ADDIS ABABA UNVIERSITY TECNOLOGY

FACALITY SCHOOLE OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

THE IMPACT OF EMPLOYING SECURITY MECHANISM ON VOIP NETWORK PERFORMANCE

MSc. Thesis Submitted
By Abebe Demissie

Addis Ababa
March 2008

ADDIS ABABA UNVIERSITY TECNOLOGY FACALITY

SCHOOLE OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL AND COMPUTER

ENGINEERING

THE IMPACT OF EMPLOYING SECURITY MECHANISM ON VOIP NETWORK PERFORMANCE

By
Abebe Demissie

A thesis submitted to the school of Graduate Studies, Addis Ababa University Technology Faculty in partial fulfilment of the requirements

For the Degree of Master of Science in Electrical Engineering (Micro Electronics Engineering).

Advisor: Dr. Mohammed Abdo

Addis Ababa

March 2008

THE IMPACT OF EMPLOYING SECURITY MECHANISM ON VOIP NETWORK PERFORMANCE

By
Abebe Demissie

Addis Ababa University, Faculty Technology, Department of Electrical and Computer Engineering

Approval by Board of Examiners

1. Dr. Mengesha Mamo _____

Chairman, Departmental Graduate Committee — Signature

2. Dr.Mohammed Abdo _____

Advisor — Signature

3. _____

Examiner — Signature

4. _____

External Examiner — Signature

DECLARATION

I hereby declare that this thesis is my original work, has not been presented for a degree in this and any other college or university to the best of my best knowledge and that all sources of materials and researches used for it has been duly acknowledged.

Name : Abebe Demissie

Signature _____

Place : Addis Ababa

Date of submission:

This thesis has been submitted for examination with my approval as a university advisor.

Dr.Mohammed Abdo
Advisor

_____
Signature

ACKNOLOGEMENT

This thesis would not have been possible without the support and assistance of many people.

First and foremost, I would like to thank my advisor Dr.Mohammed Abdo, for his continued guidance and support throughout my studies. I thank him for all his wise advice and my discussions with him and this dissertation is an acknowledgement of his tenacity and his confidence in me. Thank you.

Finally, it is a great pleasure to thank my parents, my beloved wife and my friends for their continued help and support as they were always there to guide and encourage me.

Table of Content

LIST OF FIGURES

LIST OF TABLES

Abbreviation

SSID - (Service Set Identifier) is the name assigned to the wireless network
WEP - (Wired Equivalent Privacy) is an encryption method specified by the IEEE 802.11g standard to make any intercepted communications extremely difficult to interpret by unauthorized parties
NAT -Network Address Translator generally applied by a router that makes many different IP addresses on an internal network appear to the Internet as a single address.
IEEE - Institute of Electrical and Electronic Engineers
Gateway - A gateway links computers that use different data formats together.
Firewall - Firewall is considered the first line of defense in protecting private information.
ARP - Address Resolution Protocol. ARP is a protocol that resides at the TCP/IP Internet layer that delivers data on the same network by translating an IP address to a physical address.
802.11g - An IEEE standard for wireless local area networks. It offers transmissions speeds at up to 54 Mbps in the 2.4- GHz band.
FIPS -the Federal Information Processing Standard
E-911- universally available on traditional wire line and wireless phones so that the public has access to emergency services .It has an ability to provide caller identification and location information to the call answering center.
RFC- Request for Comments uses as address allocation for private Internet.
NIST -national institute of standards and technology
ARP-address resolution protocol, which is the standard method for finding a host's hardware, addresses when only its network layer address is known.
ISAKMP-Internet security association and key management protocol
MIKEY- multimedia Internet key
RAS-registration admission status
RSVP-resource reservation protocol
ANI- automatic number identification
POTS- plain old telephone service
PBX-private branch exchange
MIME-multipurpose Internet mail extension
IETF-Internet engineering task force
ITU-international telecommunication union

ABSTRACT

Perceived conversational speech quality is an important metrics in Voice over IP (VoIP) applications. As IP networks are not designed for real-time applications, the network impairments such as packet loss, jitter and delay have a severe impact on speech quality.

Since Security has become a major concern with the rapid growth of interest in the Internet, the thesis deals with the security aspects of VoIP systems and the performance analysis of secure VOIP networks. It stresses on the underlying of VoIP protocols like Session Initiation Protocol (SIP), H.323 and Media Gateway Control Protocol (MGCP). This thesis stresses more on issues regarding the firewall and the problem faced in using it's for VoIP; it further discusses the results obtained when security mechanism are applied.

This thesis develops an algorithm of an adaptive method codec selection mechanism which changes the voice encoding scheme in the middle of an active call based on the network conditions. The proposed mechanism involves establishing a three-way handshake process in mid-call to re-negotiate station capabilities, making the switch at a determined sequence number in a real time transport protocol (RTP) packet to solve the problems and then to increase the performance of VOIP.

Chapter One

1. Introduction

1.1. Motivation

In this age of information communication, Voice over Internet Protocol (VOIP) has taken the major portion of the interactions and it is dominating in modern daily life. It is more likely that the VOIP will dominate our daily life in the future much more than today. In contrast in developing economy VOIP is being used scarcely. Its access is limited to offices, educational institutions and Internet café (business centres where a user can browse Internet). Many causes for such situations can be listed as such, high cost of computers in low living standard, unavailability of infrastructure, and computer illiteracy. But there exist high demand for VOIP and other packet data services access with low cost devices and simple connectivity.

VOIP networks are facing a trend of exponential traffic increase and growing importance to users. In some countries, like Ethiopia by Ethiopian Telecommunications corporation (ETC), since still not officially introduced, but IP telephony service demand has grown rapidly and it is about to exceed the number users.

This increased demand can be attributed to:

- VOIP access eliminates the difficulty associated with fixed line payment.
- VOIP allows service anywhere there is Internet.
- IP telephone is easy to upgrade the infrastructure for the future.
- It supports all forms of communication standardization (it has an integrated infrastructure).
- Access terminal portability and its consequences such as it's being personal and cost effective.
- The new business model used to commercialize it.
- New and enhanced service provisions.

Besides the growth of the need for IP telephony service, VOIP networks are supposed to give data services such as Internet. But Internet is optimized to operate efficiently on packet-switched data network and hence the circuit-switched data service has limitations.

In order to meet these changing traffic patterns, more and more network operators are adapting their strategies and plan to migrate to IP-based backbone networks and wireless packet data accesses. This enable them serve more customers and, introduce new and enhanced services. The most important services include Internet browsing, Multimedia messaging, and streaming audio/video services and can be provided with low cost IP terminals as well as with long plan and deployment of new infrastructure.

The most widely used network standard is Public Switched Telephone Network (PSTN). Operators have a choice to implement either the add-on option called the Public Switched Telephone Network or the more advanced Voice over IP (VOIP) standards. Due to low resource requirement and less cost of VOIP licenses as well as allows service anywhere there is Internet and the general economy is favouring increased access to services, operators are focusing in resource optimization techniques to effectively utilize the deployed resources. Thus the most followed strategy was to evolve the network into VOIP system bringing the packet-switched carrier service to the existing PSTN system. VOIP provides an end-to-end connectionless packet service that includes packet mode transfers over an IP-based packet-switched core network.

Initially VOIP protocols are optimized for connectionless end-to-end packet data services, especially for Internet connectivity, while the existing PSTN network can still be used for circuit switched voice and data calls. Development of new services and requirements to provide those services over existing networks has lead to a continuous evolution of system features. VOIP has been designed having real-time services in mind, but the need for non-real-time services, has arisen. There have been a number of features that have improved the

capabilities of VOIP to support services with more severe delay requirements. Those features enabled operators to introduce service using streaming quality of service (QoS) class. Thus packet data services that can be provided by VOIP range from wireless Internet access to streaming audio/video services.

To get the maximum benefit from VOIP implementation, services designed to run on must be provided with the requested Quality of Service (QoS). Thus it is crucial to assess the capabilities of VOIP in terms of throughput, delay and bit error rate, and point out the features that are suitable for providing requested service parameters. The other motivation for VOIP study is it's being new, lesser experience of operation, and address various issues (automatic routing of incoming calls to the VoIP phone or using other Internet facilities for video conferencing, audio conferencing and file exchange in parallel to conversations).

1.2. Research problem

The most employed migration strategy from PSTN to Universal Mobile Telecommunication Systems (UMTS) is through VOIP. Many operators upgraded their PSTN network by introducing VOIP enhancements with the aim of improving efficiency and service diversity. To complement this aim, thesis attempts to address the problem of services performance over VOIP with delay sensitive service and investigate ways of improving service quality.

Since Voice over IP also known as Internet Telephony or IP Telephony is the flow of voice data over both the public Internet and the private intra-nets. It serves as a very cost efficient alternative to the traditional telephone networks. It provides additional facilities such as automatic routing of incoming calls to the VoIP phone or using other internet facilities for video conferencing, audio conferencing and file exchange in parallel to conversations. But there are a few drawbacks

that prevent its widespread deployment. IP neither provides any Quality of Service guarantees nor does it ensure that packets are delivered in sequential order. This leads to problems such as latency, jitter, packet loss, delay and voice encoding due to limited bandwidth. This thesis looks into the details of each these quality issues and their solutions.

1.3. Thesis objectives

VOIP currently delivers only 5.3–64kbps on average. During peak hours, the data traffic performance can further degrade, as voice traffic still takes priority. On top of that, applications and services have to challenge with latency in the network, which can impact synchronous applications like audio/video streaming.

The objective of this thesis is to study the Voice service, the possible and effective packet data services over Internet Protocol and indicate the performance with respect to the different parameters that affect Quality of Service (QOS). This thesis studies the introduction of packet delay (latency), packet loss, and jitter due to security mechanisms at network level, and content format and terminal at user level.

1.4. Thesis contribution

The thesis contribute

- Assist in careful planning & implementation of VOIP system for full utilization of its capacity. E.g. Planning, optimizations and enhancement of ETC's PSTN and VOIP can be done by us.
- Adds on the understanding of VOIP system, its operation, capacity and provision of local solution. It provides re-description of the system in a view so that it reflects the solutions to our problems.
- Enables to choose the effective services that generate income and satisfaction. This points out relevant services to our context such as browsing, messaging

and streaming audio services and proposes suitable implementation and operation.

- Points out challenges and initiates further studies about VOIP.

1.5. Thesis organization

This thesis is organized as follows

Section 1 contains motivation, research problem, objective and contribution.
Section 2 contains brief overview of the VoIP system
Section 3 contains the protocols used
Section 4 discusses the different security mechanisms, quality parameters and their impacts
Section 5 describes the different simulation metrics used for measuring voice quality over a network, and a currently approaches used to ensure quality in voice over IP.
Section 6 the conclusion and the recommendation

1.6 Literature Review on VOIP

1.6.1 Introduction

The multimedia content of streaming service has the capacity to deploy in VOIP system. The packet loss and delay introduced in VOIP system is due to human factors (individual perception of audio/video quality), device factors (operating system, firewall, processor and memory capacity), and network factors.

1.6. 2 Literature review

Like other real time applications, VoIP requires service guarantees which cannot be provided by the best-effort structure of today's networks. However different possibilities are being explored to utilize the existing networks in best possible ways to meet the quality of service guarantees of a VoIP system.
Service differentiation is one possible mechanism to ensure quality. This involves configuring the routers to dedicate resources and delivering service guarantees to VoIP traffic. The main drawback of this method is the difficulty in deploying such a mechanism over the entire Internet and the additional complexity that is imposed on the routers.
Vern Paxson examined this issue in his PhD thesis [1]. But he concluded that the linkage between delay variation and loss was weak, though not negligible. From the overall data that he presented, he concluded that many elements in the network, such as the individual

delay variations at each hop of the path, the buffer spaces at each router, would contribute to delay variation, but not to loss

Shu Tao et al. use Path Switching to improve Quality [2]. The Voice over IP traffic is configured to dynamically select a path from several paths based on their performance. An application driven path switching system is used which takes into account application specific parameters and network performance. At each end of communication it determines the path on which voice packets are to be forwarded. The frequency with which path switching is done is dynamically decided by taking into account the inherent path diversity of the Internet. The popular commercial VoIP application Skype introduced a similar idea in their peer to peer system. They maintain multiple connections for each session and at a given time choose the one that is best suited.

With VoIP, quality encompasses both call establishment parameters and voice quality (VQ). Call establishment parameters are the service availability and call-setup time. Voice Quality suffers due to IP network impairments. Using prioritization and reservation frameworks can minimize these impairments' impact on real time applications. However a common policy or a set of protocols need to use by the service providers to deploy reservation frameworks to improve quality. However there is little interaction between the service providers and hence they enforce their own policies in their respective networks. The lack of a standard implies that end-to-end QoS depends heavily on the strategies and control mechanisms implemented at the endpoints.

Michael Manousos et al. implemented end-point based architecture – named Call Quality Monitoring and Control (CQMC) [3]. The CQMC framework consist of an agent that collects call data forwards them to the core which then improves each call's voice quality (VQ) by adapting the system's operational parameters. CQMC describes necessary adaptation using a set of operation control commands that the core sends to the agent for execution.

Victor Frost [4] focuses on the temporal characteristics of congestion. In his paper he characterizes network congestion events and develops analytic methodology to predict expected number of congestion events per unit time. This metric could then be used for efficient routing thus controlling the quality of service for real time applications. High congestion events could be pushed on to the network elements at the boundary while the core network needs to be kept relatively free of them. The methodology proposed in this paper illustrates a mechanism for doing the same. Thus characterizing network congestion events and their rate of occurrence helps in improving Quality of Service in real time data transmission.

Velloso, Duarte and Rubinstein in their paper 'Analyzing voice transmission capacity on ad hoc networks' analyze the voice

transmission capacity on ad hoc networks [5]. They perform simulations on the mobile network and calculate delay and jitter. They also evaluate the QoS provision and mobility on the number of transmitting sources. As ad hoc networks have no infrastructure to support QoS, the increase in node complexity makes it even more difficult to transmit real time traffic. The also report that increase in network load causes a large reduction in voice transmission capacity in multihop ad hoc networks. They further advocate a distributed mechanism of admission control to reduce degradation in voice traffic capacity.

Anna Watson and Angela Sasse in their paper question the ITU-recommended methods for subjective analysis of quality of real time data such as speech and video [6]. According to the authors they are multiple factors that influence user's perception of quality. They establish critical quality boundaries i.e. minimum and maximum threshold for a particular dimension.

Cole and Rosenbluth of AT&T laboratories propose a method for monitoring VoIP application by using transport level, measurable quantities which are deduced from ITU-T's E-model [7]. The transport level quantities that were identified to be relevant in measuring quality are delay, packet loss, decoder de-jitter buffer packet loss. The method that they advocate is to first measure transport level parameters like delay, loss, variation etc. Then combine packet loss, delay variation, packet size de-jitter buffer operations and coder frame size into an error mask. Error mask consist of exact sequence of good and bad coder frames. Then combine the error mask with coder to get the E-model factor I.e. the E-model could then be used to estimate a quality score. The R factor is derived from these parameters. This R factor is directly related to Mean Opinion Score (MOS). Monitors used for measurement could be placed either within VOIP gateways or within the transport path. There exists trade-off between the two. There are a few drawbacks with their approach. The limitations of their work are:
No data that covers the wide range of coders, loss concealment algorithms and error masks giving rise to insufficient subjective information.
Te most important sets of error masks are yet to be characterized. It would require more data on Internet loss and delay variation behaviour.
No standard model for de-jitter buffer implementation exists.
With increasing demand for real time data for multimedia applications and telephony applications, there exists a need for greater research in this field to improve Quality of service provisions for such services.

Chapter Two

2. Introduction to VOIP

VoIP (Voice over Internet Protocol) is an IP network based voice transmission technology, instead of the traditional analogy telephone line; it allows people to make telephone calls through broadband Internet connections. At the beginning stage, VoIP existed as software. It was restricted to communication from PC to PC. In other words, just installing network telephone software on the PCs at each end, people can talk through to each other through the IP (Internet Protocol) network. With the development of network technology, network IP telephony grew from PC-PC to IP-PSTN (public switched telephone network), PSTN-IP, PSTN-PSTN and IP-IP, etc. The common characteristic is using the IP network as the transmission medium. Therefore, the tradition of using circuit-switched network as transmission medium is smashed gradually.

Compared to the circuit-switched telephony network; the major advantage of deploying VoIP is the lower cost, Extra requirements for VoIP to complete a VoIP call are limited, because it uses the existing network which satisfies most requirements of VoIP. Besides the low cost, there are many new features of VoIP available, such as communication with PSTN call. This is an obvious advantage, because from the end user's point of view, the location of the end user is not important any more. Whenever the user can access Internet, he can make a call to anywhere in the world. This is more useful, especially when the wireless local network is available. I.e. the benefits of implementing VoIP have become especially attractive. Some of the benefits are listed in the following:

Comparing with the traditional PSTN network, VoIP has lower cost and investment in terms of money and time than PSTN.

- Easier to integrate with other services and applications (audio, video, fax, data and multimedia information).
- By using cheaper voice delivery media – IP network, customer can save money.
- Making use of the existing network more efficiently.
- Extend the service to remote locations more economically.
- Create new service opportunities.

Generally, VoIP communication gives demonstrable benefits to end-users, and can be a long-term success. The benefits of using this technology can be divided into the following four categories?

1. Cost Reduction.
   - Reducing long distance telephone costs.
   - The sharing of equipment and operations costs across both data and voice users improves network efficiency.
2. Simplification.

   It is an integrated infrastructure that
   - Supports all forms of communication and allows more standardization.

- Reduces the total equipment balance.
- Supports dynamic bandwidth optimization and a fault tolerant design.

3. Consolidation.
   - Universal use of the IP protocols for all applications holds out the promise of both reduced complexity and more flexibility.
   - Related facilities such as directory services and security services are easily shared.
4. Advanced Applications.
   - The longer-term benefits of VoIP are expected to be derived from multimedia and multi-service applications

Besides all these benefits, VoIP also puts forward a new security and performance evaluation question: “Is the security sufficient for the needs in VoIP?” VoIP is easier to attack because of its special character. It needs extra security mechanisms in addition to the standard security methods for data networks.

Many readers who have a good understanding of the Internet and data communications technology may have little background in transmitting voice or real-time imaging in a packet-switched environment. One of the main sources of confusion for those new to VOIP is the (natural) assumption that because digitized voice travels in packets just like other data, existing network architectures and tools can be used without change for voice transmission. VOIP adds a number of complications to existing network technology, and these problems are compounded by security considerations. Most of this thesis is focused on how to overcome the complications introduced by security requirements for VOIP.
For several years, VOIP was a technology prospect, something on the horizon for the “future works” segment of telephony and networking papers. Now, however, telecommunications companies and other organizations have already, or are in the process of, moving their telephony infrastructure to their data networks. The VOIP solution provides a cheaper and clearer alternative to traditional PSTN phone lines. Although its implementation is widespread, the technology is still developing. It is growing rapidly throughout the world specially North America and Europe, but it is sometimes awkwardly implemented on most legacy networks, and often lacks compatibility and continuity with existing systems. Nevertheless, VOIP will capture a significant portion of the telephony market, given the fiscal savings and flexibility that it can provide [8].

2.1 VOIP Equipment

VoIP is one of the emerging technologies in telecommunications. “VoIP also called “packet telephony” technology translates analog voice signals into a stream of digitized packets and sends them over

data networks" .The VoIP network combines both voice and data communications technologies.

VoIP traffic can be classified into call signalling, call control and media communications. Call signalling is a process that is used to set up a connection in a telephone network. Call control decodes addressing information and routes telephone calls from one end point to the other. The communications used between the devices might use either one channel or many channels depending on the VoIP protocols. The channels used for the connection between devices are transport control protocol (TCP)/user datagram protocol (UDP).
Voice networks on PSTN are connection-oriented networks where the path from the source to destination is established before the information transfer. One of the advantages is that once the connection has been established, the sequence of information and delay should be constant. One of the disadvantages is the consumption of resources while "signalling".

In connectionless networks like data networks, source and destination addresses are attached to the packets and put in the networks for delivery to the destination. The packet might take any route upon availability and there is no guarantee that the packet arrives at the destination. The delay can vary to a long extent. This unreliable protocol is called UDP, which uses IP.
TCP/IP is not adequate for VoIP. New protocols should be added to support time sensitive applications like Voice and Video.

In VoIP, audio signals are sent over IP networks to another computer. The sound samples are recorded and compressed so that they require less space.
CODEC (Coder/Decoder) is a circuit that compresses audio signals. This reduces bandwidth considerably. "After compressing into small samples, these samples are collected together into larger chunks and placed into data packets for transmission over IP networks".

VoIP has its own components and parts and each component has a specific function. Figure 2.1 shows the components of VoIP network

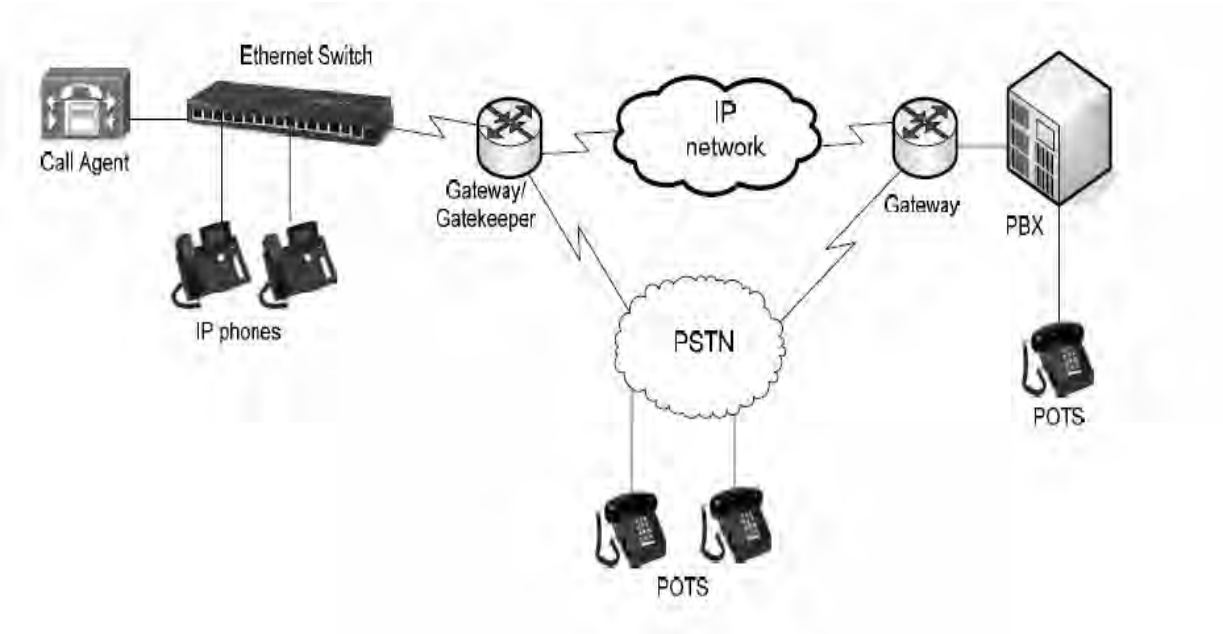


Fig2.1 VoIP Components

Call Agent: Call Agent performs call processing and manages the Gateways.
IP Phones: A telephone, which has built-in IP signalling protocols to support VoIP. It converts analog voice into IP packets and vice versa.
Gateway: It is a node that serves as an interface between two or more networks. Gateways can forward calls between different types of networks.
Gatekeeper: Gatekeeper provides bandwidth control.

In general, though, the term Voice over IP is associated with equipment that provides the ability to dial telephone numbers and communicate with parties on the other end of a connection who have either another VOIP system or a traditional analog telephone. Demand for VOIP services has resulted in a broad array of products, including:

1. Traditional telephone handset – Usually these products have extra features beyond a simple handset with dial pad. Many have a small LCD screen that may provide browsing, instant messaging, or a telephone directory, and which is also used in configuring the handset to gain access to enhanced features such as conference calls or call-park (automatic call-back when a dialled number is no longer busy). Some of these units may have a "base station" design that provides the same convenience as a conventional cordless phone.

2. Conferencing units – These provide the same type of service as conventional conference calling phone systems, but since communication is handled over the Internet, they may also allow users to coordinate data communication services, such as a whiteboard that displays on computer monitors at both ends.

3. Mobile units – Wireless VOIP units are becoming increasingly popular, especially since many organizations already have an installed base of 802.11 networking equipment. Wireless VOIP products may present additional challenges if certain security issues are not carefully addressed. The Wired Equivalent Privacy (WEP) security features of 802.11b provide little or no protection. The more recent WiFi Protected Access (WPA), a snapshot of the ongoing 802.11i standard, offers significant improvements in security, and can aid the integration of wireless technology with VOIP.

4. PC or "soft phone" – With a headset, software, and inexpensive connection service, any PC or workstation can be used as a VOIP unit, often referred to as a "soft phone". If practical, soft phone systems should not be used where security or privacy are a concern. Worms, viruses, and other malicious software are common on PCs connected to the Internet, and very difficult to defend against. Well-known vulnerabilities in web browsers make it possible for attackers to download malicious software without a user's knowledge, even if the user does nothing more than visit a compromised web site. Malicious software attached to email messages can also be installed without the user's knowledge, in some cases even if the user does not open the attachment. These vulnerabilities result in unacceptably high risks in the use of "soft phones", for most applications. In addition, because PCs are necessarily on the data network, using a soft phone system conflicts with the need to separate voice and data networks.
In addition to end-user equipment, VOIP systems include a large number of other components, including call processors (call managers), gateways, routers, firewalls, and protocols. Most of these components have counterparts used in data networks, but the performance demands of VOIP mean that ordinary network software and hardware must be supplemented with special VOIP components. The unique nature of VOIP services has a significant impact on security considerations for these networks, as will be detailed in later chapters [8].

2.2 Overview of VOIP Data Handling

Before any voice can be sent, a call must be placed. In an ordinary phone system, this process involves dialling the digits of the called number, which are then processed by the telephone company's system to ring the called number. With VOIP, the user must enter the dialled number, which can take the form of a number dialled on a telephone keypad or the selection of a Universal Resource Indicator (URI), but after that a complex series of packet exchanges must occur, based on a VOIP signalling protocol. The problem is that computer systems are addressed using their IP address, but the user enters an ordinary telephone number or URI to place the call. The telephone number or URI must be linked with an IP address to reach the called party, much as an alphabetic web address, such as "www.nist.gov" must be linked

to the IP address of the national institute of standards and technology (NIST) web server.
A number of protocols are involved in determining the IP address that corresponds to the called party's telephone number that a complex series of packet exchanges must occur, i.e. Voice over IP can be thought of as voice signals that are transmitted through the IP network. Thus, VoIP is a technology, used 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, which is digitalizing analog voice signal. Digitalizlization can be completed by various coding scheme. The current voice-coding standard is mainly ITU-T G.711, The source and destination must use the same coding algorithm, so that the digitalized bit stream can be reverted to understandable analog voice data. The Telephone Company, Internet Service Provider (ISP), or PC can do digitalization on the desk or the IP telephone set.

Step 2: Digital data to IP transformation

After digitalizing voice data into bit stream, the next step is compressing and coding the voice packet into specific frame, this is done by using complex algorithms. For example 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 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. In difference from circuit switching network, IP network doesn't have dedicated link between transmitter and receiver, the control header provides network navigation information for the packet, and the payload includes voice data, timestamp, and other additional information. Also, the PC on the desk or the IP telephone set can do by the telephone company, Internet Service Provider (ISP), or the reassemblation.

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 different values in 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, 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; bigl buffer generates small latency (the time it takes for a voice transmission to go from its source to its destination), but cannot adjust big jitter (non-uniform packet delay). Address information and other control information will be removed. Only the original data can be reserved. The reserved original data will be sent to the decoder, and the decoder will decode and decompress the voice data into new voice data.

Step 5: Digital voice to analog voice transformation

The media player driver samples the voice data and sends it to the sound card. The sound card plays it with pacific frequency. This process is depicted as follows:

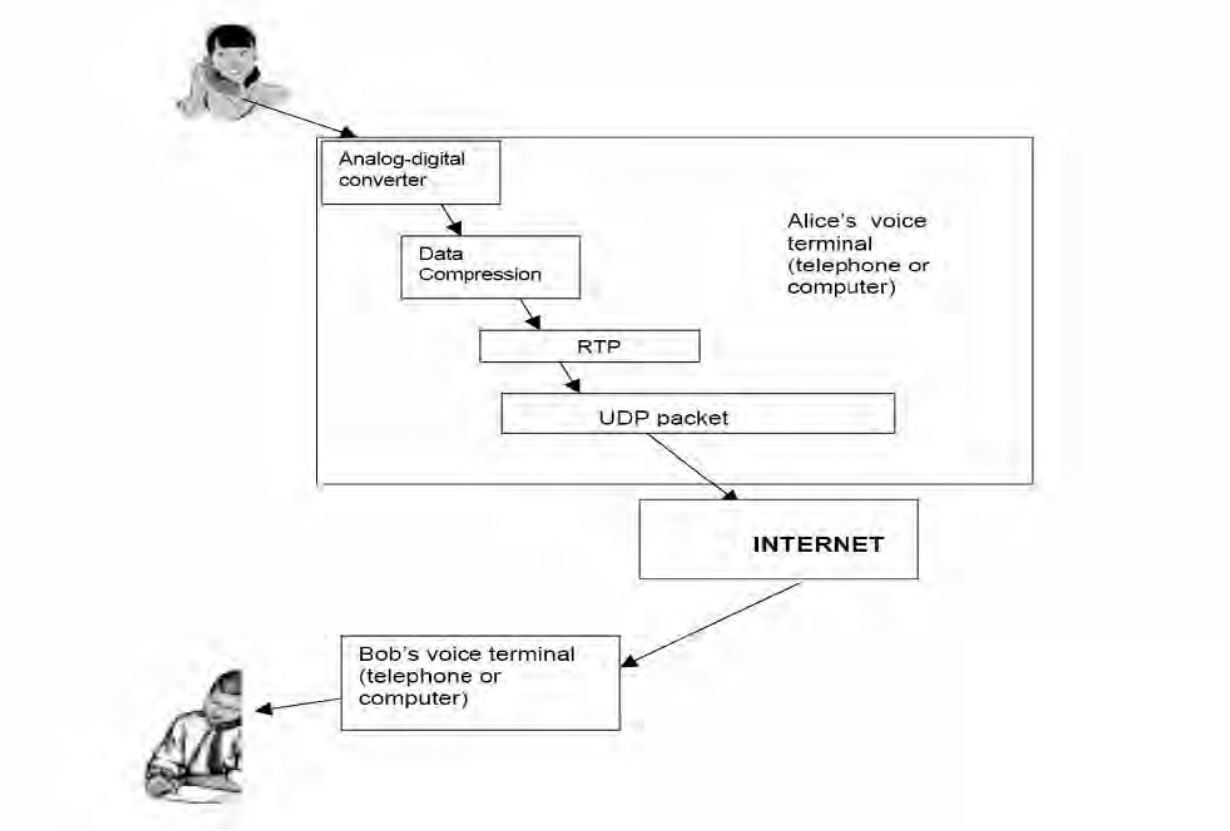


Fig.2-2.Voice data process flows between end points

Figure 2-2 illustrates the basic flow of voice data in a VOIP system. Once the called party answers, converting the voice into digitized form, and then segmenting the voice signal into a stream of packets must transmit voice. The first step in this process is converting analog voice signals to digital, using an analog-digital converter. Since digitized voice requires a large number of bits, a compression algorithm can be used to reduce the volume of data to be transmitted. Next, voice samples are inserted into data packets to be carried on the Internet. The protocol for the voice packets is typically the Real-time Transport Protocol, RTP (RFC 3550). RTP packets have special header fields that hold data needed to correctly re-assemble the packets into a voice signal on the other end. But voice packets will be carried as payload by user datagram protocols (UDP) that are also used for ordinary data transmission. In other words, the RTP packets are carried as data by the UDP datagram's, which can then be processed by ordinary network nodes throughout the Internet. At the other end, the process is reversed: the packets are disassembled and put into the proper order, digitized voice data extracted from the packets and uncompressed, then the digitized voice is processed by a digital-to-analog converter to render it into analog signals for the called party's handset speaker

2.3 Cost

The feature of VOIP that has attracted the most attention is its cost-saving potential. By moving away from the public switched telephone networks, long distance phone calls become very inexpensive. Instead of being processed across conventional commercial telecommunications line configurations, voice traffic travels on the Internet or over private data network lines.

VOIP is also cost effective because all of an organization's electronic traffic (phone and data) is condensed onto one physical network, bypassing the need for separate PBX tie lines. Although there is a significant initial start-up cost to such an enterprise, significant net savings can result from managing only one network and not needing to sustain a legacy telephony system in an increasingly digital/data centred world. Also, the network administrator's burden may be lessened as they can now focus on a single network. There is no longer a need for several teams to manage a data network and another to mange a voice network. The simplicity of VOIP systems is attractive, one organization / one network; but as we shall see, the integration of security measures into this architecture is very complex.

2.4 Speed and Quality

In theory, VOIP can provide reduced bandwidth use and quality superior to its predecessor, the conventional PSTN. That is, the use of high bandwidth media common to data communications, combined with the high quality of digitized voice, make VOIP a flexible alternative for speech transmission. In practice, however, the situation is more complicated. Routing all of an organization's traffic over a single network causes congestion and sending this traffic over the Internet can cause a significant delay in the delivery of speech. Also, bandwidth usage is related to digitization of voice by codec's, circuits or software processes that code and decodes data for transmission. That is, producing greater bandwidth savings may slow down encoding and transmission processes. Speed and voice quality improvements are being made as VOIP networks and phones are deployed in greater numbers, and many organizations that have recently switched to a VOIP scheme have noticed no significant degradation in speed or quality.

2.5 VOIP Security Issues

With the introduction of VOIP, the need for security is compounded because now we must protect two invaluable assets, our data and our voice. Ethiopian Federal 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. In a conventional office telephone system, security is a more valid assumption. Intercepting conversations requires physical access to telephone lines or compromise of the office private branch exchange (PBX). Only particularly security-sensitive organizations bother to encrypt voice traffic over traditional telephone lines. The same cannot be said for Internet-based connections. For example, when ordering merchandise over the phone, most people will read their credit card number to the person on the other end. The numbers are transmitted without encryption to the seller. In contrast, the risk of sending unencrypted data across the Internet is more significant. Packets sent from a user's home computer to an online retailer may pass through 15-20 systems that are not under the control of the user's ISP (Internet service provider) or the retailer. Because digits are transmitted using a standard for transmitting digits out of band as special messages, anyone with access to these systems could install software that scans packets for credit card information. For this reason, online retailers use encryption software to protect a user's information and credit card number. So it stands to reason that if we are to transmit voice over the Internet Protocol, and specifically across the Internet, similar security measures must be applied.

The current Internet architecture does not provide the same physical wire security as the phone lines. The key to securing VOIP is to use the security mechanisms like those deployed in data networks (firewalls, NAT, encryption, etc.) to emulate the security level currently enjoyed by PSTN network users. These investigate the attacks and defences relevant to VOIP and explore ways to provide appropriate levels of security for VOIP networks. The threats aims at network is complicated, however, VoIP put forward new vulnerabilities, from the end user device, such as IP telephone set, soft phone, wireless communications, to signalling equipments, such as H.323 gatekeeper, SIP proxy server. The attacks include simple DoS (denies of services) attack which aims at destroying availability of some resources or services, dishonest identity, such as making toll call, etc. Security issue of VoIP is complicated, although there are some security mechanisms, but there are still other issues, which cannot be easily solved by traditional security methods.

In different steps of VoIP, there are different attack threats, these are:

1. After voice data is digitalized and compressed, it will be sent to the network either by cable, or by wireless access to a wired network. Wireless access introduces new vulnerabilities to security of VoIP. The major protocol standard in wireless network is IEEE 802.11b. IEEE 802.11b has some security mechanisms, such as encryption with WEP (Wired Equivalent Privacy), using SSID (service set identifier) to control the access, using key in authentication, etc. Although 802.11b has some security mechanisms, there are vulnerabilities with each mechanism, such as 802.11b using WEP to encrypt transferred message. There are many security problems with WEP. Some tools (such as Air Snort, WEP Crack) are available for hacker to be able to crack WEP keys by analyzing the traffic. The Problem with SSID is that users usually don't change the default SSID. This gives opportunities to attacker. Wireless access points can lose signals since it broadcasts over air, key exchange is also not secured.

2. After the voice data is sent to network, voice data has the same attack threats as other data packet on the network. Currently, there are several methods to secure the network security, such as by installing firewall to control access to the network, using access control list to control source of the packet, using NAT to hide intranet from un-trusted network, using encryption to protect data integrity, etc. These methods perform protection of the network to some extent, but because of the special characters of VoIP, these methods are not strong enough to secure the VoIP based network. Since the signalling protocols in VoIP use dynamic ports, such as H.323, packet-filtering firewall is not a good solution, since it needs to open and close ports dynamically. Moreover, since H.323 uses embedded IP address, it can not be re-written by NAT.

3. The voice packet will be sent to the VoIP system server after they come out of the network. VoIP system server is the key component in VoIP, and also a weak point in VoIP system. Since most VoIP systems are designed on open platform, such as UNIX, Windows server, etc, firewalls are used to protect server, as described above. Since VoIP needs to dynamically open ports, this enhance the complexity of firewall; besides, if there are some bugs with the underlying operating system, the operating system itself is easier to be attacked also. Although there is antivirus software and system patches can be used to protect the underlying operating system. But the VoIP system still shares same risks with it.

4. In VoIP, there is some key equipment, which takes important responsibilities in VoIP operation, such as gateway, gatekeeper, server, IP phone set. In most cases, they are not physically protected, this is an obvious opportunity for an attacker to make malicious attack, such as by seeing a user login to the server, remember password of the users, then use this user's password to login and make toll call [18]

Chapter Three

3. VOIP Protocols

There are a number of protocols in providing 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 as the follows:

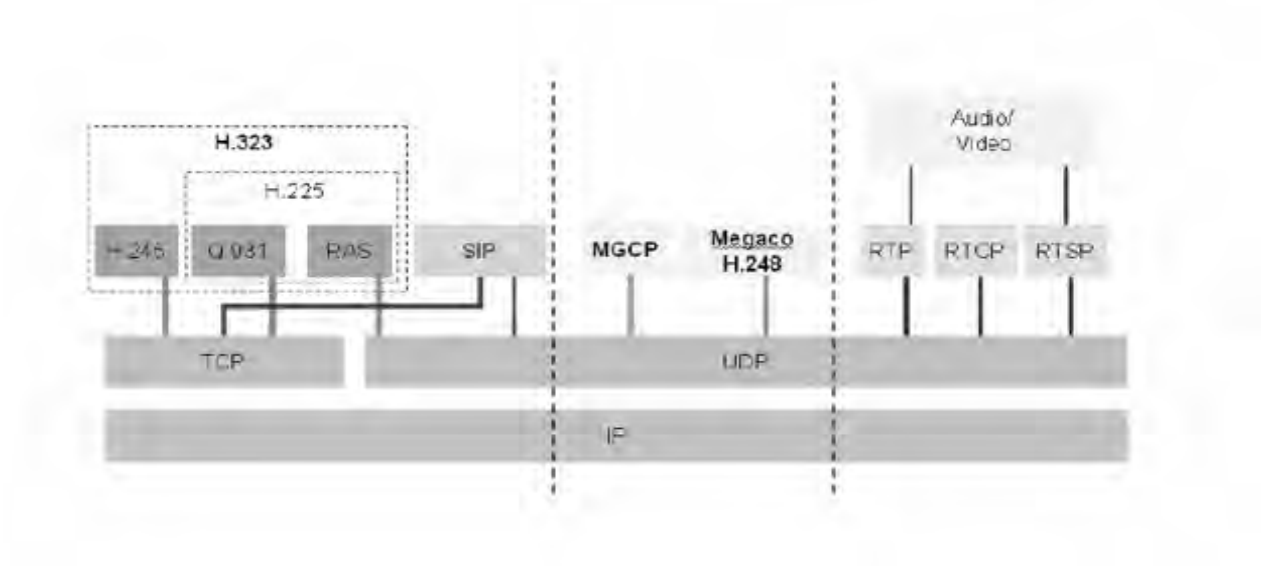


Fig 3.1. VoIP signalling protocols

3.1 RTP

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).

- 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.

- 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.

RTP doesn't include issue of resource reservation; it relies on the resource reservation protocols, such as Reservation Protocol (RSVP).

The RTP data structure is shown below:

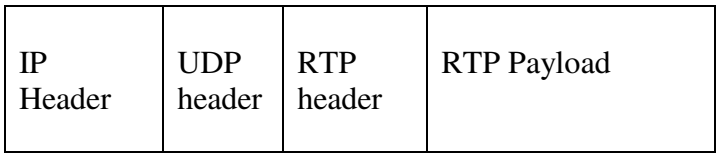


Fig 3.2. RTP data structure

The real time media data is in the RTP payload. RTP Header contains information of the payload, such as the source address, size, encoding type, etc. From figure 2-2, we can see that RTP works on top of UDP. To transfer RTP packet on 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. Secure RTP provides encryption, authentication and integrity check of the multimedia stream. In the multimedia conference, the RTCP protocol is used to transfer the control message to all participants periodically. It provides the following function:

- Provides feed back of the data transmission.
- It carries identifier to identify where the RTP data come from. RTCP may also include other information, such as email.
- Since each participant send control packets to others, so there is a number, which indicates how many users, there are. This number is used to calculate the packet transmission speed. More people in the session mean each people send packets in less frequency.
- RTCP packet includes sender and receiver's identifiers, statistics of the network traffic such as jitter, delay, and packet loses, etc.

3.2 H.323

H.323 is a standard published by the International Telecommunications Union Telecom Sector (ITU-T) for audio and video communication across packetized networks. H.323 gateway protocol is an umbrella-like standard that encompasses many sub-protocols within it such as H.225, H.235, H.245 and others, each performing a specific action in the process. It is a collection of protocols that perform functions such as setting up and tearing down VoIP calls, channelling for messages, commands, and encoding voice conversations etc.

3.3 H.323 Architecture

An H.323 network is made up of several endpoints (terminals), a gateway, and possibly a gatekeeper, Multipoint control unit, and Back End Service. The gatekeeper is often one of the main components in H.323 systems. It provides address resolution and bandwidth control. The gateway serves as a bridge between the H.323 network and the outside world of (possibly) non-H.323 devices. This includes session initiation protocol (SIP) networks and traditional PSTN networks. This brokering can add to delays in VOIP, and hence there has been a movement towards the consolidation of at least the two major VOIP protocols. A Multipoint Control Unit is an optional element that facilitates multipoint conferencing and other communications between more than two endpoints. Gatekeepers are an optional but widely used component of a VOIP network. If a gatekeeper is present, a Back End Service (BES) may exist to maintain data about endpoints, including their permissions, services, and configuration.

Before multimedia data can flow from a device to another device, various protocols are used to define how to transfer the stream. The protocols aimed at this functionality are called call-signalling protocol. The two major protocol standards for VoIP signalling are: H.323 protocol (ITU) and Session Initiation Protocol (SIP) (IETF). Both protocol standards define how VoIP technology works. However, each standard uses different methods for call signalling and call control. More importantly, they are not interoperable.

From the figure, one can easily see that VoIP signalling 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, and 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

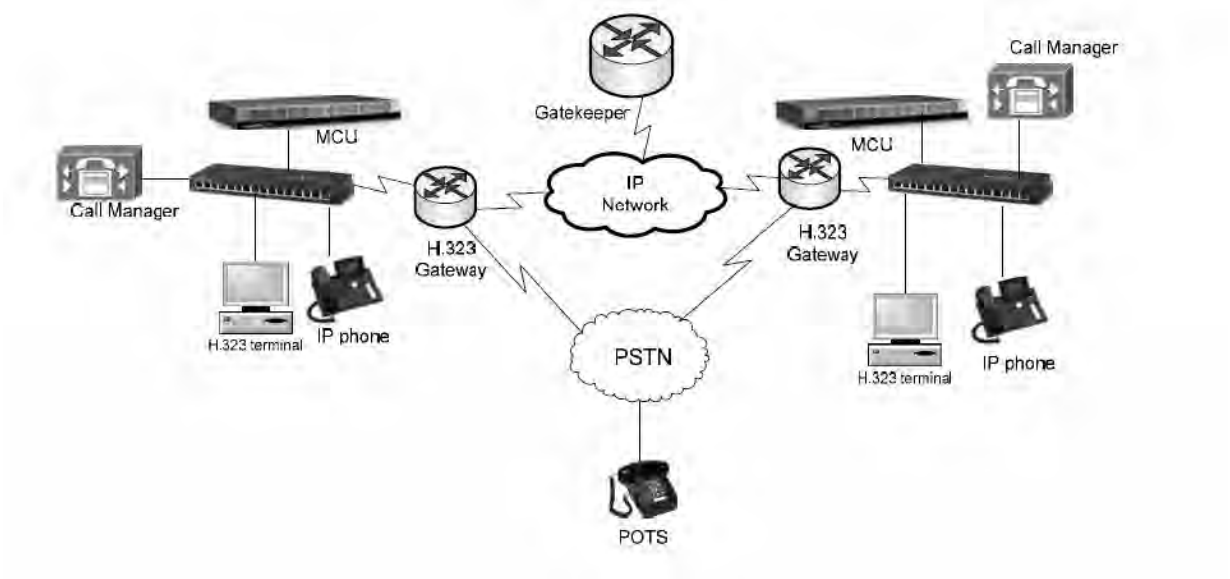


Figure 3.3. H.323 components

H.323 is a protocol, which applies intelligence everywhere. The H.323 network contains elements like H.323 Gateway, H.323 Gatekeeper, and H.323 Multipoint control unit (MCU) and H323 terminal.

H.323 Gateway acts like a bridge and serves as a communication point between H.323 and non-H.323 networks. Any traffic that comes from a PSTN network or that supports SIP or any other protocol routes through the H.323 Gateway. The Gateway routes calls from an H.323 terminal to an outside network or from an outside network to an H.323 network.

The Gatekeeper is an optional but important entity. It provides bandwidth control and serves as a "traffic cop" for the IP network. The Gateway or a call manager requests permission from the Gatekeeper before processing a call. The Gatekeeper grants permission to the gateway or a call manager based on the bandwidth availability in the IP WAN network. The Gatekeeper could deny a call request if there is no sufficient bandwidth to support a call. The Gatekeeper performs address resolution; H.323 alias numbers, E.164 addresses (phone numbers) or URLs should be resolved to IP addresses in order to route the call through the IP network [14]. H.225.0 Annex G protocol that resides in a Gateway or a Gatekeeper divides the network into domains for resolving alias addresses.

Multipoint Control Unit manages the signalling to add or remove participants in a conference. Multipoint Control Unit contains Multipoint controller and Multipoint processor for handling call control and media exchange in a conference.
H.323 terminal is an end point on a network; it communicates with either another H.323 end point or Gateway or MCU. H.323 terminals

have the entire set of H.323 features. The figure below shows all the components of H.323.
There are different ways that calls are completed in the H.323 network, these are:

- H.323.0-Q.931 Gateway to Gateway calls.
- H.323.0-RAS calls through a Gatekeeper.

3.4 H.323 Gateway to Gateway

The Gateway-to-Gateway calls do not require a Gatekeeper. Instead they communicate directly to each other. H.225.0-Q.931 performs the initial call set up by either establishing a TCP connection. It also manages the addressing information for the H.245 protocol. The H.245 protocol does the call control mechanism and is responsible for establishing the channels where the actual media transfer occurs. Another fast method may often be used where the call setup and control messages exchange in a single exchange of messages between the Gateways. Figure 3 shows how a call is setup in a Gateway-to-Gateway model by the exchange of H.323.0-Q.931 messages that are necessary for call setup and termination.

. Endpoint 1 obtains the address of the Endpoint 2 Gateway from the URL. It then fetches the Endpoint 2 IP address 192.168.0.101 in this case and the TCP port number 1720 from the Endpoint 2 Gateway. Port 1720 is usually used for call signalling for TCP.

Endpoint 1 opens a TCP connection with Endpoint 2 for call signalling. It sends a call setup message to the Endpoint 2 Gateway. The setup message has the source's IP address 129.168.0.3 and port numbers 1070 for call signalling (H.225.0-Q.931) and 1080 for call control (H.245) messages.

The Endpoint 2 Gateway upon receiving the setup message alerts Endpoint 2 and sends an alert message to Endpoint 1. The Endpoint 1 Gateway stops alerting Endpoint 1 if the user takes the call.

The connect message contains the Endpoint 2 IP address 192.168.0.101 and the TCP port 2060 for H.245 messages.

Once the H.245 messages have been exchanged and the connection has been established, RTP and RTCP streams flow through them.

Once the call is completed, either endpoint sends the release complete

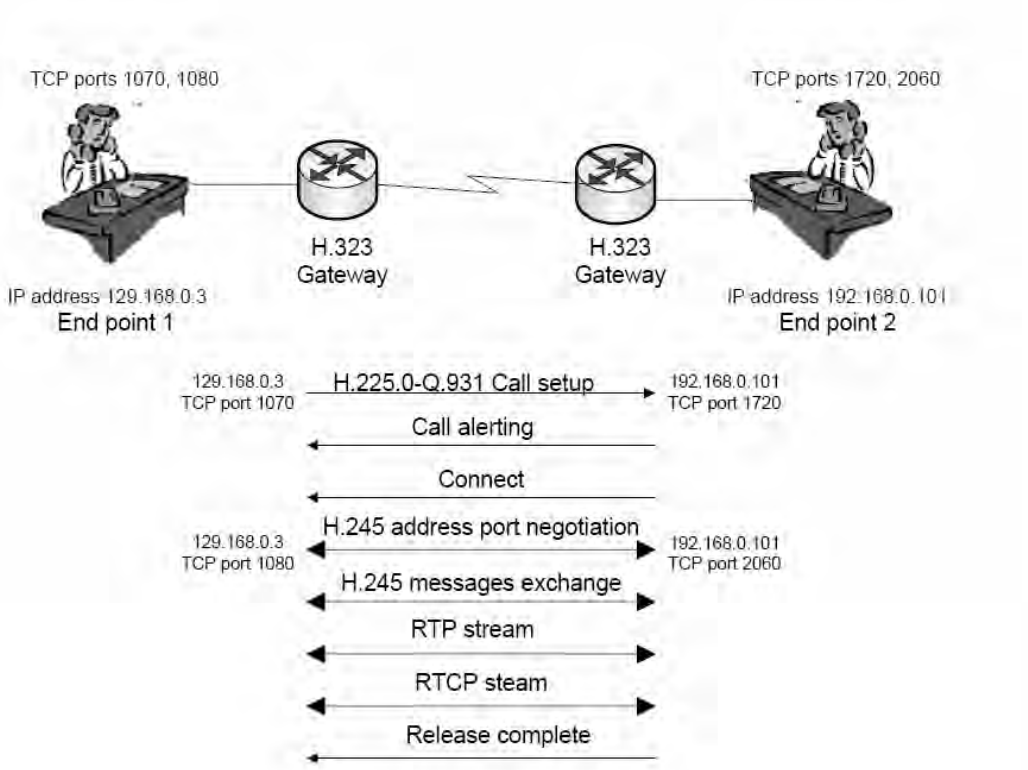


Figure 3.4 Call setup from Gateway to Gateway

3.5 H.323.0-RAS calls through a Gatekeeper

For this topology, the H.225 defines RAS (Registration Admission Status) protocol for communicating with the Gatekeeper. Gatekeepers provide address translation and bandwidth management. Endpoints send call signalling messages directly to the peer endpoints. Gatekeeper has an optional feature of call signal routing in which end points send call signalling to the Gatekeeper and Gatekeeper sends it to the destination end point.
The figure 2.5 shows how a call is setup through a Gatekeeper.

Before calling endpoint 2, Endpoint 1 discovers its Gatekeeper by sending a Gatekeeper request (GRQ) message. There are two kinds of Gatekeeper discovery: manual and automatic. In the manual discovery the transport address (IP address and port number) is preconfigured in the endpoint. Endpoint sends the GRQ message at that preconfigured transport address. In the automatic discovery, if Domain Name Server is used, endpoint initiates a DNS resource record query for the transport address using the Gatekeeper's domain. UDP ports 1718 and 1719 are generally used.

Once Endpoint 1 discovers its Gatekeeper, the gatekeeper either responds with Gatekeeper Confirm (GCF), if it can serve the endpoint or Gatekeeper Reject (GRJ) if it cannot.

If the Gatekeeper responds with a (GCF) message, Endpoint 1 registers with the Gatekeeper by sending Registration Request (RRQ), Gatekeeper responds by either Registration Confirm if it registers the user or Registration Reject (RRJ) message if it doesn't.

. Once it is registered, the originating endpoint requests permission from the Gatekeeper to place a call using Admission Request (ARQ). The Gatekeeper responds to it either by sending an Admission Confirm (ACF) or an Admission Reject (ARJ) basing on the bandwidth availability.

. The Gatekeeper does the address translation by translating the alias or E.164 addresses to transport addresses. The destination endpoint also sends a request (ARQ) to the Gatekeeper. If the Gatekeeper grants the permission for both the originating and destination Gateways, call setup proceeds. Figure 2.5 below shows how calls are completed for H.323 calls using a Gatekeeper

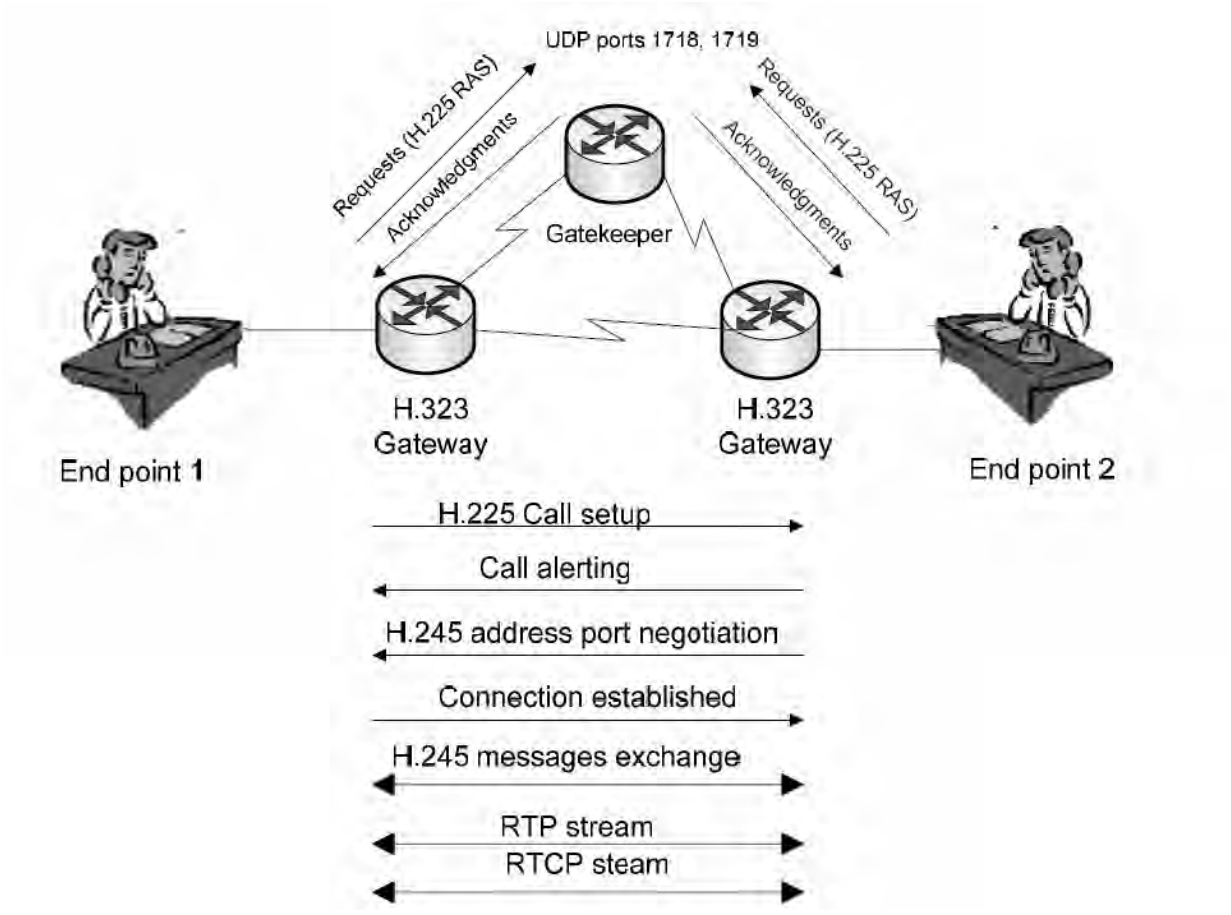


Figure 3.5 Call setup through a Gatekeeper

3.6 SIP (Session Initiation Protocol)

SIP (Session Initiation Protocol) is an application layer-signalling protocol developed by the Internet Engineering Task Force (IETF) to set up, modify and terminate multimedia sessions across packet networks. SIP is text based and designed to be extensible. It can be extended to accommodate features like instant messaging; video and services like call control services, interoperability and more. SIP can be carried by TCP, UDP or various other IPs. SIP is used to identify, locate and join entities who wish to communicate using peer-to-peer media type.

3.6.1 SIP Components

The SIP network is composed of the following types of entities each having specific functions to perform:
User agent: User agent is a SIP endpoint entity. User agents are an interface between the user and the SIP environment. They initiate and terminate sessions by exchanging responses and requests. A user agent can be a client or a server.

- User Agent Client (UAC): an application that initiates and sends SIP requests.
- User Agent Server (UAS): an application that receives SIP requests.

SIP devices can communicate directly if the endpoints know each other's URI (Uniform Resource Identifier) or IP address, but SIP servers are used for providing a means for routing, registration and authentication services.
Registrar Server: The Registrar Server authenticates and registers users when they come online. It updates the location database with the contact details and stores them.
Proxy Server: "A Proxy Server is an entity that acts both as a server and a client". A proxy takes the requests and interprets them; it can modify a SIP request, if necessary, before forwarding it. A proxy is involved only in setup and termination sessions.
Redirect Server: A Redirect Server obtains the actual address from the location server, takes a SIP request and maps the SIP address of the destined user to the address of the device closest to the user and returns it to the client.
Location Server: A Location Server is a database that keeps track of users and their locations. The location server gets its information from the Registrar and provides address resolution for Proxy and Redirect Servers [16].

3.6.2 SIP operation

3.6.2.1 SIP call without Proxy Server

The setup process shown in fig 2.9 describes the basic SIP setup between two user agents. It is assumed that each user agent knows the other's URI or IP address (sip: 585-424-8633@company.com; user phone is an example of an URI)

User Agent Client (UAC) originates the call. The destination server (UAS) responds if it wishes to join the session.

UAC begins the message exchange by sending an invite message to the called user, known here as User Agent Server (UAS). The invite message contains Session Description Protocol (SDP) that defines details of the type of session that is requested.
The fields listed in the invite message are called headers. There are different headers like via header that has the sending SIP device's address, SIP version and the default port number, which is 5060. The next headers To and From indicate the originator and destination of the SIP request. "The other headers like Content-Type and Content-Length indicate that the message body is SDP which has the connection IP address, media format, port number, media transport protocol, media encoding and sampling rate".
Invite message is a SIP request message; the next message is the ringing message from the called party in response to the invite message. The Ringing message signifies that the called party received the invite message and is ringing.
When the destination party answers the call, an ok message is sent to the calling party. This also indicates that the caller's type of media session is acceptable by the called party. The ok message body contains the called party's media information and a SDP message body that is added to the response.

The called party sends an acknowledgment request to confirm the media request.
The media session is established and the media session takes place-using RTP (Real Time Transport Protocol). The figure below shows how communication takes place without the use of Proxy or Redirect Servers [15].

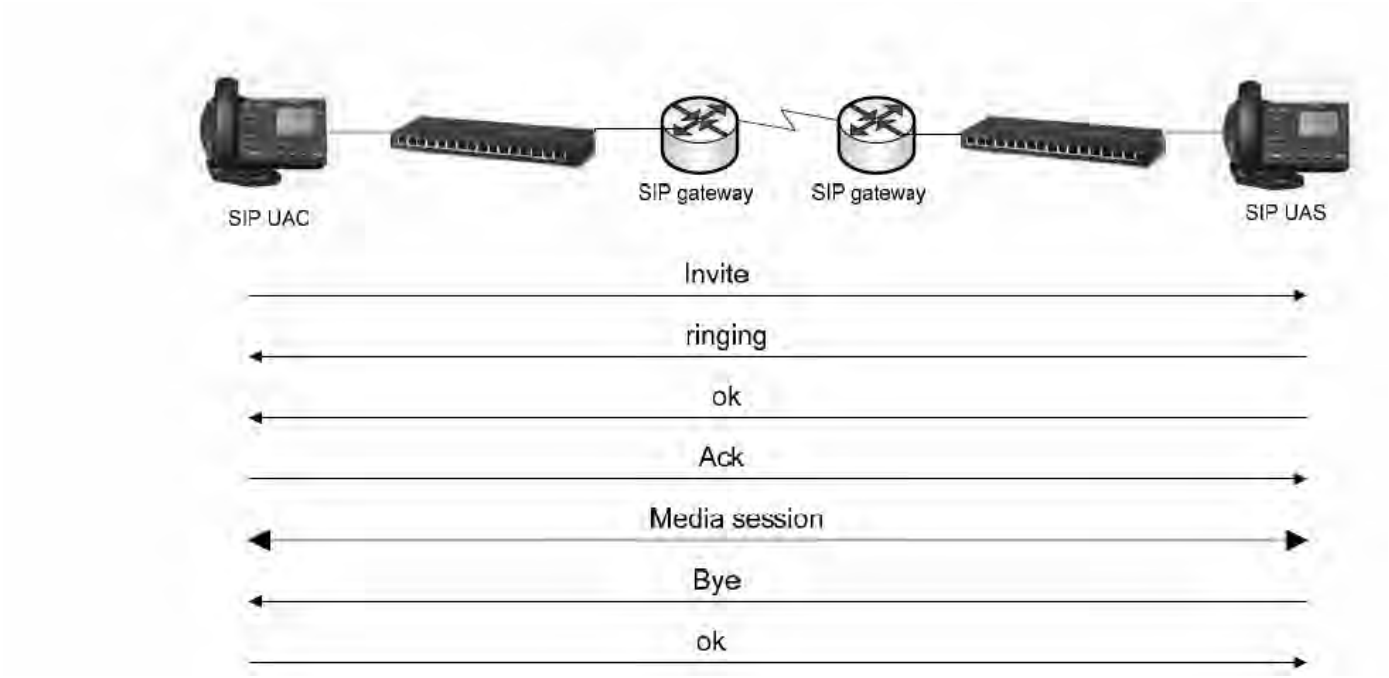


Figure 3. 6 Call setup without a Proxy Server

3.6.2.2 SIP call with Proxy Server

This section explains the typical SIP call with a SIP proxy server. "A SIP proxy does not setup or terminate SIP sessions, but it resides in the middle of a SIP message exchange, receiving messages and forwarding them".
In this mechanism the calling user agent has no idea of where the called party is; a SIP proxy server routes the Invite message to the destination. A DNS lookup of the called party's URI domain name is performed; it returns the IP address of the proxy server that handles the domain.
"The Invite message is then sent to that proxy's IP address. The proxy then looks up the URI in the request in its database and locates the called party".
The Invite message is then forwarded to the called party's IP address with the addition of a second via header attached with the address of the proxy.
The Session Description Protocol (SDP) in the header field has all the details about the type of the session. The Ringing response is sent back by the called party to the proxy. The response contains the via headers and the To and From, call-ID headers from the Invite request.

The Proxy receives the response, checks that the first via header has its own address, removes that via header and forwards the response to the address in the next via header.
"If the call is accepted by the called party, an ok message is sent to the proxy. The proxy forwards the ok message to the called party after removing the first via header".

The presence of contact header with the URI address of the calling party in the ok message allows the called party to send the acknowledgement directly to the destination without going through the proxy. This request and all further requests continue to use the contact tag in the header.
The media is always end to end and not through the proxy server.

The Bye message from the called party ends the Media Session. The calling party confirms it by sending an ok response. The figure below shows how communication takes place between user agents with a proxy server in between them.

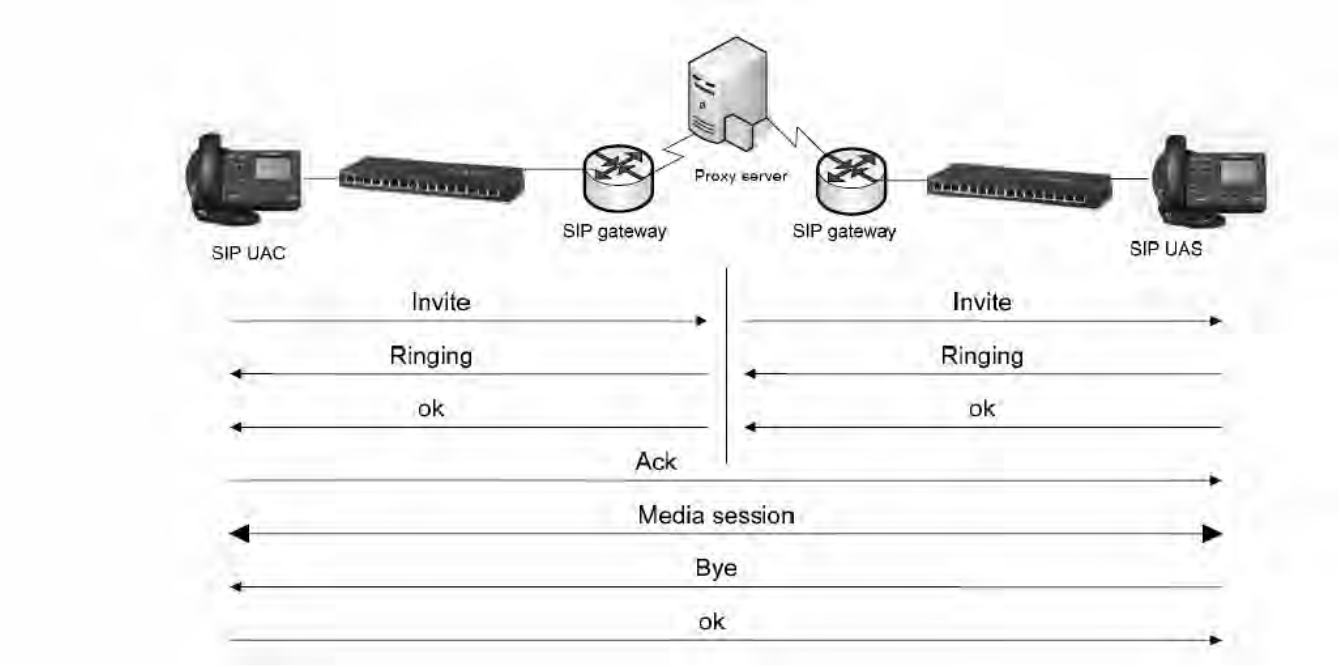


Figure 3. 7Call setup with a Proxy Server

SIP is a session layer protocol, it has two basic functions: signalling and session control.
Signalling is used to translate signals between different networks.
Session control is used to control the attributes of the end-to-end call.

SIP has the ability to:

- Provide address resolution, name mapping and call redirection. It can also 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.
- Establish session between two nodes if the call can be completed.
- Provide transferring calls. A call can be transferred from one target node to another target node without terminating the call. The session between the origination and the old target node will

be terminated; a new session will be set up between origination and the new target node

3.6.3 Voice enabled IPSec VPNs

"IPSec can be used to encrypt the traffic between two hosts known as transport mode or to build virtual tunnels between two networks which could be used to provide secure communication across the network which is known as tunnel mode or virtual private network (VPN)". The traffic flowing across the shared network can be secured as if it were flowing on a private network. A VPN tunnel is created as a logical point-to-point secured connection over a VoIP network. IPSec is widely used to secure the VPN link. IPSec provides a mechanism for the hosts to agree on an encryption key and a virtual network link needs to be created before implementing IPSec. The communicating Gateways need to have public IP addresses and private IP addresses. The Gateway needs to know how to reach the IP address of other Gateway. The packets routed from one Gateway to another Gateway should appear as if they are from a public IP address and they have to be sent to a public IP address although they route between private IP addresses. Each packet is wrapped up within another packet so that they appear to be are routing between public IP addresses. This process is called encapsulation. Once the packet reaches the destination public IP address, it needs to be un-encapsulated and delivered to the private IP address.
The following procedure provides an overview of implementing a VPN using IPSec:

- Generic routing encapsulation (GRE) tunnel endpoints are created.
- The hosts need to agree on the encryption mechanism to use.

- Security and authentication policies are established. The mechanism for specifying which traffic needs to be encrypted is defined. Crypto maps are defined.
- Crypto maps are associated with GRE tunnels. The encrypted voice traffic is routed through one of the tunnels to the destination
- The voice traffic on gateways is identified and classified.

The voice traffic across the tunnel is assigned a higher priority using appropriate queuing methods. The following figure shows a VPN using IPSec tunnel:

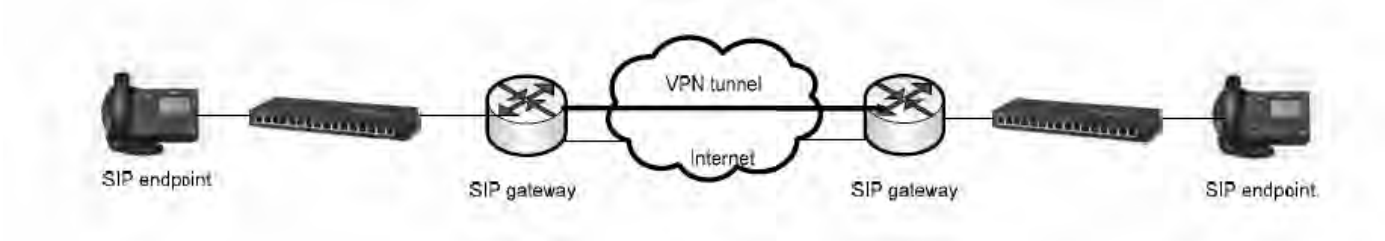


Figure 3.8 VPN architecture in SIP

3.7 Gateway Decomposition

Media gateway control protocols address the requirements of IP telephony networks that are built using "decomposed" VOIP gateways. Decomposed VOIP gateways consist of Media Gateways (MGs) and Media Gateway Controllers (MGC), and appear to the outside as a single VOIP gateway. MGC handles the signalling data between the MGs and other network components such as H.323 gatekeepers or SIP Servers, or towards SS7 Signalling Gateways. MGs focus on the audio signal translation function, performing conversion between the audio signals carried on telephone circuits and data packets carried over the Internet or other packet networks. A single MGC can control multiple MGs, which leads to cost reductions when deploying larger systems. Common examples are the Media Gateway Control Protocol (MGCP) and Megaco/H.248.

3.7.1 Media Gateway Control Protocol (MGCP)

3.7.1.1 Overview

MGCP (Media Gateway Control Protocol) is a complementary protocol to H.323 and SIP". "MGCP is a protocol that is used for controlling gateways from external call control units called Media Gateway Controllers (MGCs) or call agents". MGCP gateways do not have call forwarding intelligence. Call control intelligence is outside the gateways and handled by external call control agents. The call agents synchronize with each other to send commands to the gateways under their control; they manage the calls and conferences and support the services provided. "MGCP is a master/slave protocol with a tight coupling between the Gateway and call agent/server"

3.7.1 .2 System Architecture

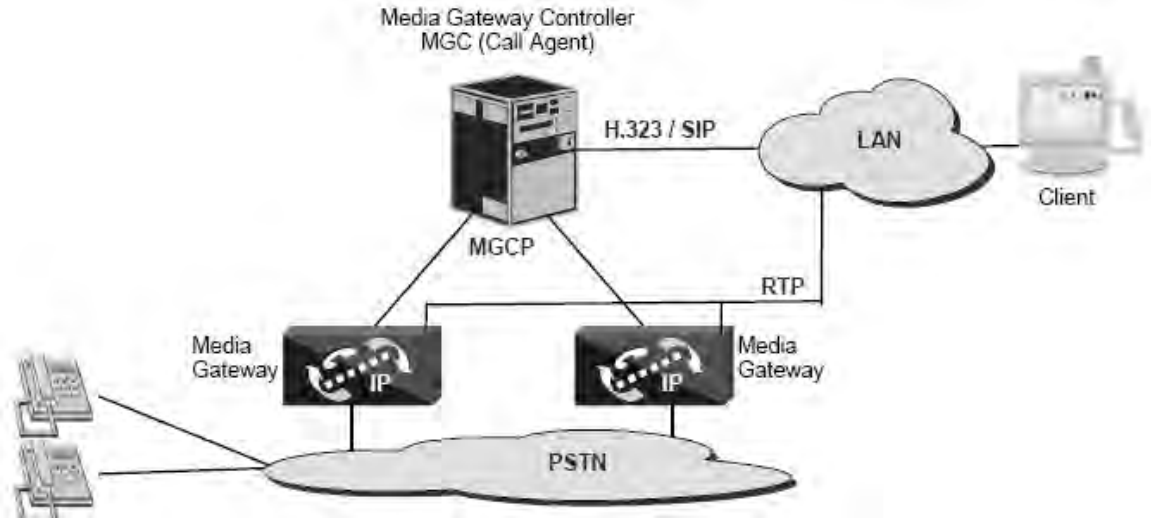


Fig 3.9 General Scenario for MGCP Usage

MGCP components
The components of an MGCP network include:

Endpoints: An endpoint is an interface between the VoIP network and the traditional telephony network. Endpoints are sources and sinks of data and could be physical or virtual.

Gateways: MGCP categorizes different types of gateways, which convert audio between different types of networks. The Gateways follow the commands issued by their call agents that control them.

Call agents: MGCP call agents have the intelligence of MGCP networks; they control the gateways and their endpoints. Call agents handle signalling and call processing functions; they take the responsibility for setting up calls and establishing the rules for communication and then back out once the calls are established.
After the call has been established, the call agent backs out and the RTP (Real time Transport Protocol) data is exchanged directly between the communicating gateways. The use of MGCP by the call agent provides the description of connection parameters like IP addresses, UDP ports and RTP profiles to the gateways [18].

3.8 VOIP Security Quality of Service Issues

Quality of Service (QoS) is fundamental to the operation of a VOIP network. Despite all the money VOIP can save users and the network elegance it will provide little added value , if it cannot deliver at least the same quality of call setup and voice relay functionality and voice quality as a traditional telephone network,. The implementation of

various security measures can degrade QoS. These complications range from delaying or blocking of call setups by firewalls to encryption-produced latency and delay variation (jitter). QoS issues are central to VOIP security. If QoS was assured, then most of the same security measures currently implemented in today's data networks could be used in VOIP networks. But because of the time-critical nature of VOIP, and its low tolerance for disruption and packet loss, many security measures implemented in traditional data networks just aren't applicable to VOIP in their current form. The main QoS issues associated with VOIP that affects security are presented here [23]:

3.8.1 Latency

Latency in VOIP refers to a processing time. Ideally, we would like to keep latency as low as possible but there are practical lower bounds on the delay of VOIP. The ITU-T Recommendation G.114 establishes a number of time constraints on one-way latency. The upper bound is150 ms for one-way traffic.

For international calls, a delay of up to 400 ms was deemed tolerable, but since most of the added time is spent routing and moving the data over long distances.

VOIP calls must achieve the 150 ms bound to successfully emulate the QoS that today's phones provide.

Delay is not confined to the endpoints of the system. Each hop along the network introduces a new queuing delay and possibly a processing delay if it is a security checkpoint (i.e. firewall or encryption/decryption point). Also, larger packets tend to cause bandwidth congestion and increased latency. In light of these issues, VOIP tends to work best with small packets on a logically abstracted network to keep latency at a minimum.

End-to-End Delay

Different types of conversation require different switching speed and thus have a different sensibility to delay .for example, a business call might require a higher level of interactivity than a social call.
Besides, various studies insist on the fact that, for natural hearing the end to end delay should be approximately 150ms. While the benefit of delays lower than an average listener cannot really appreciate 100ms, delays above 150ms are noticed by the users and become a hindrance to interactivity. In order to account for the great impact of the interactivity threshold (of 150ms) on a conversation quality. We consider three different delay impairment functions Id1(d) considered is based on the utility function and characterizations a user with strong interactivity requirements. The delay impairment is expressed as follows.

$$Id1(d) = \begin{cases} \gamma 1d \rightarrow if d \leq 150 \\ b1\tanh(\beta(d-b2)) + b3 \rightarrow if 150 < d < 300 \\ \delta + \gamma 2d \rightarrow if d \geq 300 \end{cases} \quad (3.1)$$

Where d is the mouth to ear delay expressed in ms, $\gamma 1=\gamma 2=0.01$, $\beta=0.02$, $\delta=50$ and b1, b2 and b3 are constancies selected to ensure the continuity of Id1.
The second and the third delay impairment functions considered are based on the E- model: Id2(d) =Id1 (d, 51) represents a user that is annoyed by delay because of echo, but with out a clear threshold effect and Id3(d) =Id1(d) represents a user attaches a small importance to delay (in an echo-free conversation) or the value of network parameter can be calculated by using a Linear Regression Trendline (LRT) method to predict the network delay, and then to deduce it. The network delays of latest n packets are chosen for the calculation of LRT, and the equations of the LRT can be expressed as

$\phi=\alpha T+\theta$

Where ϕ the failure of network delay, T is the VoIP system running time, and ∂ as well as θ can be obtained from

$$\partial = \frac{\sum_{i=1}^{n} xy - \frac{\sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n}}{\sum_{i=1}^{n} x^2 - \frac{\left(\sum_{i=1}^{n} x\right)^2}{n}} \quad (3.2)$$

$$\theta = \bar{y} - \partial \bar{x} \quad (3.3)$$

Where y is the network delay of last n packets and x is the corresponding send-time of each packet [29].

3.8.2 Jitter

Jitter refers to non-uniform packet delays. It is often caused by low bandwidth situations and security effects in VOIP and can be exceptionally detrimental to the overall QoS. Variations in delays can be more detrimental to QoS than the actual delays themselves. Jitter can cause packets to arrive and be processed out of sequence. RTP, the protocol used to transport voice media, is based on UDP .So packets out of order are not reassembled at the protocol level. However, RTP allows applications to do the reordering using the sequence number and timestamp fields. The overhead in reassembling these packets is non-

trivial, especially when dealing with the tight time constraints of VOIP.

When jitter is high, packets arrive at their destination in spurts. This situation is analogous to uniform road traffic coming to a stoplight. As soon as the stoplight turns green (bandwidth opens up), traffic races through in a clump. The general prescription to control jitter at VOIP endpoints is the use of a buffer, but such a buffer has to release its voice packets at least every 150 ms (usually a lot sooner given the transport delay) so the variations in delay must be bounded. The buffer implementation issue is compounded by the uncertainty of whether a missing packet is simply delayed an anomalously long amount of time, or is actually lost. If jitter is particularly unpredictable, then the system cannot use past delay times as an indicator for the status of a missing packet. This leaves the system open to implementation specific behaviour regarding such a packet.

Jitter can also be controlled throughout the VOIP network by using routers, firewalls, and other network elements that support QoS. These elements process and pass along time urgent traffic like VOIP packets sooner than less urgent data packets. However, not all network components were designed with QoS in mind. An example of a network element that does not implement this QoS demand is a crypto-engine that ignores Type of Service (ToS) bits in an IP header and other indicators of packet urgency. Another method for reducing delay variation is to pattern network traffic to diminish jitter by making as efficient use of the bandwidth as possible. This constraint is at odds with some security measures in VOIP. Chief among these is IPsec, whose processing requirements may increase latency, thus limiting effective bandwidth and contributing to jitter. Effective bandwidth is compromised when packets are expanded with new headers. In normal IP traffic, this problem is negligible since the change in the size of the packet is very small compared with the packet size. Because VOIP uses very small packets, even a minimal increase is important because the increase adds across all the packets, and VOIP sends a very high volume of these small packets.

The window of delivery for a VOIP packet is very small, so it follows that the acceptable variation in packet delay is even smaller. Thus, although we are concerned with security, the utmost care must be given to assuring that delays in packet deliveries caused by security devices are kept uniform throughout the traffic stream. Implementing devices that support QoS and improving the efficiency of bandwidth with header compression allows for more uniform packet delay in a secured VOIP network.

3.8.3 Packet Loss

VOIP is exceptionally intolerant to packet loss. Packet loss can result from excess latency, where a group of packets arrives late and must be discarded in favour of newer ones. It can also be the result of jitter,

which is, when a packet arrives after its surrounding packets have been flushed from the buffer, making the received packet useless. VOIP-specific packet loss issues exist in addition to the packet loss issues already associated with data networks; these are the cases where a packet is not delivered at all. Compounding the packet loss problem is VOIP's dependence on RTP, which uses the unreliable UDP for transport, and thus does not guarantee packet delivery. However, the time constraints do not allow for a reliable protocol such as TCP to be used to deliver media. By the time a packet could be reported missing, retransmitted, and received, the time constraints for QoS would be well exceeded. The good news is that VOIP packets are very small, containing a payload of only 10-50 bytes, which is approximately 12.5-62.5 ms, with most implementations tending toward the shorter range. The loss of such a minuscule amount of speech is not discernable or at least not worthy of complaint for a human VOIP user. The bad news is these packets are usually not lost in isolation. Bandwidth congestion, security and other such causes of packet loss tend to affect all the packets being delivered around the same time. So although the loss of one packet is fairly inconsequential, probabilistically the loss of one packet means the loss of several packets, which severely degrades the quality of service in a VOIP network.

In a comparison of VOIP quality versus traditional circuit switched networks, the reported data from a Telecommunications Industry Association (TIA) study that showed even a fairly small percentage of lost packets could push VOIP network QoS below the level users have come to expect on their traditional phone lines. Each codec the TIA studied experienced a steep downturn in user satisfaction when latency crossed the 150 ms point. However, even with less than 150 ms of latency, a packet loss of 5% caused VOIP traffic encoded with G.711 (an international standard for encoding telephone audio on a 64 kbps stream) to drop below the QoS levels of the PSTN, even with a packet loss concealment scheme. Similarly, losses of 1 and 2 percent, respectively, were enough to place quality in VOIP networks encoded with G.723.1 (for very low bit rate speech compression) and G.729A (for voice compression on an 8kbps stream) below this threshold. At losses of 3 and 4 percent, respectively, the performance of these networks resulted in a majority of dissatisfied users. The study said that "tolerable loss rates are within 1-3% and the quality becomes intolerable when more than 3% of the voice packets are lost". The studies found that greater payload compression rates resulted in a higher sensitivity to packet loss. On the bright side, the implementation of forward error correction, and packet loss concealment schemes produced a VOIP network that was less sensitive to packet loss. The percentages presented in both studies did not take into account varying packet sizes and several other properties that can affect the relationship between packet loss and QoS.

Despite the infeasibility of using a guaranteed delivery protocol such as TCP, there are some remedies for the packet loss problem. One cannot guarantee all packets are delivered, but if bandwidth is available, sending redundant information can probabilistically annul the chance of loss. Such bandwidth is not always accessible and the redundant information will have to be processed, introducing even more latency to the system and ironically, possibly producing even greater packet loss. Newer codecs such as Internet Low Bit-rate Codec (iLBC) are also being developed that offer roughly the voice quality and computational complexity of G.729A, while providing increased tolerance to packet loss [27].

Audio Packets Loss Process

There have been many research efforts in the measurement and modelling of end-to-end Internet characteristics. The main result is that the correlations structure of the loss process of audio packets can be modelled with low order Markov Chains [35, 36]. In

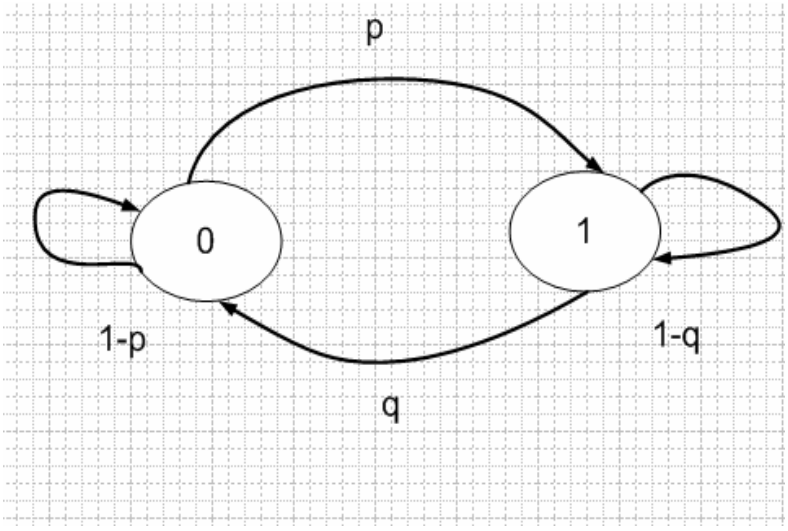


Fig 3.10 the Gilbert Model

Particular, a two state Gilbert model was found to be an accurate model in many studies. The Gilbert model is a two state model in which state 1 represents a packet loss and state 0 represents a packet reaching the destination. The parameters p and q denote respectively the probabilities of passing from state o (no loss) to state 1 (loss) and from state 1 to state 0. The stationary probabilities to be in state 1(Π1) and in state o (Π0) are given by: -

$$\Pi 1 = \frac{P}{P+q}$$

$$\Pi 0 = \frac{q}{P+q}$$

In absence of redundant information, the packet loss rate is given by the unconditional probability to be in state 1:

Packet loss rate PLR=Π1, the Gilbert model also allows to compute the packet loss rates after reconstruction when forwarded error correction (FEC) is used.

The n- stage transition matrix

Pn = [Pi j (n)], i, j Є {0, 1} is given

$$Pn=1/p+q \begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1\text{-}p\text{-}q)}{P+q} \begin{pmatrix} p & -p \\ -q & q \end{pmatrix} \qquad (3.4)$$

Let S be defined as the mean packet size and Ps be the packet drop probability computed at the gate way for this mean packet size. Now consider two rate controlled flows sending packets of different size. The first flow sends Ns packets of size S per rounds trip time and the second one sends Ns packets of size $s = S/\eta$ per round trip time.
To ensure fairness between flows sending packets of different sizes, we must drop packets based on their size in a way that all the flows achieve the same throughput (in bytes/sec). The problem can thus be expressed as follows. For any packet size s, find the corresponding drop probability Ps such that

$$Ns = \eta NS$$

Where η is a decrease in packet size.

Under the Bernoulli loss assumption, results allow expressing equation as a function of the packet loss rate (as opposed to the loss event rate); consequently NS can be expressed as follows,

$$NS= \frac{1}{\sqrt{\frac{2}{3(Ns-1+\frac{1}{Ps}}} + (12\sqrt{\frac{3}{8(Ns-1+\frac{1}{Ps}}})\frac{1}{Ns-1+\frac{1}{Ps}}(1+32(\frac{1}{Ns-1\frac{1}{Ps}})2)} \qquad (3.5)$$

One can see that NS is only a function of Ps.

The packet drop probabilities can be modified according such that large packets are more likely to be dropped than small packets. The final packet drop probability has to be weighted by the ratio of the size of the current packet si to a maximum packet size S.

$$Pa \longleftarrow \frac{PbSi}{(1-count.Pb)S}$$

Note: Pa is the arriving packets dropped probability.
And count needs to be increased by the appropriate fraction of a large packet

$$Count \longleftarrow count + \frac{Si}{S}$$

Note: count is the number of transmitted packets (i.e. packets that were not dropped)

In this case, the probability distribution of the number of packets between packet drops is

$$P(L=M) = \begin{cases} 0 \rightarrow if \sum_{i=1}^{m} \frac{si}{S} > \frac{1}{Pb} \\ Pb\frac{sm}{S} \rightarrow if \sum_{i=1}^{m} \frac{si}{S} \leq \frac{1}{Pb} \end{cases} \quad (3.6)$$

Note: si is the size of ith incoming packet after a drop and Pb is the current packet drop probability [33, 34].

3.8.4 Bandwidth & Effective Bandwidth

In any network, the obvious first concern is whether the network is available for use. Since a network can be broken down into nodes and links between nodes where traffic flows, the quest for an available network boils down to the availability of each node, and the availability of each path between the nodes. Later on, we will consider the nodes themselves, in cases where firewalls, CPUs, or other endpoints are unavailable, but for now we will concentrate on the availability of the edges: the bandwidth of the VOIP system.

As in data networks, bandwidth congestion can cause packet loss and a host of other QoS problems. Thus, proper bandwidth reservation and allocation is essential to VOIP quality. One of the great attractions of VOIP, data and voice sharing the same wires, is also a potential headache for implementers who must allocate the necessary bandwidth for both networks in a system normally designed for one. Congestion of the network causes packets to be queued, which in turn contributes to the latency of the VOIP system. Low bandwidth can also contribute to non-uniform delays (jitter), since packets will be delivered in spurts when a window of opportunity opens up in the traffic.

Because of these issues, VOIP network infrastructures must provide the highest amount of bandwidth possible. On a LAN, this means having modern switches running at 100M bit/sec and other architectural

upgrades that will alleviate bottlenecks within the LAN. The reports suggest that if network latencies are kept below 100 milliseconds, maximum jitter never more than 40 milliseconds, then packet loss should not occur. With these properties assured, one can calculate the necessary bandwidth for a VOIP system on the LAN in a worst-case scenario using statistics associated with the worst-case bandwidth congesting codec. This is fine when dealing simply with calls across the LAN, but the use of a WAN complicates matters. Bandwidth usage varies significantly across a WAN, so a much more complex methodology is needed to estimate required bandwidth usage.

Methods for reducing the bandwidth usage of VOIP include RTP header compression and Voice Activity Detection (VAD). RTP compression condenses the media stream traffic so less bandwidth is used. However, an inefficient compression scheme can cause latency or voice degradation, causing an overall downturn in QoS. VAD prevents the transmission of empty voice packets (i.e. when a user is not speaking, their device does not simply send out white noise). However, by definition VAD will contribute to jitter in the system by causing irregular packet generation.

The bandwidth requirements are designed for a basic VOIP system. Adding security constraints significantly increases the bandwidth usage, causing more latency and jitter, thereby degrading the overall QoS of the network. In addition, these requirements do not explicitly take into account the heterogeneous data flow over the network. Since voice and data streams are sharing the same finite bandwidth, and data streams tend to contain much larger packets than VOIP, significant amounts of data can congest the network and prevent voice traffic from reaching its destination in a timely fashion. For this reason, most new hardware devices deployed on networks support QoS for VOIP. These devices, such as routers and firewalls, make use of the IP protocol's Type of Service (ToS) bits to send VOIP traffic through before less time urgent data traffic. VOIP phones often also include QoS features.

Not only is the available bandwidth of the system affected by the introduction of security measures, but in addition the effective bandwidth of the VOIP system is significantly depreciated. Effective bandwidth is defined as "the percentage of bandwidth carrying actual data with regard to the total bandwidth used." The introduction of IPsec or other forms of encryption results in a much larger header to payload ratio for each packet, and this reduces the effective bandwidth as the same numbers of packets (but larger sized) are used to transport the same amount of data. The consequences of this reduction include decreased throughput and increased latency.

3.8.5 The Need for Speed

The key to successful QoS issues like latency and bandwidth congestion is speed. By definition, faster throughput means reduced

latency and probabilistically reduces the chances of severe bandwidth congestion. Thus every aspect of network traversal must be completed quickly in VOIP. The latency often associated with tasks in data networks will not be tolerated. Chief among these latency producers that must improve performance are firewall/NAT traversal and traffic encryption/decryption. Traditionally, these are two of the most effective ways for administrators to secure their networks. However, they are also two of the greatest contributors to network congestion, jitter and throughput delay. Inserting traditional firewall and encryption products into a VOIP network is not feasible, particularly when VOIP is integrated into existing data networks. Instead, these data-network solutions must be adapted to support security in the new fast paced world of VOIP.

3.8.6 Power Failure and Backup Systems

Conventional telephones operate on 48 volts supplied by the telephone line itself. This is why home telephones continue to work even during a power failure. Most offices use PBX systems with their conventional telephones, and PBXs require backup power systems so that they continue to operate during a power failure. These backup systems will continue to be required with VOIP, and in many cases will need to be expanded. An organization that provides uninterruptible power systems for its data network and desktop computers may have much of the power infrastructure needed to continue communication functions during power outages, but a careful assessment must be conducted to ensure that sufficient backup power is available for the office VOIP switch, as well as each desktop instrument. Costs may include electrical power to maintain UPS battery charge, periodic maintenance costs for backup power generation systems, and cost of UPS battery replacement. If emergency/backup power is required for more than a few hours, electrical generators will be required. Costs for these include fuel, fuel storage facilities, and cost of fuel disposal at end of storage life.

3.9 VoIP Security Solutions

As described above, firewalls devices pose problems for incoming calls. Call setup procedures require opening up a wide range of UDP ports, and firewalls have to open “pinholes” through which the outbound traffic flows. So firewalls degrade Quality of Service (QOS) and introduce latency and jitter. Firewalls have to inspect each packet that passes through them, inspecting the packets for their validity may degrade the service

3.9.1 Solutions

3.9.1.1 VoIP-aware firewalls

VoIP firewalls have the special feature of sensing and distinguishing voice packets from regular packets. When an incoming call arrives, SIP-aware firewalls analyze the dynamic rules or assign a specialized SIP device into the firewall device itself allowing SIP traffic after passing through VoIP-specific security rules. They have the ability to inspect the packets coming into the firewall and recognize SIP traffic. SIP dynamically assigns the UDP ports for RTP/RTCP streams. SIP-aware firewalls comprehend each SIP message and extract the RTP/RTCP port information.

3.9.1.2 VLANs

Virtual LANs are a good choice for logically separating voice and data traffic to prevent the data network problems from affecting voice traffic. VoIP components like IP PBXs, IP phones and VoIP servers should be isolated at layer 2 and be placed on their own VLAN, segregating them from other traffic that has multicasts and broadcasts, which might cause latency and jitter. VLANs create a logical segmentation of broadcast and collision domains thus improving the performance of network. In addition, VLANs provide security and reduce broadcast traffic to voice applications by creating a separate VLAN for voice applications. Switched networks are recommended over hubs to connect different VLANs because they provide more security than hubs.

3.9.1.3 Session Border Controllers

Session Border Controllers (SBCs) are dedicated network devices that are located at the network borders and offer services like firewall/NAT traversal and signal and media control functions. They are complicated devices that serve as VoIP-aware NATing firewalls. SBCs are divided into two types of architectures: stand-alone and distributed. Stand-alone SBC contains all the control and intelligence needed to process media and signalling of the VoIP call. Distributed SBC has two types. The signalling SBC controls access of VoIP signalling messages and manages the contents of these messages. The Media SBC controls access of media messages and manages the quality of services for various media streams. SBC provides security in the following ways:

- SBC performs the Firewalls/NAT traversal of SIP traffic with the existing NATs and firewalls and without any need for additional equipment.
- SBC provides security to core elements by identifying any malicious traffic before it reaches the core. It performs a

topology-hiding function to prevent outsiders from knowing about the internal details.

- SBC provides quality service by monitoring call admission control and Service Level Agreements (SLAs). It checks the bandwidth usage in order to maintain quality.
- SBC will be capable of handling emergency calls and monitoring lawful intercepts.

3.9.1. 4 Application Level Gateway

Application Level Gateway (ALG) is another solution for firewall/NAT traversal problems. ALGs, which are embedded in firewalls, can understand the type of traffic and can open/close ports dynamically. They have the capability of changing the header information in the packets and routing the information to the correct internal IP addresses in the private network or to the external IP addresses in the public network. They also map RTP/RTCP traffic into ports so that they can read and send to the correct destination.

Middlebox Communication Architecture and Framework (MIDCOM), a protocol being developed by IETF, is yet another solution for firewall/NAT traversal. The disadvantage of an ALG is that it resides in the firewall, which adds to the latency. MIDCOM is a device that sits outside the firewall and performs many of the functions associated with an ALG. It parses the incoming VoIP traffic and advises the firewall to open/close ports accordingly. This is advantageous over ALG for several reasons: The firewall doesn't have to change if the VoIP protocol changes and the firewall do not have the burden of processing the VoIP traffic.

An ALG is embedded software on a firewall or NAT, which allows for dynamic configuration based on application specific information. A firewall with a VOIP ALG can parse and understand H.323 or SIP, and dynamically open and close the necessary ports. When NAT is employed, the ALG needs to open up the VOIP packets and reconfigure the header information therein to correspond to the correct internal IP addresses on the private network, or on the public network for outgoing traffic. This includes modifying the headers and message bodies (e.g., SDP) in H.323 and SIP. ALG implementations are discussed for H.323 and SIP. The NAT problem is alleviated when the ALG replaces the private network addresses with the address of the ALG itself. It works by not only changing the IP address, but also mapping RTP traffic into ports the ALG can read from and forward to the correct internal machine. The need for consecutive ports for RTP and RTCP can cause a problem here because all VOIP traffic on the network (and data traffic as well) is being routed through the ALG, so as call volume increases, finding enough consecutive ports may become an issue. So although both endpoints may have adequate ports to convene a conversation, the firewall's deficiencies may cause the call to be rejected as "busy" by the ALG itself.

There are significant performance and fiscal costs associated with the implementation of an ALG. Performance-wise, the manipulation of VOIP packets introduces latency into the system and can contribute to jitter when high call volumes are experienced. Depending on the firewall architecture, this can also slow down throughput in the firewall, contributing to general network congestion. A firewall with ALG support can be expensive, and would need to be upgraded or replaced each time the standards for VOIP change. Also, the addition of application intelligence to a firewall can introduce instabilities into the firewall itself. Some firewalls have been found vulnerable to an attack in which a high rate of call setups can be sent, depleting the connection tables of the firewall. These half-open VOIP sessions may not time out in the firewall for more than 24 hours. Still with all these detractions, an ALG remains the simplest and safest workaround to allow the coexistence of VOIP, firewalls, and NAT.

3.9.2 Codec and (De-) Packetiser

Transmission of analogue signal is not acceptable in IP networks. As abovementioned, voice analogue signals are first coded/converted into digital format. Encoder's carry out these conversion processes, and decoders can accomplish reversed processes. Both two sorts of mechanism are usually called codec.

Pulse Code Modulation (PCM) and Adaptive Differential PCM (ADPCM) are examples of waveform codec techniques, and can be used in traditional PSTNs. Waveform codecs are compression techniques that exploit the redundant characteristics of the waveform itself. Different to waveform codecs, source codecs compress speech by sending only simplified parametric information about voice transmission, which require less bandwidth. Source codecs include Linear Predictive Coding (LPC), (Conjugate Structure-) Algebraic Code-Excited Linear Prediction ((CS-) ACELP) and MultiPulse-MultiLevel Quantization (MP-MLQ).

Coding techniques for VoIP are standardized by the ITU-T, such as widely used G.711, G.723.1 and G.729 codecs. Table 7-1 presents some information of certain representative codecs using different coding techniques.

Table 4-1 General information of codecs [24]

Codec	Bite rate(kpbs)	Frame size(ms)	Coding technique
G.711	64	0.125	PCM
G723.1	5.3	30	ACELP
	6.3	30	MP-MLQ
G.729	8	10	CS-ACELP

In general, the use of G.711 can obtain high quality voice output, but low compression rate and high bit rate limit its development in VoIP

application, currently required services for low bandwidth. Both G.723.1 and G.729 can however provide drastic rate reduction, and also achieve the acceptable voice quality, as the MOS (mean opinion score) is from 3.65 to 3.9.
Packetiser is a type of mechanism, which follows the encoder. The function of it is to place voice frames after encoding into RTP, (RTCP), UDP and IP contained packets.
De-packetiser reverses this process before VoIP packets are decoded.

3.9.3 Playout Buffer

In VoIP systems, the voice packet must be played out at the receiver site in a timely manner and in the order they were emitted from the sending site. The basic function of a playout buffer is to collect packets and to store them, and then to send specified number of packets to the next mechanism.
Playout buffer is located at the end of VoIP system, and the main intention of it is to smooth speech. When variable delays take place, playout buffer can allow some later-arrived packets to be played out, which depends on the set-up of packet playout time, to keep the completeness of speech. However, any packet, arrived later than the playout time, will be simply discarded.
The set-up of playout time can be fixed or adaptive, but both need synchronization between sender and receiver due to the changing network delay. Fixed playout time set-up/scheduling is simply, but normally causes a constant delay and cannot follow the change of network delays. Adaptive playout scheduling was introduced to overcome these problems, and is controlled by corresponding playout buffer algorithm, which can utilize the silence time between two successive voice periods, referred to as talkspurt, to slow down or speed up playout time of each talkspurt [24].

Chapter Four

4. Simulation results and Discussion

4.1 Secure VoIP Network Simulation Model

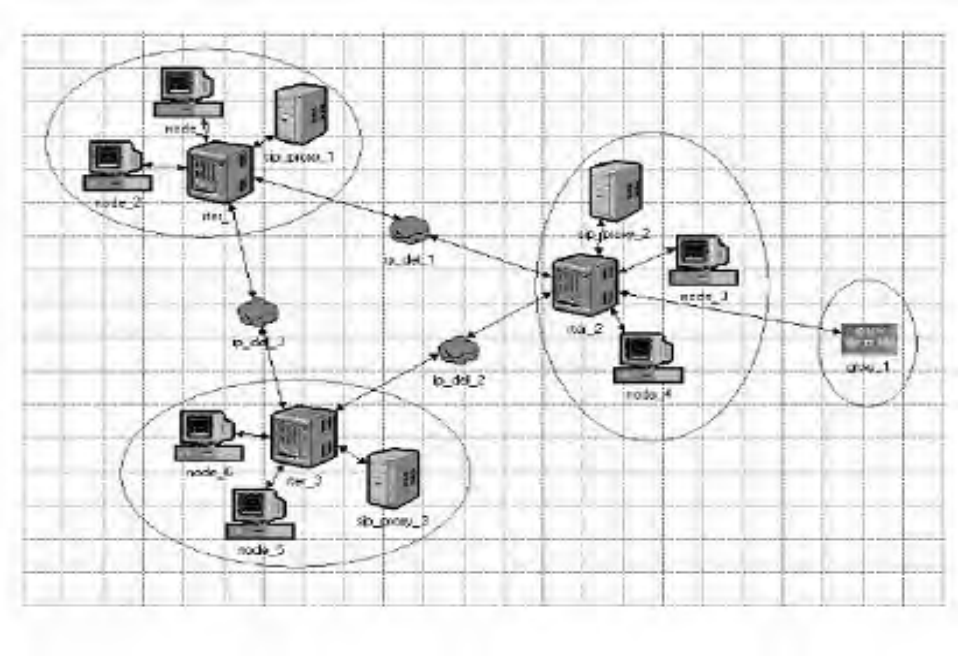


Fig 4.1. VOIP Network Topology

Figure 4.1, shows a general VoIP network model developed using the OPNET Modeller network simulation tool. The network illustrated in this figure simulates a small scale of VoIP networks, only for the clarity of illustration. The model can, and has been, scaled to simulate the operation of more complex secure VoIP networks. The VoIP stack is implemented over an UDP/IP infrastructure. SIP/H.323 is used as the signalling protocol for VoIP call establishment, and RTP is used to carry the media stream. The main components that build up the network model are the SIP/H.323 Terminals, Servers, IP Routers, VoIP-PSTN Gateways and IP Cloud models.

4 .2 Simulation result discussion

The objective of this OPNET simulation is to study the performance evaluation of VOIP in providing security to shared-public network such as the Internet.

Computer networks are typically a shared resource used by many applications for many different purposes. Sometimes the data transmitted between application processes is confidential, and the applications would prefer that others not be able to read it.

A firewall is a specially programmed router that sits between a site and the rest of the network. It is a router in the sense that it is connected to two or more physical networks and it forwards packets from one network to another, but it also filters the packets that flow through it. A firewall allows the system administrator to implement a security policy in one centralized place. Filter-based firewalls are the simplest and most widely deployed type of firewall. They are configured with a table of addresses that characterize the packets they will, and will not, forward.

The virtual private network (VPN) is an example of providing a controlled connectivity over a public network such as the Internet. VPN utilizes a concept called IP tunnel. IP tunnel is a virtual point-to-point link between a pair of nodes that are actually separated by an arbitrary number of networks. The virtual link is created within the router at the entrance to the tunnel by providing it with the IP address of the router at the far end of the tunnel. Whenever the router at the entrance of the tunnel wants to send a packet over this virtual link, it encapsulates the packet inside an IP datagram. The destination address in the IP header is the address of the router at the far end of the tunnel, while the source address of that of the encapsulating router.
In this simulation you will set up a network where customers who have different privileges access servers over the Internet. You will study how firewalls and VPNs can provide security to the information in the servers while maintaining access for customers with the appropriate privilege. Different simulation scenario is stated below.

1) No_ Firewall Scenario: - in this network no security mechanisms is applied to have secure communication between the sender and the receiver, that is any body has an opportunity to access the information from the server.

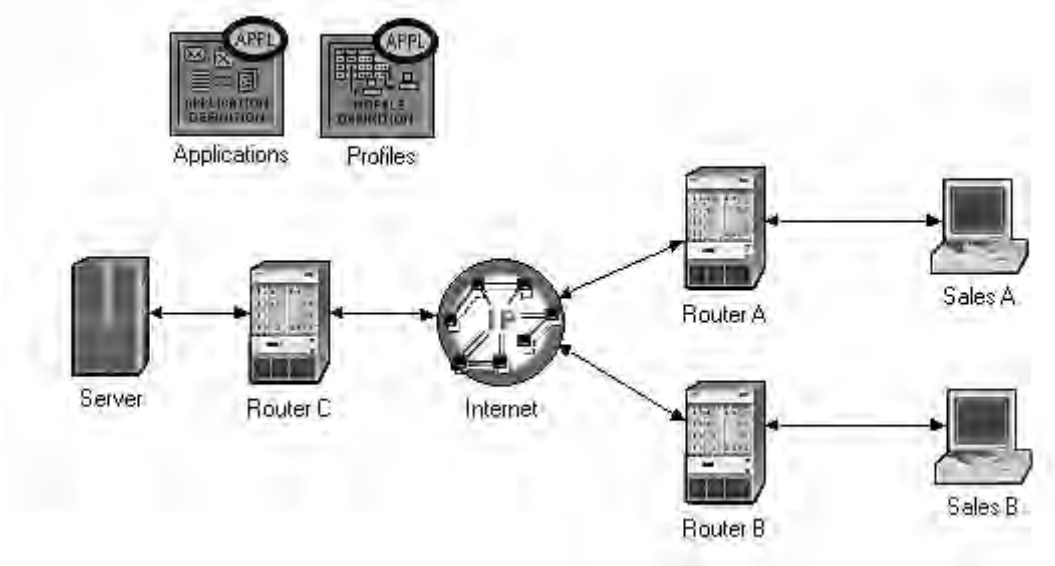


Fig 4.2 No Firewall Scenario

2) The Firewall Scenario

In the network we just created the Person profile allows both sales sites to access applications such as Database Access, Voice from the server. Assume that we need to protect the database and Voice in the server from external access, including the sales persons. One way to do that is to replace Router C with a Firewall as follows:

Our Firewall configuration does not allow database-related traffic to pass through the firewall (it filters them out). This way the databases and the Voice in the server are protected from external access. My Firewall scenario should look like the following figure.

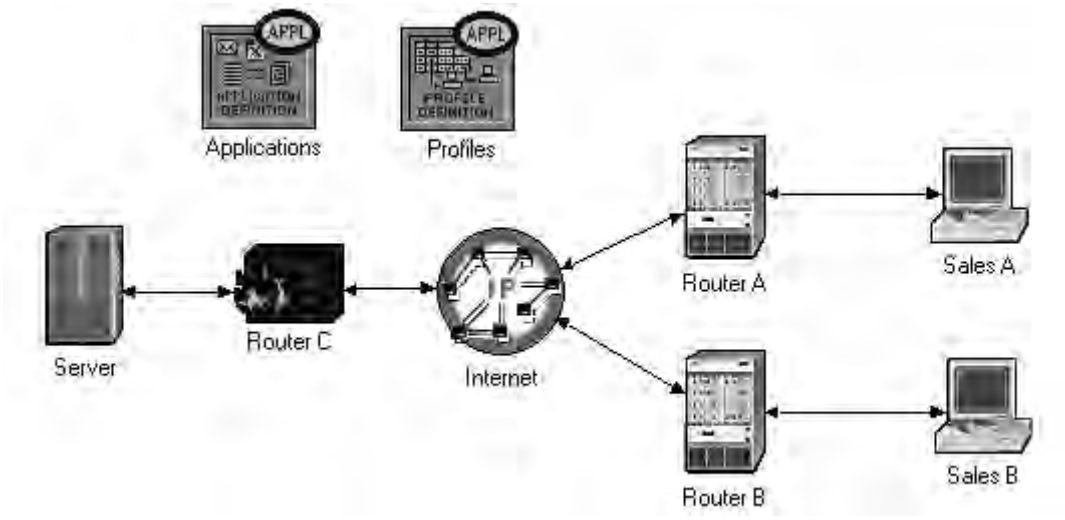


Fig 4.3 Firewall Scenario

3) The Firewall_VPN Scenario

In the Firewall scenario, we protected the databases in the server from "any" external access using a firewall router. Assume that we want to allow the persons in the Sales A site to have access to the databases in the server. As the firewall filters all data and voice base related traffic regardless of the source of the traffic, we need to consider the VPN solution. Sales A to send database and voice request to the server can use a virtual tunnel. The Firewall will not filter the traffic created by Sales/voice A because the IP packets in the tunnel will be encapsulated inside an IP datagram.

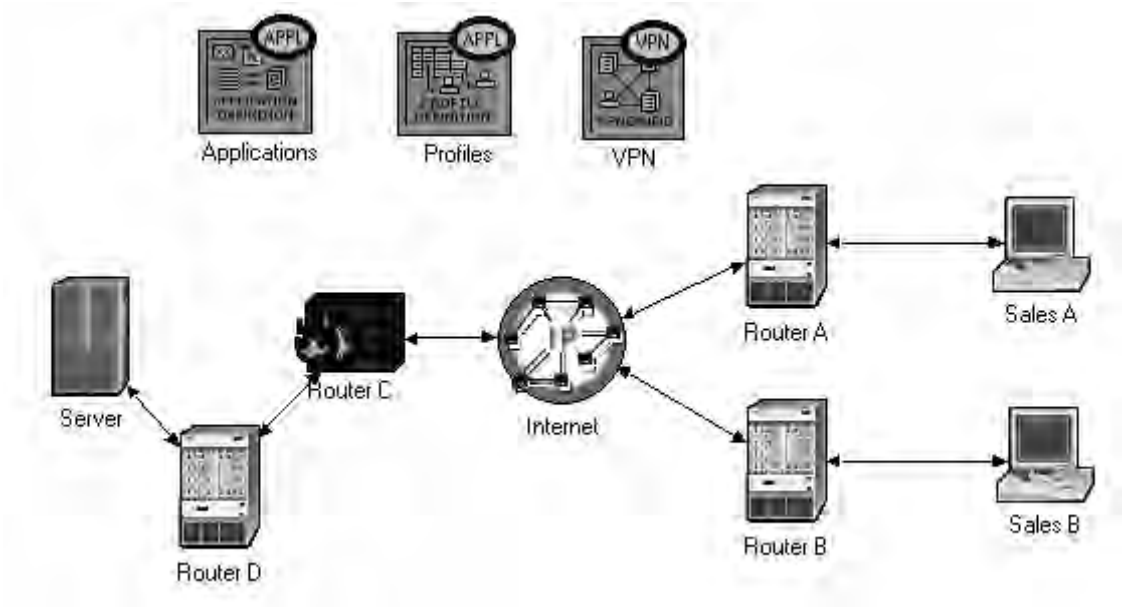


Fig 4.4 Firewall_VPN Scenario

As we have seen in the simulation diagram below different kinds of traffic is received from different design and configuration .For instance more traffic can be arrived at the destination using No-Firewall, while medium traffic is arrived at the end users when the network becomes Firewall_VPN, but less traffic is received using Firewall.

Reference from table 4.1 Standard Value

Simulation Parameters	
Delay	0.5ms
Loss	0.6%
Jitter	20ms
BW	64kB

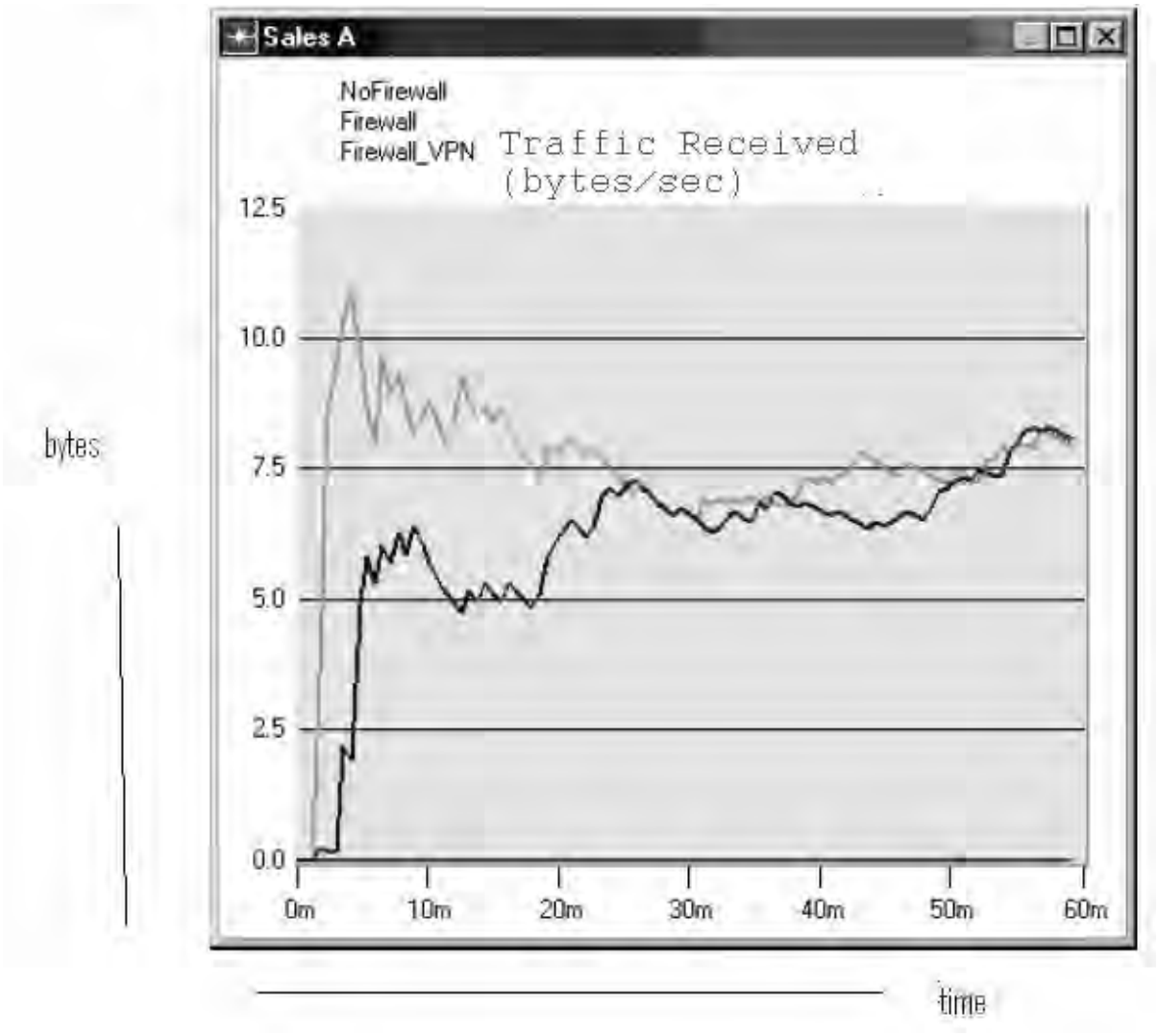


Fig 4.5 Traffic Received VS Distance Diagram

On the other hand the simulation shows how much time is required to transmit the information between the two end users, as shown on the graph below more time is needed for Firewall to deliver the voice to the receiving stations, and less amount of time is required for No Firewall to deliver the information to the end users.

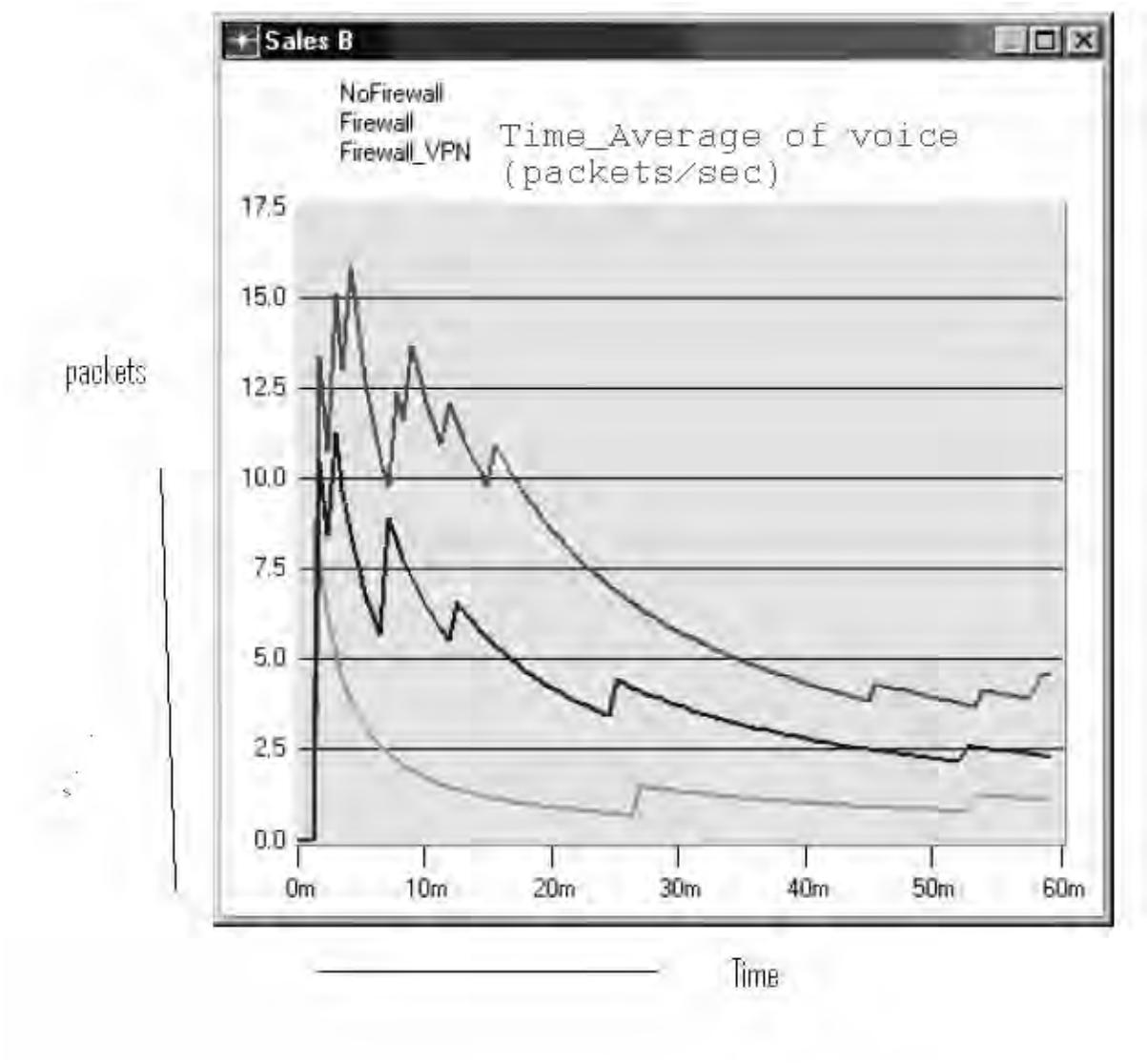


Fig 4.6 Time Average VS Distance Diagrams

Generally, as shown fig 4.7 the simulation result indicates that there are more packet delay, packet loss and delay variation because of using security mechanisms especially Firewall.
Which is, as the distance is increased the traffic received becomes more decrease, because of this the delay is increased, which makes the packet loss is more.

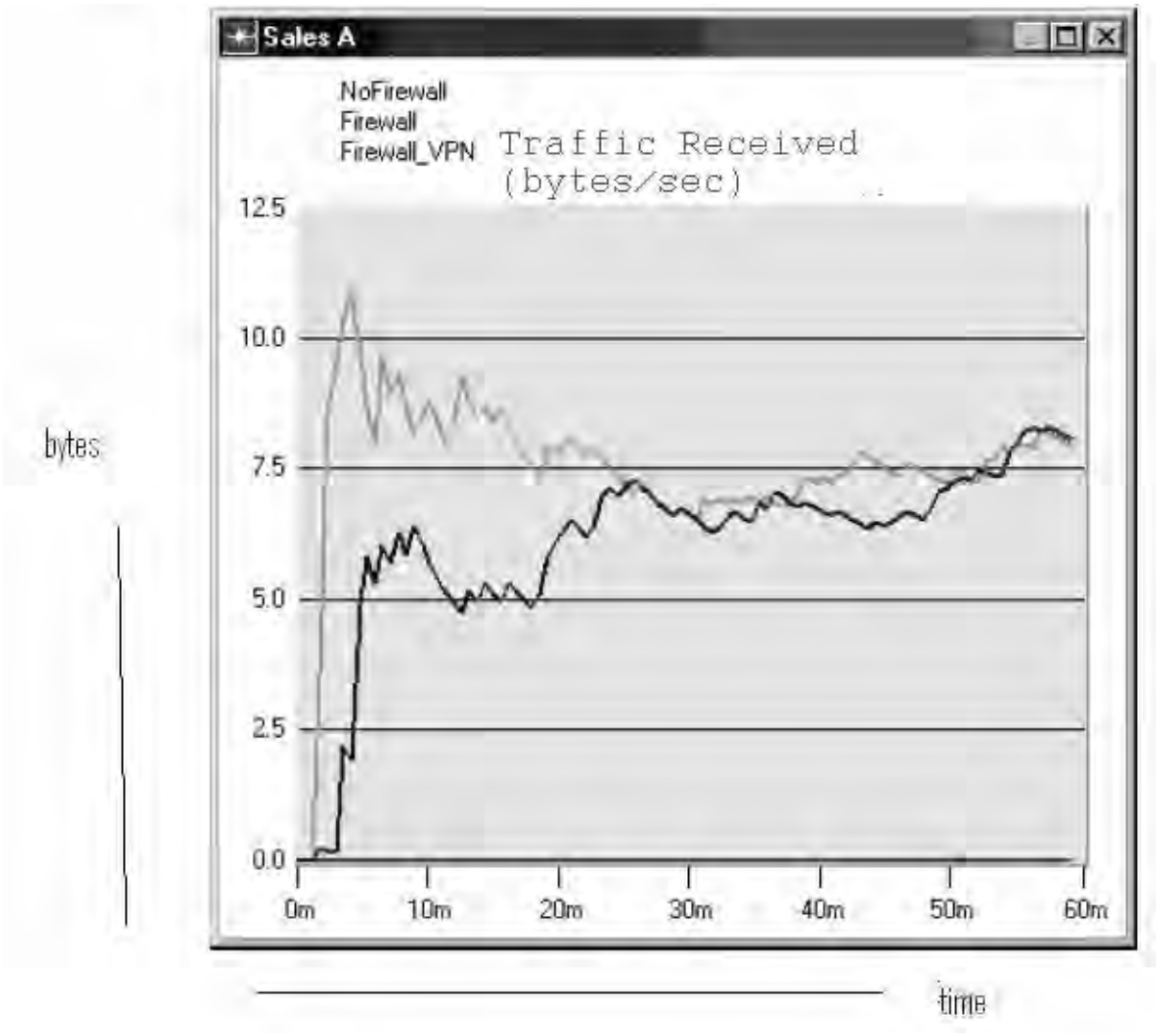


Fig 4.7 End-to-End Delay VS Distance Diagram

4.3 Comparison of the Standard values and the Simulation results

Table 4.1 Standard Value [23]

THEORETHECAILE VALUES			
	Good	Acceptable	Poor
Delay	(0-150) ms	(150-300) ms	>300ms
Jitter	(0-20) ms	(20-50) ms	>50ms
Loss	(0-0.5) %	(0.5-1.5) %	>1.5%

Table 4.2 Simulation Value

SIMULATION VALUES			
	No firewall	Firewall	Firewall_VPN
Delay	Good (0-150) ms	Poor (>300ms)	Acceptable (150-300) ms
Jitter	Acceptable (20-50) ms	Poor (>50ms)	Acceptable (20-50) ms
Loss	Acceptable (0.5-1.5) %	Poor (>1.5%)	Acceptable (0.5-1.5) %

The ultimate objective of this simulation is to test whether VoIP security communication provides reliable, high-quality and low-cost services to the end users. However, these are not guaranteed on current IP networks, primarily because of bandwidth limitations and security effects. These combined impairment factors that lead to packet loss, delay and delay variation/jitter, (the three main parameters that determine IP network QoS, as user's perceived quality).

Generally VOIP security decreases the quality of voice conversation services between the two ends because of the above and the following effects

Overall packet loss.

1. Network packet loss.
Voice packets (data), transmitted from an originating device, do not arrive at the destination. It mainly results from congestion points on transmission routes in IP networks.

2. Discarded packet loss.
Packets, which arrive later than the playout time set up by playout buffer, are discarded by system.

Packet loss causes more noticeable degradation in voice quality, compared with other impairments, and is normally mitigated by PLC, FEC etc. mechanisms.

Overall delay.

1. Fixed delay.
It consists of transport delay, coding delay, (de-)packtising delay, etc

2. Variable delay (delay variation/jitter). It is the variation in inter-packet arrival time, and is mainly resulted from queuing effects in IP networks

3. Buffer delay. It is the delay introduced by playout buffer in order to output smoothed and sequencing voice packets (data), in other words, to deal with delay variations.

Small overall delay can be tolerant, but large delay can damage VoIP conversations. Fixed delay can be minimized using low-delay VoIP mechanisms, but is difficult to decrease. Increasing corresponding transmission bandwidth, which may well increase the cost of transmission, can decline variable delay. Buffer delay can, however, be well controlled by a good-performance buffer mechanism with introducing a master buffer algorithm.

Codec performance.

1. Coding loss.
Loss introduced by encoding process.

2. Coding delay.
Delay introduced by both encoding and decoding processes

As you see on the graph, the result not satisfied, because security mechanism degrades quality of service (QOS) and introduces latency and jitter. To reduce these negative impacts on voice over Internet Protocol (VOIP), different solutions which are stated on chapter 3 can be applied on the network.
Among the different solutions chapter five states that, the developed Adaptive Codec Selection Algorithm Mechanism to minimize the effects of security mechanisms in order to increase the performance of VOIP.

Chapter Five

5 Proposed Adaptive Codec Selection Mechanism

5.1 RTP packet format

Audio, video, and multimedia services require the use of RTP, which provides the necessary end to end delivery requirements of the time sensitive data. Both RTP and RTCP were designed to run independently of the underlying transport and network layers. RTP often runs in unison with the User Data Protocol (UDP), which supports multiplexing and checksums. UDP is the exploited by RTP because the retransmission nature of TCP is detrimental to VOIP. If a packet is not received at the destination, TCP will attempt to retransmit the lost packet. However strict time constraints for VOIP, the retransmission of the packet will be of no use due to the time sensitivity of the data. A typical RTP packet includes a sequence number that allows the receiver to reconstruct the data the sender has sent in the appropriate order. Further, it allows for time synchronization through the use of timestamps and can differentiate between multiple senders in a multicast stream.
However, it does not ensure the packets are received in order; guarantee the delivery of the packets, nor address reservation. It relies on lower layer services to provide the expected quality of services. The sequence number and timestamp assist in reordering the packets in the jitter buffer. Each time a packet arrives at the receiver, the buffer will place the packet in the appropriate order based on these values. If a packet is not available when it needs to be played out, the last frame to leave the buffer is copied and played out repeatedly until the timestamp of the next available packet is arrived. This RTP sequence number will be used to determine which RTP packet will contain new codec.

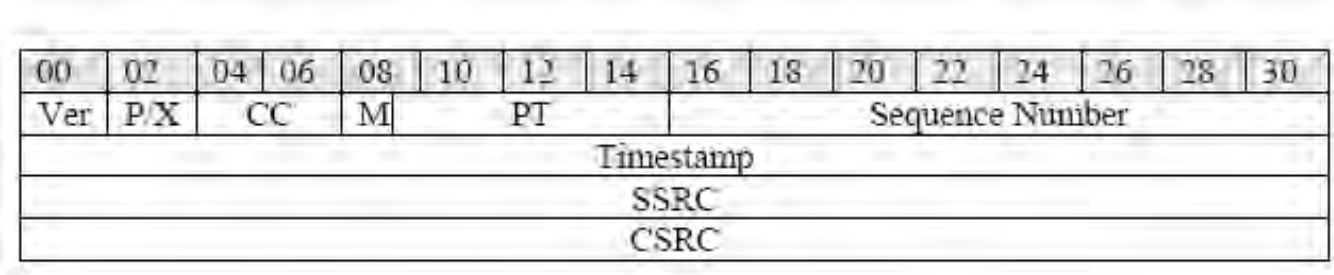


Fig 5.1 RTP packet format [30]

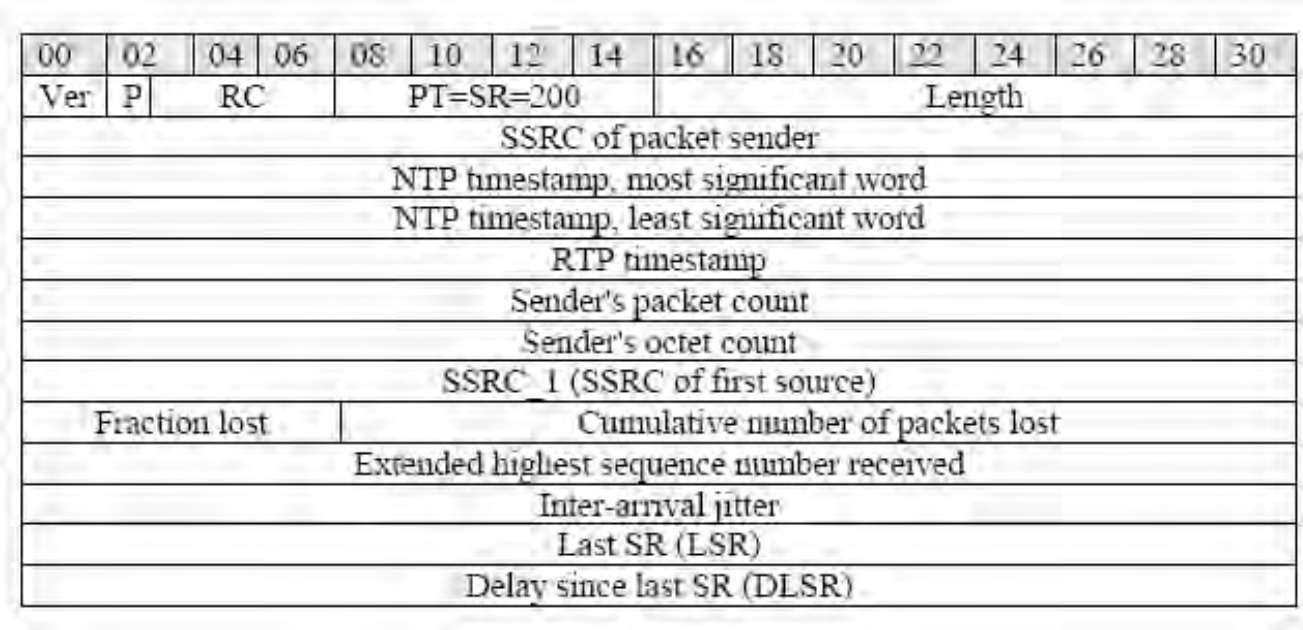
Ver, Version-Two bits specifying the RTP version number (usually version 2).
P, Padding –If this padding bit is set, it indicates padding bytes are present that are present that are not part of the payload.
X, Extension-A one bit value that indicates only one header extension follows.

CC, CSRC Count –A four bit value detailing the number of CSRC identifiers that follow the fixed header.
CSRC, Contributing Source- The contributing source is a thirty two bit values. It identifies the sources that contributed to the packet's payload. Up to fifteen elements can be represented.
M, Marker- This is a one bit value defined by a given profile that allows evens such as frame boundaries to be marked in the packet stream.
PT, Payload Type- The payload type is a seven bit value distinguishes what format the RTP payload includes and determines how the application is to interpret it.
Sequence Number – The sequence number is a sixteen bit value that increments for each RTP packet and can be used by the receiver to re-establish proper packet sequence or known plaintext attacks. Sequence numbers are unidirectional .an H.323 terminal A could send an RTP sequence number starting with 579 for the first packet sent to B. All subsequent packets sent in the direction of B would have sequence numbers incrementing by one from 579. However, packets flowing from B to A could start with an RTP sequence number by one.
Timestamp –The timestamp is a thirty two bit value that allows for synchronization and jitter calculations. It signals the instant the RTP packet is sampled. It is derived from a clock that increments linearly.
SSRC, Synchronization Source- This field is a thirty two bit value that identifies the synchronization source. This number should be a unique value so no two sources within the same RTP session have the same SSRC.

5.2 Real Time Control Protocol

RTCP allows network administrations to monitor network conditions and provides minimal control and identification functionality. It operates in unison with RTP, but does not transport data. It utilizes the periodic sending of control packets to active terminals. These control packets provides feedback on the quality of data reception, which is directly related to the transport protocol flow and congestion. Each user sends control packets to every other, which allows the send to observe each user independently. This information enables the sender to modify the encoding and transmission rates at any time during the session. The multiplexing of these control packets and data packets is done using separate port number with UDP.
There are several types of RTCP formats specified in RCF3550. These include sender report packet (SR), receiver report packet (RR), source description RTCP packet (SDES), goodbye RTCP packet (BYE), and application specification RTCP packet (APP), which are explained below.
SR- the sender report consists of transmission and reception statistics from active senders. This packet format is as shown below.

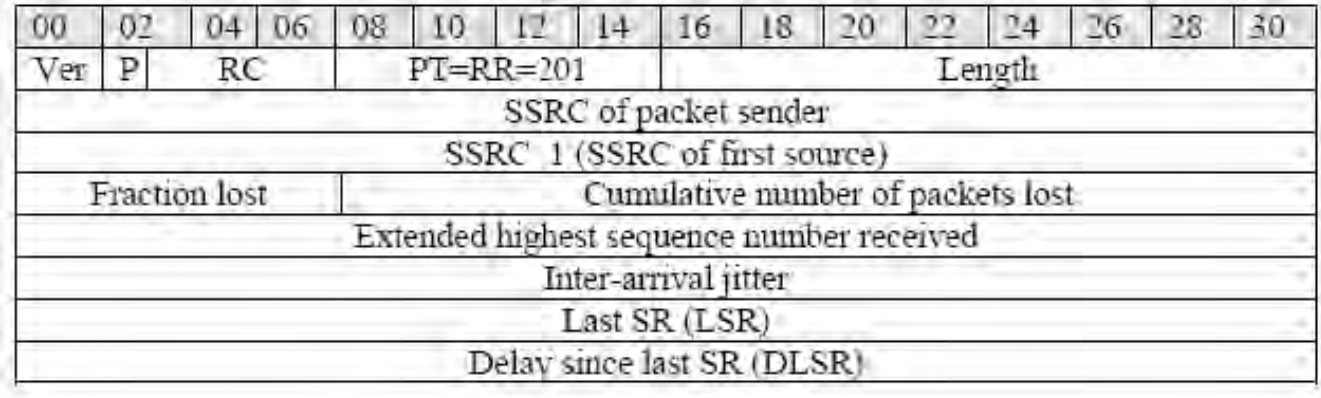
Table 5.2 RTCP sender report packet format [30]



00	02	04	06	08	10	12	14	16	18	20	22	24	26	28	30
Ver	P	RC		PT=SR=200				Length							
SSRC of packet sender															
NTP timestamp, most significant word															
NTP timestamp, least significant word															
RTP timestamp															
Sender's packet count															
Sender's octet count															
SSRC_1 (SSRC of first source)															
Fraction lost				Cumulative number of packets lost											
Extended highest sequence number received															
Inter-arrival jitter															
Last SR (LSR)															
Delay since last SR (DLSR)															

RR-the receiver report include reception statistics from non-active sender.
NTP-network time protocol

Table 5.3 RTCP receiver report packet format [30]



00	02	04	06	08	10	12	14	16	18	20	22	24	26	28	30
Ver	P	RC		PT=RR=201				Length							
SSRC of packet sender															
SSRC_1 (SSRC of first source)															
Fraction lost				Cumulative number of packets lost											
Extended highest sequence number received															
Inter-arrival jitter															
Last SR (LSR)															
Delay since last SR (DLSR)															

SDES- the source description packet contains source descriptor items such as a canonical name (CNAME) that is used as a transport level identifier for an RTP sender.
BYE- the goodbye packet indicates the end of participation
APP- the application specified packet describes application specification functions

Both the RTCP sender and receiver reports provide feedback on reception quality. If a station is only a receiver, the receiver report is generated. However, if the receiver is also a sender, then the sender report is generated. The only difference between the sender and the reviver report packet formats is the Packet Type (PT) and the five words of sender information in the sender report. The RTCP fields are explained below.

Ver, Version- Two bits specifying the RTP version number (usually version 2).
P, Padding- If this padding bit is set, it indicates padding bytes are present that are not part of the payload. The number of padding bytes to ignore is contained in the last byte of padding.
RC, Reception Report Count- This five bit represents the number of reception report blocks in the packet.
PT, Packet Type- this eight bit value is set to 200 for sender reports and 201 to identify receiver reports
Length –a sixteen bit value representing a 32-bit word, minus one, which allows for a valid zero length and avoiding potential loops in scanning of RTCP packets.
SSRC, Synchronization Source –identifies the originator of the report packet.
NTP timestamp- a thirty two bit value that assist in the calculation of RTT propagation time.
This is ideally selected to use a common wall clock that can be used as a reference for all members.
RTT timestamp- a thirty two bit value that corresponds to the NTP timestamp, but reflects the random offset as the RTP timestamp in data packets
Sender's Packet Count- a thirty two bit value that signifies the total number of RTP data packets sent by the sender from initial transmission to the current report packet being generated.
Sender's Octet Count- a thirty two bit value that indicates the total number of payload octets sent by the sender from initial transmission to the current report packet being generated.
SSRC_n, Synchronization Source- a thirty two bit SSRC identifier of the source that the information included in the report belong to.
Fraction Lost- an eight value that indicates the fraction of RTP data packets that have been lost since the previous sender or receiver report was generated. It is represented as the number of packets lost divided by the number of packets expected since the last report
Cumulative Number of Packet Lost –a twenty four-bit value that represents the total number of RTP data that have been lost since communication began.
Extended Highest Sequence Number Received- a thirty two bit that corresponds to the highest RTP sequence number received from a given source, SSRC_n, with the lower sixteen bits representing that value and the most significant sixteen bits signifying the number of sequence number cycles.

Inter arrival Jitter- is a thirty two-bit value that characterizes the statistical variance of RTP data packet inter arrival times. It is measure in timestamp units.
LSR, Last SR Timestamp- a thirty two-bit value that indicates the middle thirty two bits of the total sixty four bit NTP timestamp.
DLSR, Delay Since Last SR – a thirty-two bit value that denotes the elapsed time between receiving the last SR and sending a RR. It is expressed in units of 1/65536 seconds.

5.3 Selecting the Appropriate Codec

Codec are devices that encode or decode signals. Encoding a signal provides a more efficient form for transmission or storage, and decoding restores the signal to the original form. This compression and decompression of signals negatively impacts voice quality.
Therefore, selecting the appropriate codec is necessary to obtain best quality of voice with the lowest bandwidth requirements. Codec selection itself can make an enormous difference in voice quality. There are many codecs specified by the ITU-T. Among these are G.711, G.723.1, and G.729. The G.711 codec uses Pulse Code Modulation (PCM) and transmits at 64Kbps. This is equivalent to the traditional PSTN. Consequently, this codec uses high bandwidth, but has voice quality similar to PSTN. G.723.1 uses Algebraic Codec Excited Linear Prediction (ACELP) techniques and encodes or decodes at 6.3Kbps with a code delay of 37.5ms. G.723.1 can also decode at 5.3Kbps using Multipulse Maximum Likelihood Quantization (MP-MLQ) with similar delay. This codec uses significantly less bandwidth than G.711, but has a lower quality of voice and higher delay. Another codec, G.729.Uses Conjugate Structure Algebraic Codec Excited Linear Prediction (CS-ACELP) methods and consumes 8Kbps with a coding delay of 15ms.

Table 5.4 Values from “integrated voice and data networks” [31]

Codec	Bit Rate (Kbps)	Delay (ms)
G.711	64	0.125
G.723.1 MP-MLQ	5.3	37.5
G.723.1 ACELP	6.3	37.5
G.729 CS-ACELP	8	15

Table 5.5 illustrates the impairment attributed to various codecs, therefore, a tradeoff exists between high quality, high bandwidth

codecs and lower quality, lower bandwidth codecs that support more simultaneous calls while reducing clarity and increasing delay.

Table 5.5 Various codec impairment [32]

Codec	Bit rate (kbps)	Frame time (ms)	Look ahead (ms)	Codec impairment
G.711	64.0	10	0	0
G.723.1-MPMLQ	6.3	30	7.5	15
G.723.1-ACELP	5.3	30	7.5	19
G.729	8.0	10	5	11

Keep in mind that one-way delay must be around 150ms to maintain an acceptable voice quality for end users. Selecting the appropriate code requires a tradeoff of bandwidth for delay. Using higher bandwidth codecs will trigger higher cost, while using a lower the quality of voice. Over the LAN, a codec such as G .711 at 64Kbps could be used with ease, as these local networks have much more bandwidth and are subsequently less expensive as a result. When utilizing Wide Area Network (WAN) resources, however, one may want to consider a codec such as G.729 that has a relatively high voice quality and low coding delay over the expensive WAN link. This will reduce the necessity of large bandwidth costs over the WAN.
T1 lines, one of the more popular transport lines used by companies, can support up to 24 simultaneous telephone calls under the traditional PSTN circuit switched networks. These calls are encoded at 8 KHz with 8000 samples, or 64Kbps. When using VOIP, more than 24 simultaneous calls can be supported as well as the existing data traffic due to the selection of voice codecs that will reduce the bandwidth requirement, yet preserve call quality. Therefore, upgrading the T1 lines may not be required, depending on current network usage. The research states that 'dedicated network transports supporting computer data on traditional telephony system are generally about 30 present utilized.' Only data network analysis will determine the current utilization of the link, but with an average of 70% utilization free onT1 line, multiple
VOIP calls can be made without requiring an increase in bandwidth.

5.4 Proposed Algorithm Codec Selection

The proposed codec allocation scheme takes into account the dynamic network conditions as obtained from RTCP sender and /or receiver reports and expected call requirements, and it subsequently makes a codec selection based on those statistics. If current network conditions permit the use of higher bandwidth codec selection scheme will select a codec appropriately. However, if insufficient bandwidth is available

or the number of calls is too great, the codec selection scheme will only permit lowers bandwidth, lower quality codecs. Since this method adjusts utilization according to current network parameters, packet loss is likely to be reduced due to not oversubscribing that are more susceptible to packet loss.
The reception quality feedback provided in the RTCP reports was designed to be useful for senders, receivers, and third party monitors. Network administrator can use applications that evaluate network performance based on these RTCP reports. Similarly, receivers can perceive problems with links across a wide spectrum. In addition, senders can adjust media capabilities based on the parameters including packet loss rates, round-trip delays, and jitter obtained from RTCP reports. These parameters are maintained completely for overall performance measurements or at intervals between reports. This allows for measurements at shorter intervals that represent recent quality of the distributions or over long intervals that represent overall reception quality.
With respect to packet loss rates, the fraction lost and cumulative numbers of packets lost RTCP fields in the reports provide useful measurements of loss rates. The fraction lost field indicates the percentage of packets lost between the two most recent reports. The cumulative number of packets lost field indicates the total number of packets over the duration of the call. Provided packet loss and delay parameters approximately follow a normal distribution, 95% of all future values will lie within two standard deviations of the loss rate or delay averages. The interval between the reports varies based upon the number of participants. As the number of members' increases, reception information might not be kept for all receivers or the interval between reports could become exceptionally long. It calculates the interval between RTCP reports to be Max (Tmin, n*c).

Tmin is set to 2.5 seconds if the participant has not sent an RTCP packet or 5 seconds if it has already sent an RTCP packet. N corresponds to the number of members, and C is a constant that represents the average RTCP packet sized divided by 75% of the RTCP bandwidth. For purpose of this research, Tmin will be taken as the maximum, which will be 5-second intervals. The worst case scenario for obtaining two consecutive reports would be 10 seconds.
Since it is desirable to keep one-way delay less than 150ms, the RTCP report can be used to help select the appropriate codec to keep the delay to a minimum. The source SSRC_n can compute the Round Trip Time (RTT) propagation delay to the receiver, SSRC_r, by nothing the time, T. when the reception report is received and is local to the router. The sum of the last SR timestamp (LSR) and the delay since last SR (DLSR) is subtracted from the time T. The round trip propagation time is therefore represented as

$$RTT=T-LSR-DLSR \qquad (5.1)$$

This resulting different in time approximates the round trip time delay, as some links experience exceptionally asymmetric delays. The RTT is divided by two to approximate the one-way delay.
Provided RTT/2 is less than 150ms, no codec change is required.
The dynamic codec selection scheme allows for the most optimal codec to be used for the call based on the feedback provided by the RTCP reports that represent current network conditions.
In traditional implementations, calls are configured to begin using a predefined codec. If the network normally has considerable bandwidth available for calls, the predefined codec could be G.711. If the network normally experiences a grate deal of congestion; the codec could be G.729 or either of the G.723.1 codecs. In the first case, without the ability to renegotiate codecs mid call, the call participants would be subject to worse call quality as network conditions worsen since it is using a high bandwidth codec. In the other case calls, using only the quality codecs are not making full use of the available bandwidth and quality. The proposed scheme is designed to ensure calls use the highest quality codec based on network conditions.
Using current implementations, the codec used to initialize the call must be used throughout the call, as no change in codec is permitted.
Take for instance using the G.711 codec for a call –the codec exhibits extremely high call quality when packet loss is minimal. However, once packet loss 3% or higher, the network factor is higher than that of the G.729 codec with 0%packets loss.
Recall G.711 uses 64Kbps and G.729 uses only 8Kbps. If the call is not allowed to renegotiate to codec that requires less bandwidth, packet loss rates will remain at high levels or continue to rise, thus affecting the quality of the call more severely. However, the dynamic codec selection algorithm proposed in this research allows a change in codec to maximize the call quality by using optimal codec.
The proposed algorithms will initiate a codec change request based on feedback received from RTCP reports, as shown by known results for codec with various packet loss rates. G.711 is used as the default codec to begin the call, as it requires the highest amount of bandwidth. Essentially, each codec represents a state, S, and each state has three future states, SF that can be reached based on the feedback received. The future states can either be an increase to a higher bandwidth codec, a decrease to a lower bandwidth codec, or no change. Using the values of G.711 can remain as the codec provided packet loss rates are not greater than 3%. However, if loss rates are higher than 3% and the one way delays are within the bounds of the other codecs. It is possible that using either G.729 or G.723.1 6.3Kbps codec can free up bandwidth to reduce packet loss.
The use of either codec varies based on the one way delay.
Similarly, if the current codec is G.729, the codec has three future states. It can remain as the codec if the packet loss rates are less than or equal to 1% or if the flexibility in delay does not permit the use of a lower bandwidth codec with higher delay. On the other hand, if loss rates are greater than the tolerable 1%, the codec will migrate to

G.723.1 6.3 Kbps codec, provided the one way delay is less than 82.5ms(150-67.5). If packet loss rates exhibit relief, or if the one way delay begins to exceed 150ms, G.711 will be selected as new codec.

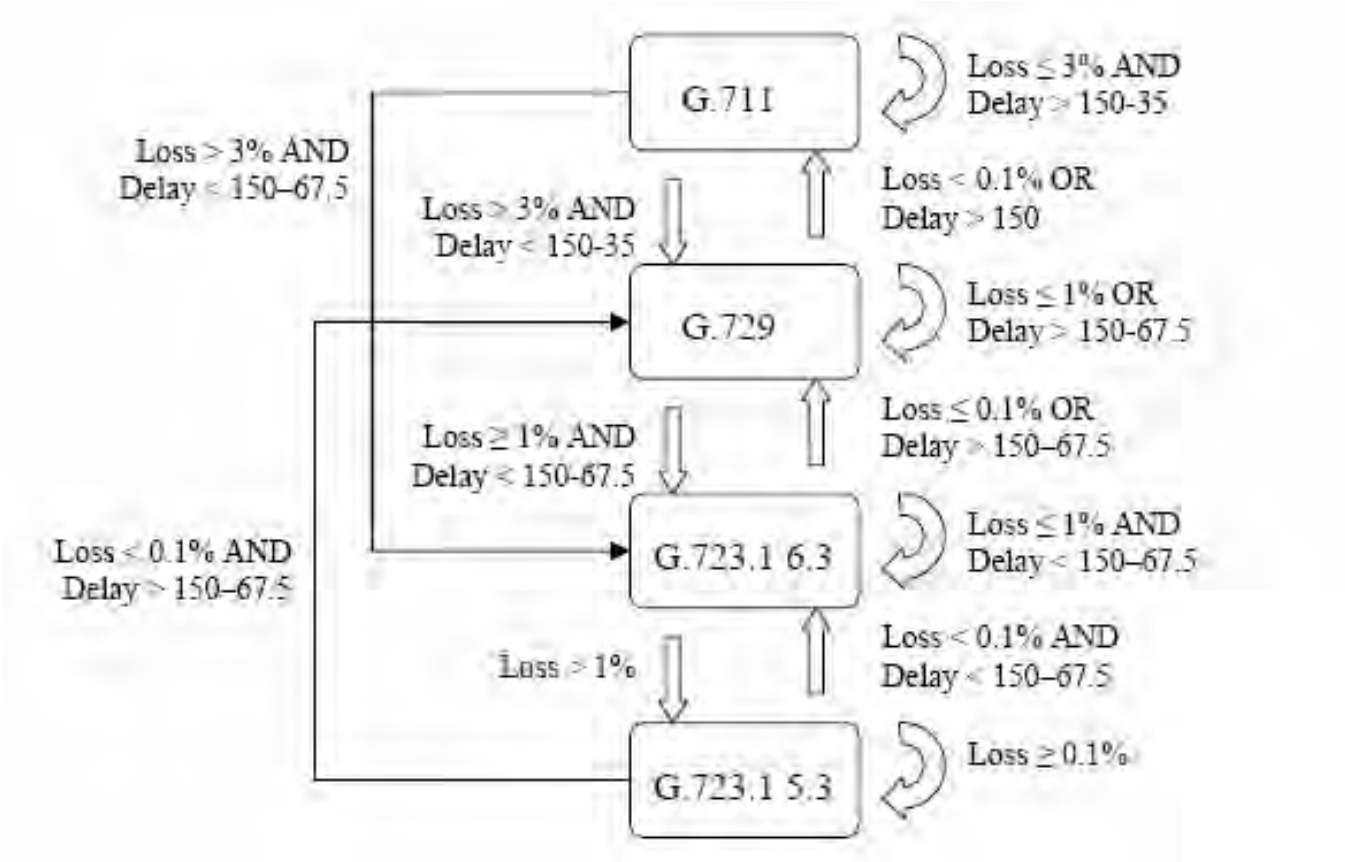


Fig 5.2 Codec selection flow chart

If the current state of the system is using the G.723.1 6.3 Kbps codec, there are also three possibilities for future states. If the packet loss rates remain relatively low, less than 1% and the one-way delay does not exceed the 82.5ms allowed, the system will remain in its current state. However, if loss rate begin to ease and the delay increase above 82.5ms, the system will request a codec change to G.729. Finally, if the loss rates are greater than 1% the system will request a change to the lower bandwidth G.723.1 5.3Kbps codec.

If the system is currently using the lowest possible bandwidth codec, G.723.1 5.3Kbps, and the loss rates are negligible, or less than 0.1%, the system can attempt to increase the quality of the codec used. If the one-way delay is less than 82.5ms, the new requested codec will be G.723.1 6.3 Kbps, where as if the one-way delay is higher than 82.5ms, the new codec will be G.729. If the loss remains high, the G.723.1 5.3Kbps codec will continue to be used.

5.5 Algorithm Simulation discussion

The proposed codec allocation scheme takes into account the dynamic network conditions as obtained from RTCP sender and /or receiver reports and expected call requirements, and it subsequently makes a codec selection based on those statistics. If current network conditions permit the use of higher bandwidth codec selection scheme as we see in fig 5.3 it will select a codec appropriately. However, if insufficient bandwidth is available or the number of calls is too great, the codec selection scheme will only permit lowers bandwidth, and lower quality codecs. Since this method adjusts utilization according to current network parameters, packet loss is likely to be reduced due to not oversubscribing that are more susceptible to packet loss.

Generally, a codec change request is initiated by analyzing parameters such as available bandwidth (BW), packet loss rates, and end to end delay. Due to these the packet loss and the delay are different for different algorithm codecs.

As shown in the figure 5.3 below the sample parameter output is, loss, time (T), last sender report (LSR), delay last sender report (DLSR), loss rate, and one way delay scheme is dispelled on the simulation graph and according to, the codec will be changed.

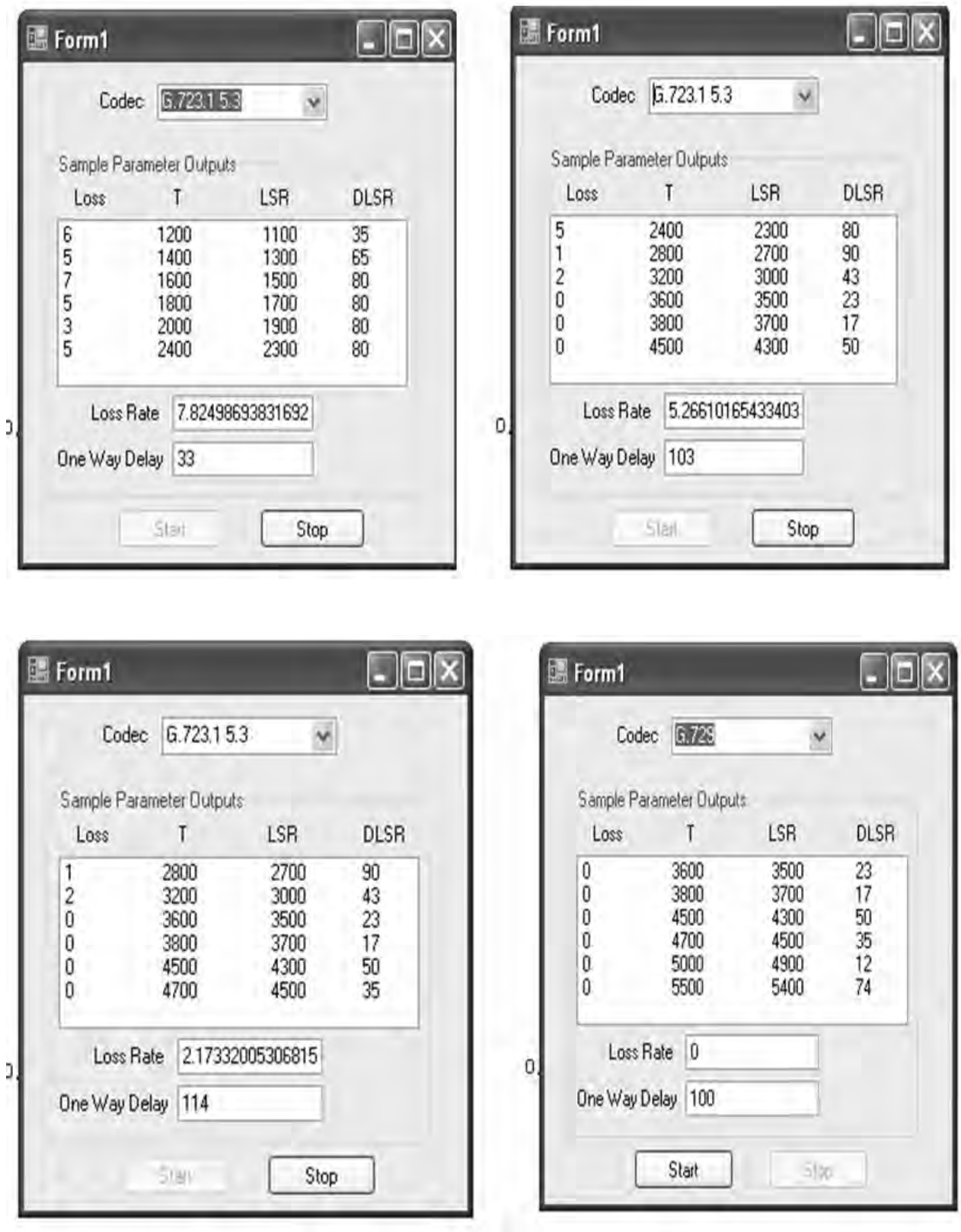


Fig 5.3 Algorithm Simulation Diagrams

5.6 Proposed Handshake Mechanism

The proposed scheme is designed to take advantage of the resource centric method. A codec change request is initiated based upon the parameters determined from that approach. The approach continuously monitors resource availability and the requirements of current calls by analyzing parameters such as available bandwidth, packet loss rate, and overall end-to-end delay. These parameters can be obtained either by the use of Simple Network Management Protocol (SNMP) or by evaluating sender and receiver reports in RTCP, which provide feedback on current network statistics. Codec allocation is therefore determined based upon available resources obtained from network monitoring procedures. The algorithm selects the best codec option based on these available resources while attempting to maximize voice quality. Once the current network resources have been determined and a codec selected, the proposed algorithm provides a means for changing to the appropriate codec in the middle of the call.
Take the following situation for example. There are two H.323 capable stations, A and B. assume the sender always initiates the codec change request and that both the sender and receiver use different codecs for sending. A three way handshake is designed to accomplish mid call codec renegotiation as illustrate in fig 6.3 and explained in the following steps.

1) Station A determines that a codec change is needed. It will send a control packet to B indication which codec it would like to use and continues to send RTP packets with the current codec.

2) Station B receives the change request from A and checks whether it supports the requested codec .If the requested codec is not supported by station B, it replies with a negative acknowledgement (NACK). However, if the new codec is supported, station B replies to A with a message indicating which RTP sequence number will begin with the new codec transmission. The RTP sequence must be calculated at a minimum of 3 round trip times (RTT) from the current point in time, allowing station A to acknowledgement. A minimum of 1.5 RTT's are required to allow the packet to reach the destination and an acknowledgement to return. Therefore, 2xRTTP is used to compensate for any unforeseen delays due to the inconsistency of delays over the Internet.

3) Station A receives the response from B and replies with an ACK to confirm the sequence number or send a more realistic sequenence number back .If the response from B is never received, perhaps because of packet loss; it re-tries the request as in (1).

4) Station B receives the ACK from station A and becomes ready to accept packets with the new codec starting at the previously determined RTP sequence number. All of the packets using the new

codec will then be placed into the same jitter buffer used for the original codec. Packets in the jitter prior to the codec switch will continue to be played out, followed by packets using the new codec. If station B never receives the ACK from A., it re-calculates the RTP sequence number to be used and send a new ACK with that sequence number back to A as in (2).

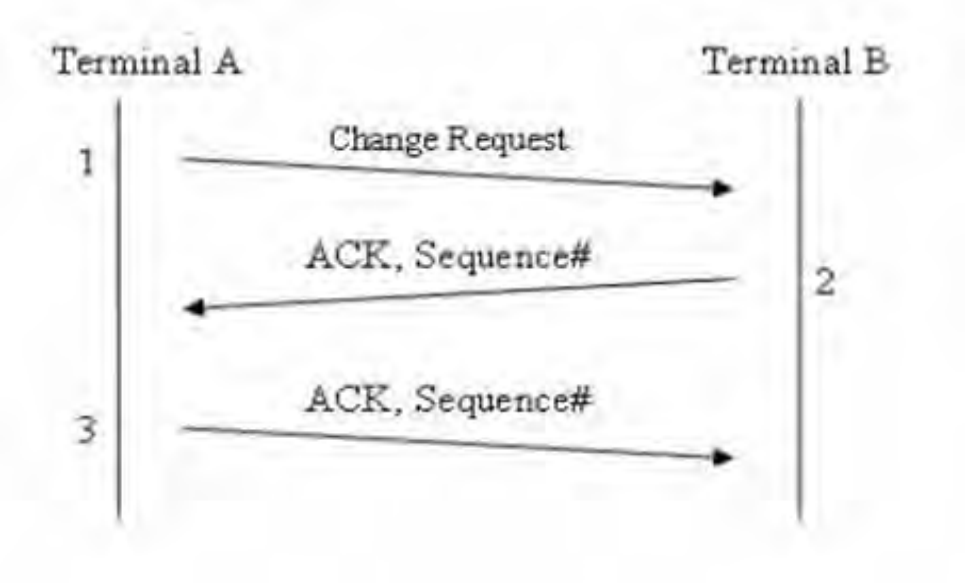


Fig 5.4 proposed three way handshake mechanisms for codec change

Chapter Six

6. Conclusion and Recommendation

6.1 Conclusion

The addition of VoIP to the data networks adds risks to the security; however, through careful and appropriate use of security mechanisms, security can be maintained. VoIP is gaining an edge and becoming mainstream. New issues continue to arise with security; however, the future of VoIP seems promising.

As we have seen in the simulation result different amounts of traffic is received from different networks. More traffic can be arrived at the destination using No-Firewall, while medium traffic is arrived at the end users when the network becomes Firewall_VPN, but less traffic is received using security mechanisms.
On the other hand the result shows more time is needed when the network uses security mechanisms to deliver the voice to the receiving stations, and less amount of time is required when there is no security mechanisms is applied on the network to deliver the information to the end users.

Generally, the simulation result indicates that there is more packet delay, packet loss and delay variation because of using security mechanisms, especially when the distance is increased the traffic received becomes more decreases, because of the delay and the packet loss is increased, the quality of voice becomes more degrade.
Due to this drawback, more VoIP developments and security solutions are anticipated in the near future as its prevalence increases in the telecom industry because the existing security mechanisms like Firewall /NAT, and traffic encryption /decryption are the greatest contributors to network congestion, jitter and throughput delay. These security mechanisms are
Which is a latency producers and network congestion generators, but they have effective means to secure a network. This has the “good news” and “bad news”

The proposed codec allocation scheme takes into account the dynamic network conditions as obtained from RTCP sender and /or receiver reports and expected call requirements, and it subsequently makes a codec selection based on those statistics. If current network conditions permit the use of higher bandwidth codec selection scheme it will select a codec appropriately. However, if insufficient bandwidth is available or the number of calls is too great, the codec selection scheme will only permit lowers bandwidth, lower quality codecs. Since this method adjusts utilization according to current network parameters, packet loss is likely to be reduced due to not oversubscribing that are more susceptible to packet loss.

Since a codec change request is initiated by analyzing parameters such as available band width (BW), packet loss rates, and end to end delay. Due to this the packet loss and the delay are different for different algorithm codecs.

So the proposed adaptive codec selection mechanism ensures voice continuity while switching codecs by filling the play out buffers appropriately. Actually the proposed idea appears theoretically feasible, but it needs to be practically implemented and tested in a production level environment by using E-Model.

6.2 Recommendation

This research explored changing a codec unidirectional, so bidirectional codec re-negotiation should be investigated for synchronization between both end points

The effect of the proposed an adaptive codec selection algorithm mechanism on voice over IP quality should be determined by using different speech quality measurement or by using E-model, like mean opinion score (MOS), comparison mean opinion score (CMOS), and degradation mean opinion score (DMOS).

Finally, designing an adaptive playout buffer algorithm since fixed playout time set-up/scheduling is simply, but normally causes a constant delay and cannot follow the change of network delays. Adaptive playout scheduling should introduce to overcome these problems, and is controlled by corresponding playout buffer algorithm, which can utilize the silence time between two successive voice periods, referred to as talkspurt, to slow down or speed up playout time of each talkspurt.

References

[1] V. Paxson, “Measurements and Analysis of End-to-End Internet Dynamics”, PhD Thesis, Computer Science Division, University of California, Berkeley, 1997.

[2] Shu Tao, Kuai Xu, Antonio Estepa, Teng Fei, Lixin Gao, Roch Guerin, Jim Kurose, Don Towsley, Zhi-Li Zhang. “Improving VoIP Quality through Path Switching”.

[3] Micheal Manousos, Sypros Apostolacos, loannis Grammmatikakis, Dimitrios Mexis, Dimitrios Kagklis, Efstathois Sykas. “Voice-Quality Monitoring and Control for VoIP.”

[4] Victor S. Frost. “Quantifying the Temporal Characteristics of Network Congestion Events for Multimedia Services”, IEEE Transactions on Multimedia, September 2003

[5] Pedro B. Velloso, Otto Carlos M.B.Duarte, Marcelo G. Rubinstein. “Analyzing voice transmission capacity on ad hoc networks.”

[6] Anna Watson, M. Angela Sasse. “Measuring Perceived Quality of Speech and Video in Multimedia Conferencing Applications”, ACM Multimedia’98, Bristol, UK.

[7] Cole, R.G. and Rosenbluth, J.H. Voice Over IP Performance Monitoring, ACM SIGCOMM, Computer Communication Review,

[8] Bruner, S. & Ahklaq, A. (2004).Voice Over IP 101: Understanding VoIP Networks. Retrieved March, 2006 at URL:http://www.juniper.net/solutions/literature/white_papers/200087.pdf-jitter

[9] Alberts, Christopher. , et al. (2002). Managing Information Security Risks. Pearson Education Inc.

[10] Thermos, P. (2006). Two attacks against VoIP. Retrieved April, 2006 at http://www.securityfocus.com/infocus/1862/1.

[11] Schneier, B. (1996). Applied Cryptography (2nd Ed). John Wiley & Sons, Inc, 1996,

[12] Thermos, P. (2006). Two attacks against VoIP. Retrieved April, 2006 at http://www.securityfocus.com/infocus/1862/1.

[13] M. Hassan et al., “Internet Telephony: Services, Technical Challenges, and Products,” IEEE Communications, Apr 2000.

[14] A.Thom, “H.323: The Multimedia Communications Standard for Local Area Networks,” IEEE Communications, Dec 1996.

[15] Schulzrinne and J. Rosenberg, “The Session Initiation Protocol: Internet- Centric Signaling,” IEEE

[16] H. Schulzrinne et al., “SIP: Session Initiation Protocol,” IETF RFC 2543, Mar 1999.

[17] M. Handley and V. Jacobson, “SDP: Session Description Protocol,” IETF RFC 2327, Apr 1 9 9 8.

[18] R. Atkinson and S.Kent, “Security Architecture for Internet Protocol,” IETF R F C 2 4 0 1, N o v 1 9 9 8.
[19] R. Atkinson and S. Kent, “IP Authentication Header,” IETF RFC 2402, November 1998.
[20] Carrel and D. Harkins, “The Internet Key Exchange (IKE),” IETF RFC 2409, Nov 1998
[21] Maughan et al., “Internet Security Association and Key Management Protocol (ISAKMP),” IETF RFC 2408, Nov 1998
[22] Sue B. Moon, Jim Kurose, and Don Towsley, “Packet audio playout delay adjustment: performance bounds and algorithms”, Multimedia Systems (1998) 6:17-28.
[23] Mona Habib and Nirmala Bulusu, “Improving QoS of VoIP over WLAN (IQ-VW)”, Project Research Paper, for CS522 Computer Communications, University of Colorado at Colorado Springs, December 2002.
[24] N. Shacham and P. Mc Kenney. Packet recovery in high-speed networks using coding and buffer management. In Proceedings of IEEE Infocom, volume 1, pages 124–131, June 1990.
[25] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A transport protocol for Real-Time Applications. INTERNET-DRAFT - draft-ietf-avt-rtp-new-08.ps, July 2000.
[26] Nobuhiko Kiatawaki and Kenzo Itoh. Pure delay effect on speech quality in telecommunications. IEEE Journal on Selected Areas in Communications,9(4):586–593, May 1991.
[27] Wcuyu Jiang and Henning Schulzrinne. Modeling of packet loss and delay and their effect on real-time multimedia service quality. In Proceedings of ACM NOSSDAV,
Jun 2000.
[28] W. Jiang and H. Schulzrinne. Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss. In Proceedings of ACM NOSSDAV, pages 73–81, May 2002.
[28] Bolot, J. C. (1993) End-to-end Packet Delay and Loss Behaviour in the Internet. Proc. 1993 ACM SIGCOMM Conf. pp.289-298.
[29] One-way transmission time. ITU-T Recommendation G.114, February 1996.
[30] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson,” RTP: A Transport Protocol for Real-Time Applications,” IETF RFC 3550, July 2003,http://www.ietf.org/rfc/rfc3550.txt
[31] Scott Keagy, Integrating Voice and Data Networks,
Cisco Press, Aug.2001.
[32]Assessing VOIP Call Quality Using the E-Model,”IXIA, 2005 Http://www.ixiacom.com/pdfs/library/white papers/voip/quality.pdf

[33] J.-C. Bolot. End-to-end packet delay and loss behavior in the internet. In Proceedings of ACM Sigcomm, pages 189–199, August 1993. San Francisco, CA.
[34] J-C. Bolot and A. Vega Garcia. The case for FEC-based error control for packet audio in the internet. In to appear in ACM Multimedia Systems [20].

[35] S. Floyd. RED: Discussions of setting parameters. http://www.icir.org/floyd/REDparameters.txt, November 1997.

[36] S. Floyd, R. Gummadi, and S. Shenker. Adaptive RED: An algorithm for increasing the robustness of RED. Under submission, August 2001.