIME type media content such as text, image, audio or video clips, application data or a combination of these. The message is sent through the packet data network to the IP Multimedia Subsystem (IMS), which locates the terminating IP client and routes the message to the recipient.

IMS enabled Voice and Video Telephony

IMS enabled Voice and Video calls are carried over a packet switched network (VoIP). Video Telephony is seen as a critical end user service in mobile networks. The Session Initiation Protocol (SIP) enables Voice and Video calls over an IP network.

Video-conferencing

IP Multimedia Subsystem (IMS) Video-conferencing service extends the point-to-point video call to a multi-point service. Videoconferencing requires an IMS conference bridge service, which links the multiple point-to-point video calls together. The video telephony connections are made point-to-point from the terminals to the conference bridge, which takes care of joining the point-to-point connections into a conference.

The conference bridge is not concerned about the underlying infrastructure and client devices and assumes that audio- and video connections are provided by the appropriate standard and that these connections are delivered over the IP network [22].

8.3 Benefits of IMS

The benefits of IMS over the existing cellular network infrastructure can be demonstrated in the following four aspects.

Time to market

One of the biggest challenges in today's communication network is to improve the long and costly process for creating a new service. Service providers are looking for ways to reduce the time-to-market for rolling out new multimedia services. Additional new services can be provided by third parties. This multi-vendor service creation industry leads to an open market, and allows service providers to choose the most effective way to roll out new services.

QoS enablement

Quality of Service (QoS) mechanisms were developed in order to provide certain guarantee levels of network bandwidth during transmission. IMS specifies enablement of Quality of Service within the IP network and takes advantage of the QoS mechanism to improve and guarantee the transmission quality.

Charge appropriately

The advantage of IMS is that it provides information about the service type being invoked by the user and allows the operators to determine how to charge the users based on service types. They can choose to charge user by the number of bytes transferred, by the session duration (time-based), or create any new type of charging scheme.

Location Independency

A typical problem when working with cellular technology is that some of the services will not be available when the user is roaming in another country. To resolve this problem, IMS uses Internet technologies and protocols in order to allow users to move across the countries and still be able to execute all the services as if they were from their home networks [23].

9. The future directions of VoIP

VoIP has started a new revolution in the history of Internet Telephony. With more software companies implementing VoIP, it is recreating itself with the advent of the new technology. It is believed that future of VoIP is VoIP 2.0, which will focus on services instead of cut-rate pricing. Though companies like Google, Skype and Microsoft are offering new services now, the day is not far when every company will be forced to charge for this service after adding new features and technologies.

Due to inevitable competition, traditional service providers try to erode the revenues gained by new service providers by initiating a new effort known as Next Generation Network (NGN). It is a packet-based network able to provide telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies [1]. IP multimedia subsystem is at the core of NGN, thus NGN can rightfully be viewed as a very-much voice-centric effort with no real desire to grow and encourage other non-voice services [1].

ITU-T is trying to overcome that limitation through the development of AMS, which is a new multimedia communication system that integrated audio, video, and data communication through a new modular system design. With this new protocol, a person might use a mobile phone to initiate a phone call, use a Bluetooth headset for audio, use a PC for application sharing, and view a video stream on a separate LCD screen within a single session.

10. Discussion

Under this study, I tried to explore the key components of VoIP communications. So far in this study I discovered how VoIP works and its architecture overview. The bulk of the work is carried out by the VoIP protocols and codecs. Transmission of packets between computers is done by Speech transmission protocols where as signaling protocols help to negotiate and find remote devices. Codecs are used to compress the sampled audio so that the bandwidth can be preserved when transmitting through highly congested networks. The VoIP infrastructure within an enterprise allow voice and data networks to be merged to a single network so that network management becomes easier. The types of equipment used in this kind of an environment are ATA, Softphone, IP phone and Wi-Fi phone. The real benefit of VoIP can be gained only if it is quality and reliable. The major factors that affect the quality are delay, packet loss, jitter and codecs. Quality should be continuously monitored using tests like MOS. Security considerations for VoIP involves firewalls, NAT, SRTP and encryption. VoIP forensics investigation concerns in collecting memory images, extracting the payload from RTP packets and reconstructing the audio file. VoIP applications such as Yahoo Messenger, Skype, and Cisco Unified Communications are used to leverage the voice communication requirements within home and office environments. The next generation VoIP services are known as VoIP 2.0 in which more advanced and powerful features are introduced.

In this paper most of the aspects of VoIP are discussed. But areas such as VoIP protocols are broader subjects where some portions are still at research. These protocols such as Session Initiation Protocol are intended to study in more deeply in future work. Authentication capabilities, SIP requests and responses are needed to be further studied.

Furthermore, experiments for consolidating VoIP with WiMax are carried out worldwide. This leads to produce hybrid mobile phones with an inbuilt detector that tries to sense the VoIP availability so that phone will automatically switches to VoIP if it detects Wi-Fi hotspots. Otherwise it will operate through the regular cellular network. A city-wide wireless network like Wi-Max will support bulky transmission of data where video and audio can be shared in real time. Areas like this are still at research level and will revolutionize the global communication with free Internet Telephony and other multimedia services.

Acknowledgement

It is with great pleasure in acknowledging the support and guidance given to me in preparing this document. Firstly I would like to thank Professor A. S. Karunananda for his significant, yet supportive, supervision and guidance, and giving me the opportunity to explore a very interesting study area. Next my supervisor Prof. Dileeka Dias for her assistance, advice, cooperation and her valuable time granted in favour of preparing this report.

And I would like to thank all the lecturers and the colleagues of my Faculty for their support and corporation.

Reference

[1]http://www.packetizer.com/ipmc/papers/understanding_voip/voip_introduction.html,
Understanding VoIP, Accessed on 10 February 2008

[2] http://communication.howstuffworks.com/ip-telephony.htm
HowStuffWorks, Accessed on 11 February 2008

[3] Xinyuan Wang, Shiping Chen, and Sushil Jajodia, *Tracking Anonymous PeertoPeer VoIP Calls on the Internet,* Proceedings of the 12th ACM conference on Computer and communications security, 2005

[4] http://www.inphonex.com/main/how-voip-works.php, Inphonex, Accessed on 12 February 2008

[5]http://www.packetizer.com/ipmc/papers/understanding_voip/voip_protocols.html
Understanding VoIP, Accessed on 12 February 2008

[6] http://www.tech-faq.com/voip-codec.shtml,
The tech-FAQ, Accessed on 12 February 2008

[7] Bakre Ajay, *Intel VoIP over WLAN architecture,* Proceedings of the 2nd annual international workshop on Wireless Internet, August 2006

[8] Cisco Systems Case Study, *KB Home Responds Quickly to Business Cycles,*
Available at - http://www.cisco.com/warp/public/cc/so/neso/vvda/iptl/kbhom_ss.pdf, Accessed on 12 February 2008

[9] Intel, *An Introduction to the Basics of Voice Over Internet Protocol,*
Available at- download.intel.com/smallbusiness/products/laptop/howto_voip.pdf, Accessed on 12 February 2008

[10] Muhamad Amin A.H. , *Voip Performance Measurement Using Qos Parameters,* The Second International Conference on Innovations in Information Technology (IIT'05), 2005

[11] Yair Amir, Claudiu Danilov, Stuart Goose, David Hedqvist, and Andreas Terzis, *Using Overlay Networks to Improve VoIP Quality*, 2004,
Available at- www.cnds.jhu.edu/pub/papers/cnds-2004-2.pdf, Accessed on 12 February 2008

[12] Ana Flàvia M. de Lima, Leandro S. G. de Carvalho, José Neuman de Souza,and Edjair de Souza Mota , *A framework for network quality monitoring in the VoIP environment,*
International Journal of Network Management, Volume 17 , Issue 4 (July August 2007)

[13] Ian Marsh, Fengyi Li, and Gunnar Karlsson. , *Wide Area Measurements of Voice Over IP Quality,* 2003
Available at-
www.sics.se/~ianm/Papers/qofis2003.pdf
Accessed on 12 February 2008

[14] Catherine Boutremans, Gianluca Iannaccone, and Christophe Diot , *Impact of link failures on VoIP performance,*
International Workshop on Network and Operating System Support for Digital Audio and Video, 2002, ISBN:1-58113-512-2

[15] Howard SL, Hong JW, Katchabaw MJ, and Bauer MA, *Integrating visualization into event monitoring and analysis in distributed systems*, IBM Centre for Advanced Studies Conference, 1995

[16] Upkar Varshney, Andy Snow, Matt McGivern, and Christi Howard, *Voice over IP,* Communications of the ACM, Volume 45 , Issue 1 (January 2002)

[17] D. Richard Kuhn, Thomas J. Walsh, and Steffen Fries. *Security Considerations for Voice Over IP Systems,*
Available at-
http://csrc.nist.gov/publications/nistpubs/800-58/SP800-58-final.pdf
Accessed on 12 February 2008

[18] Cisco. *Cisco Unified Communications Solutions Blueprint,*
Available at-
www.cisco.com/en/US/solutions/collateral/ns339/ns639/ns641/net_implementation_white_paper0900aecd806002d7.pdf
Accessed on 13 February 2008

[19] Cisco. *Cisco Unified Communications Solution Overview*
Available at-
www.communications.siemens.co.uk/enterprise/downloads/bo34_unified_comms_overview_260905.pdf
Accessed on 13 February 2008

[20] *http://www.protocols.com/pbook/VoIPFamily.htm,*
Protocols.com, Accessed on 13 February 2008

[21]*http://www.ibm.com/developerworks/webservices/library/ws-soa-ipmultisub1/,*
Introduction to IP Multimedia Subsystem, Accessed on 13 February 2008

[22] 3G Americas, *IMS Overview and Applications,* (July 2004)
Available at-
http://www.3gamericas.org/pdfs/ims_july2004.pdf
Accessed on 09 May 2008

[23]*http://www.ibm.com/developerworks/webservices/library/ws-soa-ipmultisub1/,*Introduction to IP Multimedia Subsystem, Accessed on 09 May 2008

[24] Marjan Zandi, Miguel Vargas Martin, and Patrick C.K. Hung, *Overview Of Security Issues Of VoIP*
Internet and Multimedia Systems and Applications, March 2007, Accessed on 11 May 2008

[25] Jill Slay, Matthew Simon, *Voice over IP Forensics,*
1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop, (January 2008),
Available at- http://portal.acm.org/
Accessed on 11 May 2008