Masters Thesis

VoIP in the Context of Security

Fiifi Botwe Arkaah

Department of Computer and Systems Sciences

Stockholm University / Royal Institute of Technology

April, 2006

This thesis corresponds to 20 weeks of full-time work of the author

The wide spread deployment of enterprise networks and its capacity to carry voice traffic over the internet has spear-headed the march forward for IP-telephony or voice over internet protocol (VoIP). Unfortunately, the internet which is based on IPv4 public networks was not built to anticipate the security problems that could be passed on to this evolving technology. Moreover, VoIP has been challenged to provide better quality of service (QoS) in addition to its much hyped or promised value added services. A much better way to mitigate these IPv4 network security issues is to adopt a defence-in-depth approach which would include a migration to a less faulty network service protocol such as IPv6 and the separation of voice, signalling and data services into different segments in a single network. An IPv6 deployment will offer mandatory confidentiality, authentication, integrity and non-repudiation services with the help of the internet security protocol IPSec which properly addresses the issue of message origin in IPv4. Moreover, the separation of VoIP critical services into different layers would help allocate enough resources based on priority which in turn will improve QoS and security. Currently, the technology used in achieving QoS and security involves Ethernet switch based IP networks which support virtual local area network (VLAN) standards. In discussing the mentioned issues, this thesis will focus on security in VoIP especially in the context of SIP (Internet Engineering Task Force's recommendation for signalling in multimedia communication).

Contents

1 Introduction

1.1 Background

Traditional or yesterday's Plain Old Telephone Systems (POTS) were based on circuit switched technologies where, phone calls were conducted through a Public Branch Exchange (PBX) [48] and a Public Switched Telephone Network (PSTN) [4]. In circuit-switched technology, a call session is allotted a specific channel or a slot for conversation through a process called Time Division Multiplexing (TDM) [6]. The call is then communicated to its destination by a PBX through a PSTN.

Thanks to the power of the Internet, circuit switched technology is expected to give way to a system based on packet-switched technology called Voice over Internet Protocol (VoIP). VoIP technology is intended to carry voice, video and data in a digitized form through an IP network in real-time manner with the aim of cutting cost on infrastructural investment and making long distance calls cheaper. Traditional circuit-switched technologies are heavily dominated by Signal system number 7 (SS7) [3] for call establishment and termination. VoIP calls on the other hand are characterised by a stream of voice packets carried by a real time protocol and processed or regulated by standards based multimedia signalling protocols such as session initiation protocol (SIP) and H.323 [42] or proprietary based protocols like the one supported by the popular free peer-to-peer internet soft phone *Skype*. Other VoIP applications supporting SIP and H.323 include the likes of *Windows Messenger* and *Netmeeting* respectively [62].

When such unprotected calls or packet streams are routed over an insecure or hostile medium such as the internet, VoIP systems falls prey to call eaves-dropping, unauthorized free calls resulting from compromised gateways and call hijacking due to disruption of service or signal jamming due to denial-of-service (DoS) [57] attacks. Tackling these pertinent security issues contributed by weaknesses such as TCP session establishment requires a multi-dimensional approach. Such approach could begin with the securing of the entire enterprise network with a gateway firewall to regulate out-bound and in-bound traffic with filtering rules normally based on packet header information. In response to the concern of multiple openings of ports required to process calls on a larger network scale, dynamic filtering firewall [9] architecture provides a mechanism for understanding VoIP signalling enough to open and close ports after call sessions are completed. Security of packets beyond firewalls or gateways can be improved by deploying a virtual private network (VPN) [45] based on the internet security protocol IPSec [58] to provide a secured channel over the internet. IPSec deployment becomes helpful in eliminating packets from unknown sources by mostly establishing direct links between known parties. In a typical VoIP VPN deployment, a company with branches can set up a secure channel or a tunnel between branch gateways.

The next level of protection provides voice and signalling packet streams with confidentiality, integrity and authentication. Signalling packets which are used to setup, control and tear down a call session are secured by the transport layer security (TLS) protocol. TLS [30] ensures session privacy using public and symmetric key encryption and packet integrity using message authentication codes (MAC). A key exchange mechanism based on public key infrastructure in TLS presents communicating peers the opportunity for mutual authentication with the help of digital certificates which in turn provides sufficient evidence for non-repudiation.

The voice stream is basically RTP [29] packets directly exchanged between the sender and receiver. This stream is protected by providing payload (voice) encryption and packet integrity using the secure real-time protocol (SRTP). In addition to providing unique encryption to each voice packet, SRTP [31] tackles crucial voice attacks such as replay attacks with a replay list which helps receivers in a communication to verify whether a packet has already been played or not. The key derivation process in SRTP improves security by enforcing re-keying or refreshing of cryptographic keys to prevent attackers from making effective use of compromised keys.

In the case where the gateway hosts in an enterprise network risk exposing the entire network upon compromise, it will be helpful to partition the internal network into layers. Layers could be divided depending on whether they carry less or more sensitive packets. More sensitive layers are given more quality-of-service (QoS) attention such as more bandwidth allocation than their less security demanding counter-parts. Network partitioning can be efficiently achieved using switches which support virtual local area network (VLAN) and IEEE 802.1Q [36] technology. Finally, staff education on the secured use and operation of VoIP equipments is crucial in improving secure communication. Such practice will be instrumental in branding a VoIP infrastructure as having a holistic sense of security.

1.2 Goal

Even though VoIP revolution comes with its blessings in a form a more efficient bandwidth usage than traditional systems, it certainly has its security and service implications. Since VoIP is based on the internet, it implies it would inherit the security problems that confront IP data networks -making some technology experts and vendors raise fears about security of conversations in VoIP deployment. Moreover, they question the quality of service of VoIP and claim that it is inferior to PSTN systems. US based surveys (2006) such as the *ZEDNet* [8] have been quick to react to these concerns about VoIP. However, the same survey overwhelmingly points to the upward trend of VoIP to the extent of becoming the dominant force in the telecom industry.

The scope of this thesis encompasses an overview of the core functions, weaknesses, and performance issues of the underlying network foundation of VoIP. This paves the way for a comprehensive review of the current attempts at providing security in a VoIP infrastructure which also extends to include an overview of associated benefits and challenges.

1.3 Purpose

This thesis should serve as a useful document of reference for middle management, VoIP enthusiast and network administrators and designers who know little about security especially in the context of VoIP. In this regard, I am referring to developing countries with poor telecom infrastructure especially third world countries that would want to catch up with western advancement in telecommunication and security.

1.4 Methodology

The VoIP security and service issues will be qualitatively analysed with the help of secondary data sources such as company annual reports, focus group discussions and literature surveys like white papers, IT magazines and request for comments (RFC). In other words, the thesis approach to help reach a verdict on the security issues related to VoIP will be purely inductive and based on extensive literature.

1.5 Structure of Thesis

This thesis begins with an overview of internetworking protocols and related vulnerabilities. This is followed by a description of the problems facing IP-telephony network infrastructure and corresponding security solutions to them. I then touch on the benefits and challenges confronting voice-over-IP deployment in a converged network. This thesis concludes with suggestions or proposals of additional security strategies to complement the discussed security mechanism which is then wrapped-up with a brief discussion and suggestions for further studies.

2 Internetworking Services

The way to efficiently deliver information – be it data, voice or video, from one destination to the other through a network has become the bone of contention between the old and new telecommunication guards. The old guards in a sense are the pioneers and designers of the old traditional voice telephone network who proposed a dedicated channel to each user on a network. They are linked with such old technologies such as POTS (plain old telephone system), PSTN, and TDM –all of which are referred to as circuit-switched technologies.

With the rise of information technology or the dot com bubble brought into prominence the new guard. They were mostly championed by Universities such as MIT and Berkeley, IT companies such as Xerox, Digital and Cisco, backed by standards body such as the Internet Engineering Task Force (IETF) and Internet service providers. Basically, they support information delivery based on breaking data into smaller chunks and letting the network decide how best to reach its destination- a technology generally referred to as packet switching. However, carrying out this philosophy has also meant borrowing from the old ones in the form of virtual circuit switching which sees packet switching technologies behaving as a circuit-switch.

In this chapter, I will briefly touch on the mode of operation circuit-switched networks which are categorically connection oriented in the kind of services it provides. I will then focus on the overview on packet-switched technologies which embrace both connection and connectionless worlds and highlight the weaknesses they contribute towards a network system's vulnerability.

2.1 Types of Services

2.1.1 Connection Oriented Services

Circuit switching technologies are normally connection oriented meaning that, before transmission of data a connection path is first determined and established for sequential delivery. A pioneer of this service is the PSTN which in a way can be viewed as "the concentration of the world's public circuit-switched telephone networks, in much the same way that the Internet is the concentration of the world's public IP-based packet-switched networks" [4]. Fig. 2-1 illustrates the present digital state of PSTN which is a leap from an analogue system

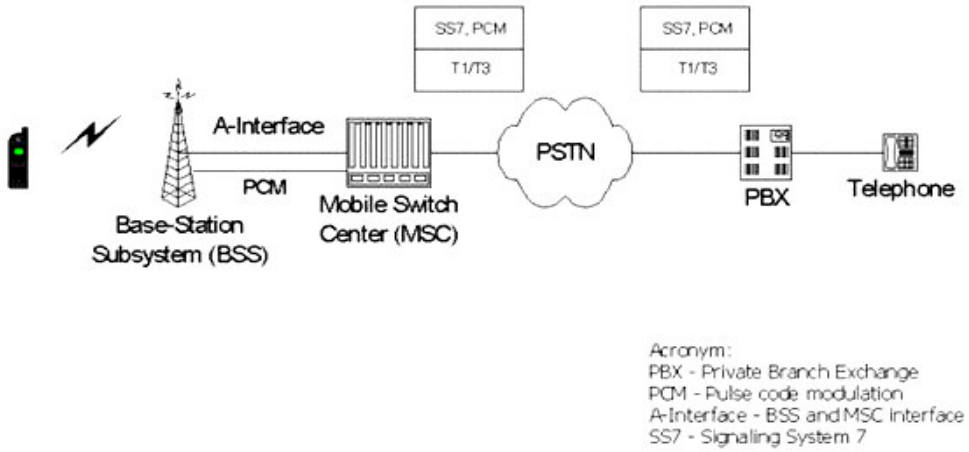


Figure 2-1: Mobile and fixed networking using PSTN [1].

Currently, PSTN is used to support and service both fixed and mobile terminals. To achieve the "connection oriented" reputation, PSTN networks uses a Public Branch exchange-PBX [48] as a switch and a coordinating mechanism normally provided by a signalling protocols like the signalling system 7 (SS7) [3]. The switch is used to determine the path of the communication with the help of the signalling protocol serving as a connection controller. In a typical call scenario, the caller's phone sends a signal using SS7 to the PBX which in turn relays the signal to the callee for communication to commence. After the connection is established, the speaker's voice, which is analogue, is digitized by a mechanism based on pulse code modulation- PCM [5] and sent through a dedicated full duplex circuit channel of 128kbps capacity. This means that, in a telephone conversation, each party is allotted 64kbps bandwidth.

PSTN uses a technology called time division multiplexing (TDM) [6] to handle the numerous calls that come through the network. TDM allocates a particular single channel to each caller at a time. This is normally referred to as time slicing. In a case where all available slots or channels are engaged, a new call will return an engage message or placed in queue until an existing user finishes his call session and frees a channel slot.

2.1.2 Connection-Less Oriented services

Packet switching technologies like the one the internet is based on are connection-less oriented. Applications used in such technologies don't normally require real-time processing of data and besides can tolerate some delay in transmission of information such as e-mail messaging and file transfer. In such a case, packet-switching is said to be offering an unreliable or datagram service [22].

Interestingly, packet switching does behave like circuit-switching when data transmission is done over dedicated circuit paths between the communicating parties using a special unique call identity number basically called virtual circuit identifier (VCI) [22]. This kind of service requires only the VCI for transmission from source to destination unlike a datagram service where relatively lengthy address formats are used in identifying communication data.

A typical example of how virtual circuits works is illustrated by *X.25* [53] which is an International Telecommunication Union-Telecommunication (ITU-T) protocol standard for WAN (Wide Area network) communications. In fig. 2-2, data terminal equipment (DTE) devices are end systems that represents computer terminals or network hosts communicating across an *X.25* network. Data circuit-terminating equipments (DCE) are devices such as modems and packet switches basically responsible for carrying data from one DTE to the other in a sequential manner after session establishment.

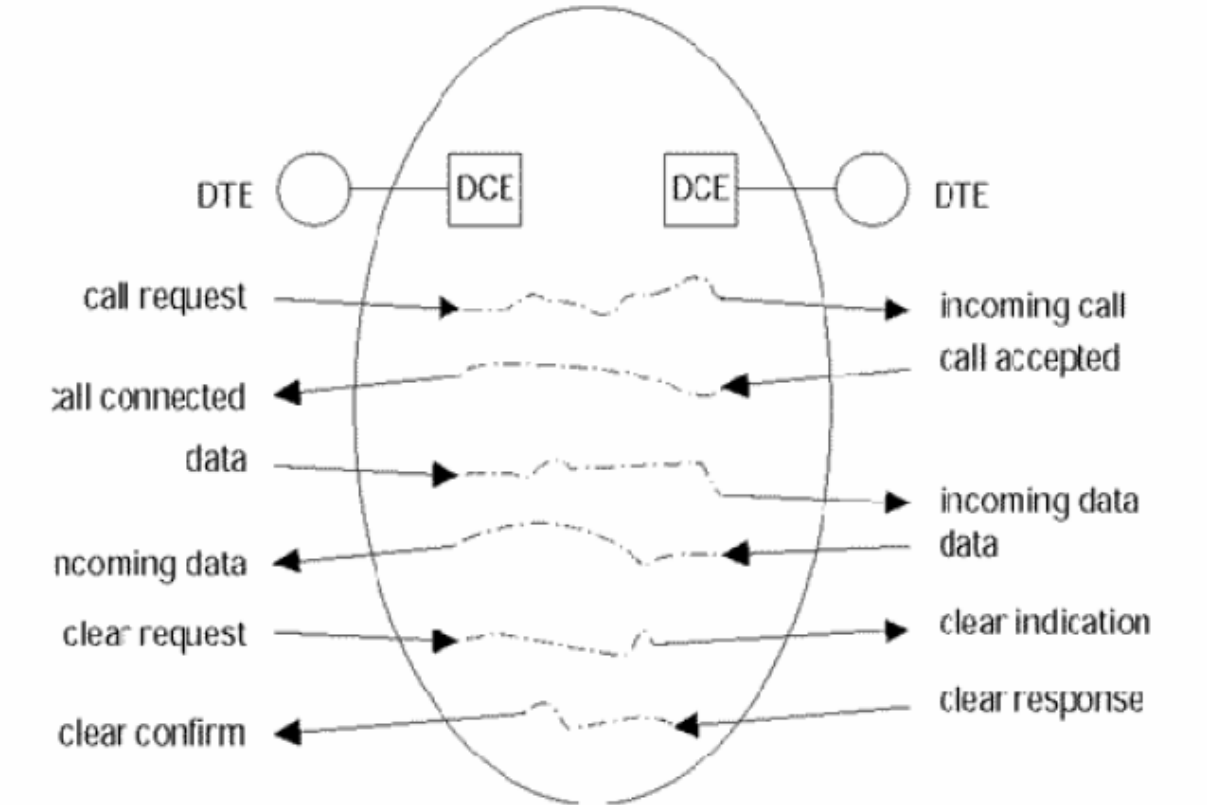


Figure 2-2: A summary of X.25 call flow [22]

A sharp contrast that can be drawn from the illustration (figure 2-2) compared to the likes of internet based IP network is that, an orderly arrival of data is not a strength of datagram services. This is because packets or data are allowed to choose their different preferred paths to their destination for reassembly into original forms - a process constrained by delay. This mechanism makes datagram services less reliable than their circuit switched counter-parts. However, the network architecture of packet switched technologies are layer based in such a way that, by adding a reliable service on top of a datagram service, a connection oriented and reliable service can be achieved.

The next section introduces how packet switched services are organized into stacks using the generic OSI (Open System Interconnect) reference model and the internet specific TCP/IP protocol suite.

2.2 Network Architecture

The mechanism by which for instance, an e-mail message is sent from one computer to another on any platform has been modelled and implemented into layers that provide services to an application. This section looks into two models namely the Open systems interconnect – OSI model and the internet's TCP/IP model. OSI promotes a generic architectural guideline or reference for network implementations of layer protocols. The TCP/IP model on the other hand is a summarized format of the OSI model that is adopted by the internet [45].

2.2.1 OSI reference Model

The OSI model is a standard defined by the ISO 7498-1 which consist of seven layers of services forming a protocol stack. The protocol stack can be divided into the lower layers which are normally implemented in hardware and the upper layer which are implemented in software. Fig. 2-3 demonstrates the OSI model.

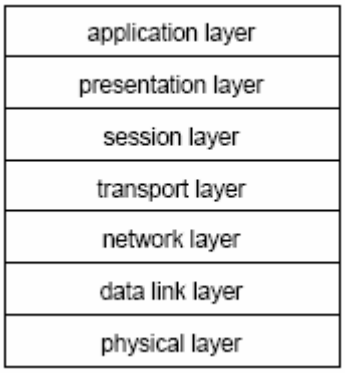


Figure 2-3: Layers of the OSI network model

The first layer counting up the stack is the physical layer. This layer carries information in the form of bit stream through hardware means such as cabling and network interface cards.
Next on the stack is the data-link layer responsible for encoding and decoding of packet data into raw bits over communication links. The third is the network layer and it provides path determination of packet transmission between end-systems – a feature also referred to as switching or routing.

The fourth layer is the transport layer which is normally responsible for establishing a reliable connection between end nodes. This also includes a mechanism for end-to-end error recovery and data flow control. The session layer takes care of setting up, the maintenance and termination of communication between applications. The presentation layer, also called the syntax layer is the sixth on the stack. It provides services to convert data from network format to application readable formats and vice versa such as encryption and compression. The last layer which is the application layer is concerned with network application and end-user interactions. Activities in the application layer ranges from user authentication to file transfer and e-mailing.

2.2.2 TCP/IP Model

The TCP/IP model (fig. 2-4) unlike the OSI model has only 4 layers. Some features in the OSI model have been either left out or merged to form one layer of service. In the TCP/IP model, the physical and link layer functionality operate as one in the first layer. Layer 2, which is the network layer deals with addressing, routing and communication services normally left to the IP and ICMP protocols. Layer 3 provides transport communication between applications using either Transmission control protocol –TCP for reliable connection or unigram data protocol for datagram services. The functionality offered at the application, network and transport layer are no much different from their equivalent layers of the OSI model.

The terminology used to refer to data moving down or up the stack can be a source of confusion sometimes. Data handled in the transport layer is referred to as a segment. At the network layer, data becomes an IP datagram and between the network and link layer, IP datagram is referred to as packet. Packets are finally packaged into a frame format before they are transmitted to another host's stack.

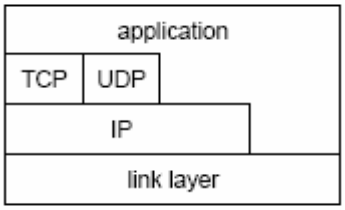


Figure 2-4: TCP/IP protocol stack

2.3 Internet Protocol – IPv4

The internet protocol in its version 4 is a standard defined by RFC 791 [25]. It operates at the networking layer of the TCP/IP protocol suite. The services offered by IP is basically datagram which means that, packets heading for a destination will not be guaranteed safe delivery or arrival in the same order they were sent. This is because, IP lacks on its own, the mechanism to recover lost packets or control flow of packets in an orderly manner. In other words, at best, IP offers "best effort" delivery service [22]. Routers, a layer three device, are used to determine how packets independently reach their destination by using information found in an IP header illustrated in fig. 2-5.

2.3.1 IP Header

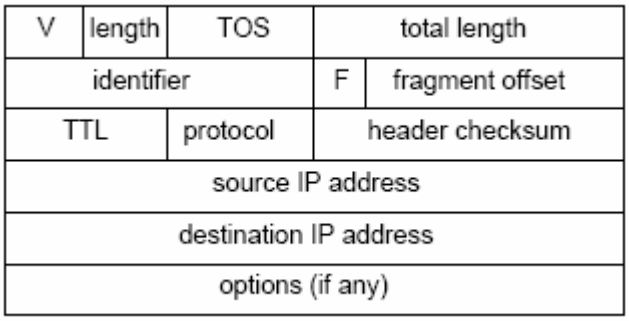


Figure 2-5: IP header packet format

IP, with the help of header information, uses two kinds of schemes to send datagram from source to destination. They are addressing and fragmentation. Both aid every single packet in choosing different paths to their destination resulting in delay or out-of-order packets sometimes.

Addressing for a datagram is provided by the source and destination IP address fields in the IP header. IP addresses are four 8 bit decimal numbers separated by dots for example 192.189.6.2. IP addresses can be categorized into five different classes ranging from A to E. The different classes are dictated by the first four most significant bits of the first octet in the address illustrated in Fig 2-6. Addresses further reveal two levels of information. One part points to network address and the other represent the point to which a host or machine is attached to, on a network.

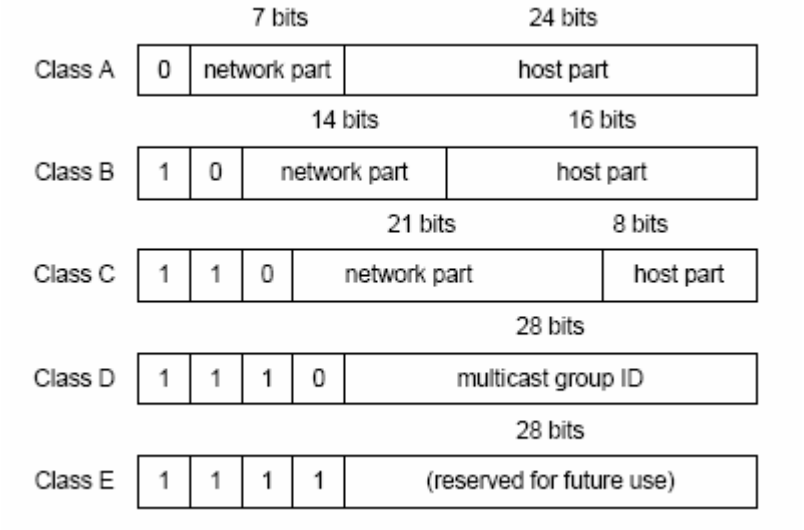


Figure 2-6: IP network address classes

2.3.2 Fragmentation

The maximum size of data or frame passing through a network can be restricted by the network link's maximum transmission unit (MTU). When frames larger than the required MTU attempt to pass through a network, IP relies on a mechanism called fragmentation to break the frame into smaller fragments to be sent separately. The fragmentation process illustrated in figure 2-7 and figure 2-8 uses the *MTU* and the size of the header to compute the required number of fragments. The header of the original datagram is attached to each of the new fragments. The total length field's value is now changed to reflect the new lengths of the smaller fragments. When packets arrive at the destination, they are reassembled with the help of the IP header's flag, identification and fragment offset field information. The flag field contains 3 bits which help determine whether a packet will undergo fragmentation or not. This is done by either setting the *DF* bit to mean no fragmentation or, setting the *MF* bit to indicate more fragments. The identification field helps group or differentiate one original packet from another. Finally, the right order of fragments in the original datagram is determined by the fragment offset field.

Figure 2-7: Links with different MTUs [2]

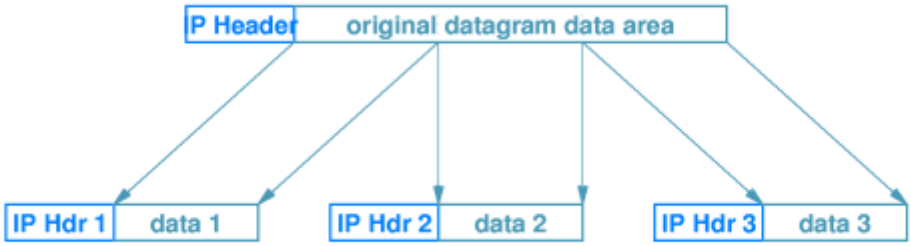


Figure 2-8: Fragmentation process of packets larger than MTU [2]

2.3.3 Encapsulation

The way in which data sent from the application layer is processed through the levels of a protocol stack into an (Ethernet) frame is called encapsulation. Hosts communicating over a network send and receive data in frame formats. The whole process starts when the application lying on top of the TCP/IP stack sends data which is also known as payload, to be first processed by the transport layer which is TCP. The processing done by the TCP service appends transport layer information called TCP header and forwards the data packet which is at this stage called a *segment*, to the IP layer. At the IP layer, the TCP *segment* is processed into an IP datagram by adding an IP header information.

The link layer finally uses the Ethernet protocol to process the datagram into a frame which is then sent to its destination host. The application has the option to use the session oriented TCP or the just the connection-less oriented IP layer. At the destination host, the frames arrive at the link layer and goes through a reverse process of de-capsulation. The Ethernet, IP and TCP headers are pealed of the frame at their various layers of the stack leaving only the payload for the application's use. In the case where the length of the frame exceeds the maximum transmission unit (MTU) of a link, the frame is split into fragments before reassembly at their destination host.

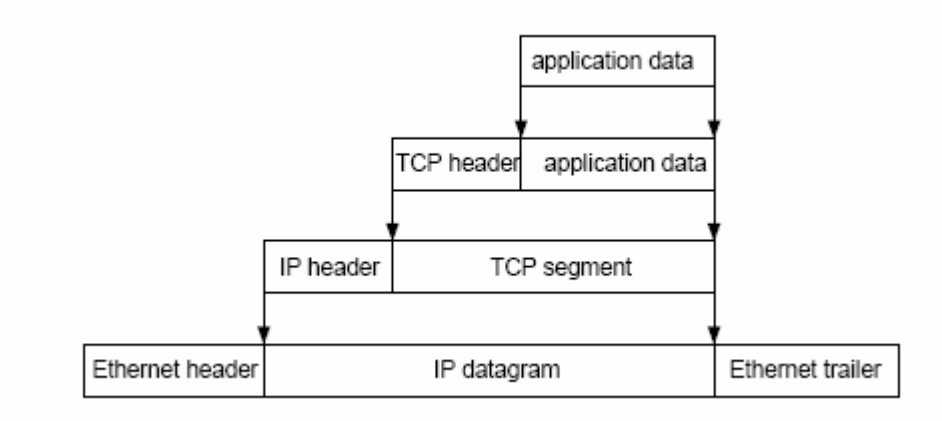


Figure 2-9: Encapsulation of application data into Ethernet frame

2.4 Transmission Control Protocol - TCP

TCP is a session transport layer protocol defined by RFC 793 [26]. When TCP runs over connection-less network layer like IP, it provides a reliable and connection oriented service between end-nodes. A typical TCP connection involves the port number and IP addresses of the sender and receiver machine.

2.4.1 TCP Header

source port number			destination port number
sequence number			
acknowledgement number			
length	reserved	flags	window size
TCP checksum			urgent pointer
options (if any)			

Figure 2-10: TCP packet header

The TCP header which is attached to a segment contains the source and destination ports. The sequence number represents a random 32 bit counter starting from the first chunk of data placed after the header. The acknowledgement number provides information to the sender of data confirming the receipt of *acknowledgement -1* size of data. The flow of segments is regulated by providing the maximum size of data the receiver can accept at a time. This is normally called the window size of the recipient.

Six flag bits are used to determine what kind of segment is transported. They could be either *URG*, *ACK*, *PSH*, *RST*, *SYN*, or *FIN*. Each bit functions as follows:

- Urgent (URG): It enables urgent mode. One end signals the other end that some form of urgent data has been placed into the stream.
- Acknowledge (ACK): It indicates that the acknowledgement number contained in a TCP packet is valid.
- Push (PSH): It asks the receiver to pass the data to the application as soon as possible.
- Reset (RST): It resets the connection.
- Synchronize (SYN): It is used during the connection establishment to synchronize sequence numbers.
- Finished (FIN): Its sender indicates that it wants to terminate the connection.

2.4.2 Three way Handshake

The *ACK*, *SYN*, and *FIN* bits when set, in the flags are used by TCP to establish a session using a three-way handshake process to establish and close transport connections. Fig. 2-11 illustrates a time line instance of the handshake process during connection setup. The process starts when the sender, *A*, initiates a connection to *B* sending a message whose *SYN* bits is set 1. *B* then sends an acknowledgement message with the *ACK* and *SYN* bit set. *A* finally sends an acknowledgement to confirm receipt of B's message.

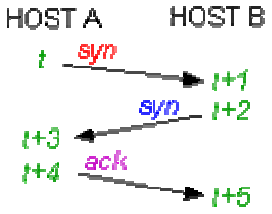
Time	Event	DIAGRAM
t	**Host A sends** a TCP **SYN**chronize packet to Host B	
t+1	Host B receives A's **SYN**	
t+2	**Host B sends** it's own **SYN**chronize	
t+3	Host A receives B's **SYN**	
t+4	**Host A sends ACK**nowledge	
t+5	Host B receives **ACK**. ***TCP connection is established.***	

Figure 2-11: Connection setup in three-way Handshake process [63]

During the connection setup stage, the segments exchanged between host *A* and *B* has a special sequence number referred to as the initial sequence number (ISN). After the connection setup, segments sent to and fro have their *SYN* bit cleared or set to zero and carry a sequence number.

Closing a connection depicted in Fig. 2-12 is controlled by the *FIN* and *ACK* bits. The connection release process is initiated when, the sender, A (Client), sends a FIN message to B (Server) signalling B of A's intention to close connection. B then replies with an *ACK* and *FIN* message acknowledging the receipt of the *FIN* message and in turn, also signalling *A* of its intension to release the release the connection too. Finally, *A* replies with an *ACK* message to bring the session to an end.

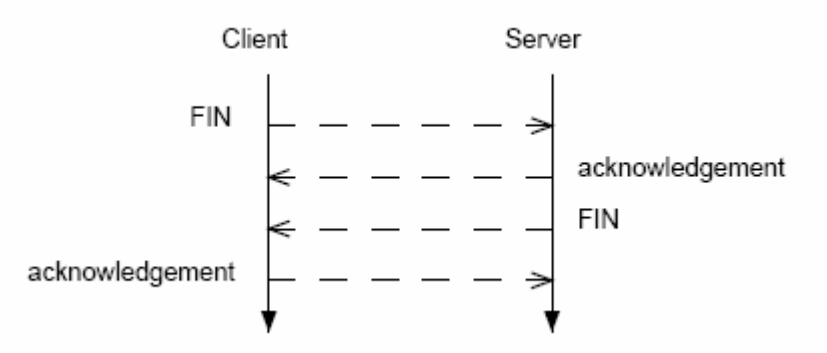


Figure 2-12: Time line termination of connection

2.5 User datagram Protocol

UDP is the optional transport layer protocol in the TCP/IP protocol suite defined by RFC 768 [27]. Together with TCP, they both provide IP layer with transport service. Unlike TCP, the service offered by the UDP is limited. In that, it does not guarantee delivery and sequencing of data. Besides, it does not add much assistance to IP since IP already offers "best effort" delivery services. However, the strength of UDP lies in its simplicity and less overhead due to a smaller header size in addition to fact that it is session-less.

The UDP header illustrated in Fig. 2-13 has four fields totalling 8 bytes as compared to TCP's 20 bytes. Its make up include 2 bytes of source and destination port number, a 2 byte packet length and a 2 byte optional checksum value for error detection. The light weight nature of UDP proves very useful in terms of throughput especially in applications that require real-time processing and can also withstand some level of data loss.

source port	destination port
length	checksum

Figure 2-13: UDP packet header

2.6 Security Issues in TCP/IP

The success of the internet has been built on the pillars of TCP and IP protocols. The growth and evolution of the internet into a global phenomenon has also come to mean more problems with security due to lack of such features. This situation is understandable since, the internet itself was only originally invented on a small scale in a very trust-worthy environment. This mind-set is reiterated by Terry Williams when he opines that, "TCP/IP and most of its protocols and utilities were not written with security as priority. They were designed for functionality and portability" [23]. As a result of the many security holes in the implementation of TCP/IP stack, it has been subject to numerous attacks. This section sheds light on five of such attacks. They include SYN flooding, ICMP attacks, source routing and UDP spoofing [57].

2.6.1 SYN Flooding

TCP uses the three-way handshake protocol to establish a session. The protocol allows a limited maximum of five connections to a host to be established by default and each connection can be completed within 75 seconds to accommodate delay in response acknowledgement. The problem with such a protocol is, when the in-coming connection SYN requests are more than five, the sixth connection is put on hold or queued for at least until another connection is released. This scenario is referred to as SYN flooding. It can be used by attackers with fake IP addresses to send multiple requests to a host. Since the attacker doesn't need to return an acknowledgement message, he succeeds in first, denying other legitimate hosts from gaining connection and second, taking advantage of the half opened connection to launch intended attacks. A pictorial representation SYN flooding is illustrated in Fig. 2-14.

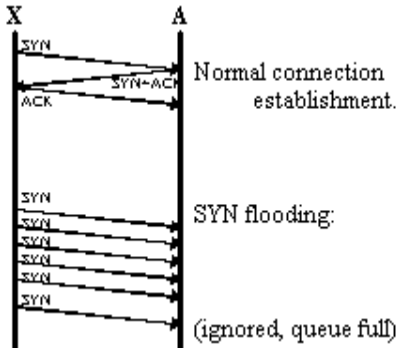


Figure 2-14: SYN flooding attack [64]

2.6.2 ICMP Attacks

The internet control message Protocol-ICMP defined in RFC is used by the IP layer to provide services such as reachability or connectivity information and error report feed-backs which are useful for network trouble-shooting. ICMP uses the *ping* utility program to send message messages from one host to the other. The message could be “echo request”, “echo reply”, “Time exceeded” and “Destination unreachable”. These messages on the other, hand when sent by an attacker, succeed in denying a host a connection to another host. This is done by sending “echo request” message larger than 64k bytes instead of the normal size of 64 bytes. This attack, which is popularly known as “ping of death”, has proved to be a major risk confronting many operating systems. Another form an ICMP attack sees an attacker sending a “time exceeded” or “destination unreachable” message to a host waiting for a response. The waiting host would interpret the message to mean the request packets have expired. Eventually, these false message signals will force connection between hosts to be misleadingly closed.

2.6.3 Source Routing

The option field in the IP header can be used by the originator of a packet (the source) to determine the path that reply packets should take from the receiver. This is known as source routing which comes in two flavours namely, strict and loose source routing. In strict source, all hosts to be traversed in a particular order are defined. In case the path is not traversed, an ICMP error message is sent back to notify the source. The second kind of routing being the loose source routing poses a threat since packets are not really restricted to a specific path. An attacker can modify routing information to include his host which helps him to redirect and monitor packets intended for the originator. Turning off source routing in networks has become popular work-around this vulnerability.

2.6.4 IP Fragmentation

IP uses fragmentation to split datagram into smaller fragment when they approach network with smaller MTU. Each of the fragments is attached to a copy of the original IP header and reassembled on arrival at the destination. The issue here is, when the data portion of an IP datagram is fragmented, only the first portion carries the TCP segment. In the case where a filtering service such as a firewall is deployed in a network, a policy could demand the dropping of packets arriving without TCP header information.

2.6.5 UDP Spoofing

The IP network layer uses UDP for connection-less transport services. Unfortunately, UDP only offers meagre services such as port numbers for identifying different connections and header checksum for preserving header integrity or protection against header forgery. It doesn't provide any sequence numbering service or identification. This deficiency makes it hard to detect whether a group of packets belong together hence, the difficulty to detect that, a forged packet has been placed between datagram.

3 VoIP- IP Telephony

When TCP/IP networks begun to process voice and video contents in addition to the pre-existing formats, the internet gave rise to concept of converged networks. The corporate world hailed it as IP telephony but the "techies" stuck with Voice over Internet protocol or VoIP for short. The "*Wikipedia*" definition of VoIP "as the routing of voice conversations over the Internet or any other IP-based network" [7], has the tendency to limit its understanding or immense scope to only cheap phone services over the internet- as many people have come to think. VoIP extends from this perception to include real-time video streaming, conferencing and killer application technologies such as unified messaging (UM) and instant messaging (IM). Popular day-to-day use of VoIP applications includes *Skype* and Microsoft's *Netmeeting* [62].

In this chapter, I will delve into how a VoIP network infrastructure processes voice and handles network design issues that must be taken into account to prevent a degraded VoIP service in communications between end-systems. This will include functions of VoIP components, signalling and media protocols (SIP and RTP).

3.1 VoIP Functions

For a VoIP system to function as a viable telephone system like PSTN, it had to have real-time qualities. Unfortunately, the network foundation of VoIP which is the internet's TCP/IP protocols was not built from the scratch with real-time capabilities in mind. However, with the standardization of media protocols by IETF, the real-time protocol-RTP [29] over UDP/IP became the way forward for real-time multimedia processing. VoIP preferred UDP to TCP because of UDP's light-weight nature with only 8 bytes of header size which contrasts with TCP's heavy-weight session based 20 bytes header size. This UDP feature, combined with some tweaking of IP "Type of service" field in the header, makes data streams carried in RTP faster to transport. The preference of UDP comes with its fears since it lacks the mechanism to guarantee packet orderly arrival through sequence numbering and retransmission of lost packets. This deficiency is compensated in RTP (Fig. 3-1) which support sequence numbering. In reality, ensuring 100% packet delivery in most real-time communication such as audio conversations is not too necessary since communicants can always tolerate some level of packet loss as long as the conversation makes sense.

For VoIP to be a force to reckon with in the telecom industry, its core services must mirror that of existing telephone infrastructure such as the PSTN by providing the following functions [11].

1. Codec operation
2. Signalling
3. Database service
4. Call control and bearer channel

3.1.1 Codec Operation

Unlike the old traditional analogue voice communication networks, VoIP networks are digital. That implies, analogue conversation would have to be converted to digital format before they are carried to their destination. On arrival at their destination, a similar process would have to be used to convert the digital format back to analogue. These processes can be achieved by analogue-to-digital or digital-to-analogue converter devices. In a call setup, the caller's side uses ADC to voice to digital form. The callee's side would use the DAC to reverse the digital format into the original voice.

After voice has been digitised, the next process compresses it further for faster transmission through the network. Different compression standards can be used communication depending on the purpose but only one standard can be used for a particular communication. The international telecommunication union- ITU defines a list of the top five compression encoding standards in fig. 3-1. Referring to the list, Stephen Brunn and Akhlaq Ali (2004) draw attention to "the trade-off between encoding efficiency, reduced bandwidth consumption and increased conversion delay" [11]. The coding and decoding feature in VoIP networks equips it to ensure effective bandwidth usage. This is not the case in PSTN systems which is always happy to assign an under utilised bandwidth for each communication.

ITU Standard	Description	Bandwidth (Kbps)	Conversion Delay (ms)
G.711	PCM	64	< 1.00
G.721	ADPCM	32, 16, 24, 40	< 1.00
G.728	LD-CELP	16	~ 2.50
G.729	CS-ACELP	8	~ 15.00
G.723.1	Multirate CELP	6.3, 5.3	~ 30.00

Figure 3-1: ITU audio encoding standards [11]

3.1.2 Signalling

The task of prompting end-nodes of the other's intension to establish a communication or call session is left in the hands of signalling protocols.
The two popular and competing signalling protocol standards for multimedia delivery are the H.323 and session initiation protocol-SIP. The H.323 [42] standard is defined by ITU whiles the SIP [32] is the brainchild of the IETF. According to Brunn and Ali, even though H.323 has been wildly deployed than SIP, the status-quo was about to change. This view seems to be shared by many in the industry who favour SIP over H.323 on the grounds of its hyped simplicity and flexibility. Figure 3-8 presents some clear differences between the two protocols in their bid to achieve the following the goals [17]:

- Provision of capabilities needed to set up, manage, and tear down calls and connections;
- Scalability support for a very large number of registered endpoints, and simultaneous calls (in the order of millions worldwide);
- Support network management features for policy control, accounting, billing, etc;
- Provision of mechanisms to communicate and set up the Quality of Service requested by the end points;
- Extensibility to promote the addition of new features easily;
- Interoperability support among different vendors' implementations, among different versions of the signalling protocol, and with different signalling protocols.

Signalling protocols go through three phases to set-up, negotiate communication parameters and tear down communication. The first phase is known as the initiation stage where the caller first contacts the callee to agree to communicate.
The second phase is used to negotiate and exchange call parameters both ends will use in the communication. The main information exchanged should contain which codec format to use, UDP port number for identification and media protocol for the conversation. Third phase is used to tear down communication after conversation.

3.1.3 Database Services

Database services are used to keep track of end-node's current address and location. End-node devices are assigned unique IP addresses for registration and identification in addition to the specific ports on which conversation will take place. Information provided by such a service become helpful in billing purposes and promotes security by assigning call capabilities to end-nodes.

3.1.4 Call control and bearer channel

Both PSTN and VoIP use different mechanisms for setting up a communication session. In a PSTN call session, end nodes are assigned a dedicated channel for the whole call duration. On the other hand, a VoIP call session is characterised by the real-time delivery of multimedia IP packet stream from the source to the destination. The call traffic contains two kinds: one for signalling or controlling conversation and the other is the conversation itself which is also known as the bearer traffic.

3.2 RTP Header

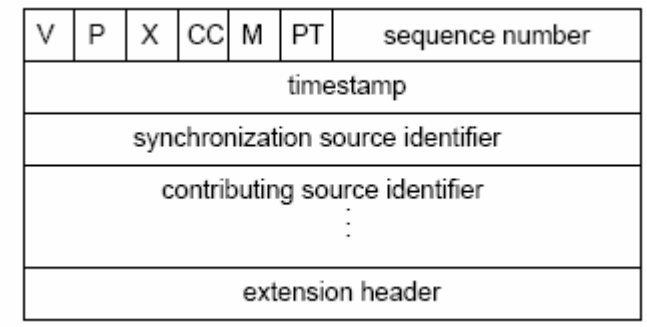


Figure 3-2: RTP packet header

RTP is a transport protocol for real-time applications defined by RFC 1889 [28] and 1890 [29]. The services offered by RTP include payload type identification through encoding scheme format, timestamps for replay detection, delivery monitoring or packet loss detection through sequence numbering and reconstruction of packets using special network devices. In figure 3-2, the *V* represents the first two bit version number of the RTP protocol. The next bit is *P* which stands for padding which becomes necessary in cases such as encryption algorithms that require plaintext to be of a specific length. The *CC* is a 4 bit value specifying the number of contributing sources. The *PT* represents the payload type which specifies the encoding scheme used by the media gateway to digitize voice format.
The PT field identifies the RTP payload format be it audio, video or application and determines its interpretation by the *CODEC* in the media gateway. A profile (RTP Profile for Audio and Video Conferences with Minimal Control) specifies a default mapping of payload type codes to payload formats.

Users participating in an RTP session are uniquely identified by a 32 bit randomly generated synchronizing source-*SSRC*. In cases where session involves audio and video conferencing, RTP provides two relay mechanisms to process packets. The first mechanism uses a network device called a mixer to connect communicating networks. Mixers tend to be useful when some participants in a session are behind a low speed link in the mist of the other high speed link participants. The mixer channels the packets of the low link users through it to resynchronize or reconstruct the in-coming stream-also known as contributed source stream.

The RTP header information at this point, changes to reflect the mixer as the synchronising source. The list of users contributing to the out-going packets stream is identified by their 32 bit contributing source identifiers. The second mechanism uses a translator device to, for instance, process packet encoding conversion from one network to the other. This basically does not require the translator to change the *SSRC*.

3.3 VoIP Components

The components that make up a VoIP deployment may be dictated the signalling traditions and architectural philosophies of the standard bodies and VoIP vendors in question. This means that, the set-up can either be IETF, ITU-T, or Level 3 Communication based architecture, to say the least. Whereas IETF implementations are mainly governed by a SIP server in different modes, the ITU-T implementation favours gatekeeper and media gateway collaboration.

The third emerging kind is *Level 3 communication*'s media gateway control protocol–MGCP based architecture which supports a media gateway controller and media gateway. This is indicative of the potential non-interoperable environment that VoIP operates in. However, there has been some level of cooperation between IETF and ITU in their bid to support and promote a newer or improved version of the MGCP. In the ITU world, this protocol is known as H.248 while the IETF world knows it as Megaco [43]. The H.248/Megaco standard was designed with standardization of IP telephony equipment in mind to achieve the following [11]:

- Meeting basic needs of business users.
- Rapid expansion which help support sophisticated telephony features
- Categorisation of telephony devices from very simple to very feature rich.
- Minimal and simple design.
- Device cost which matches capabilities.
- IP phones which meet Megaco/H.248 protocol requirements.

Figure 3-3 presents a self descriptive MGCP network architecture followed by a brief description of the afore-mentioned components.

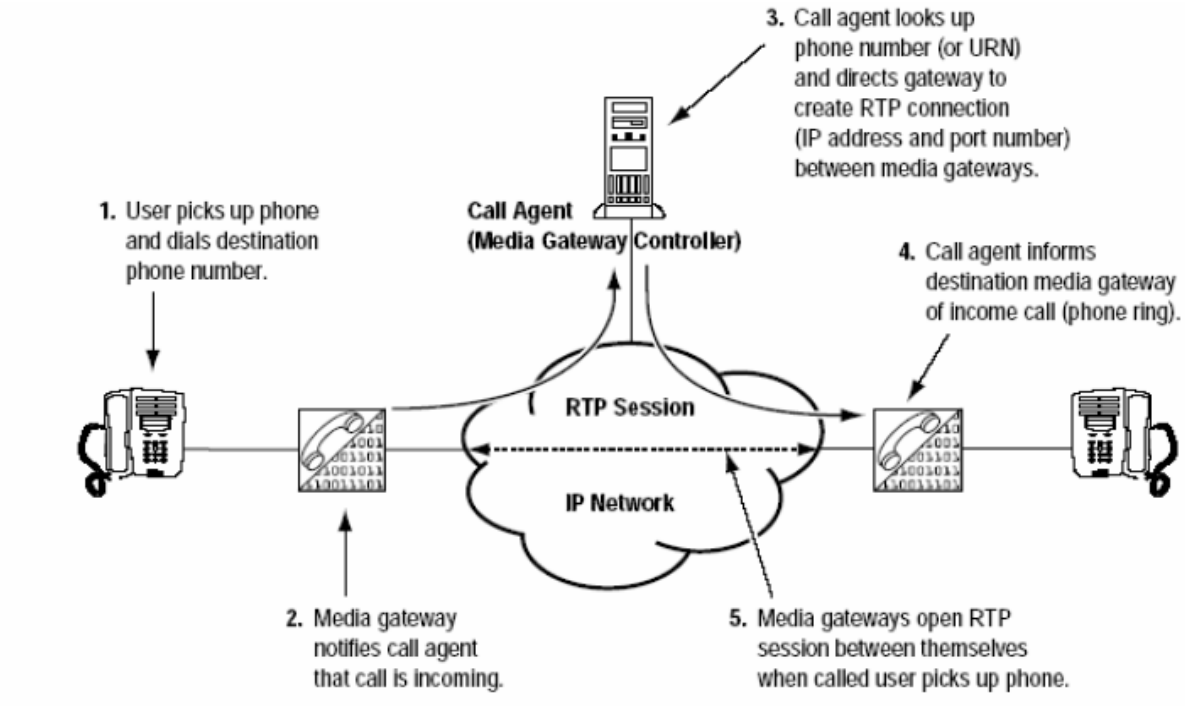


Figure 3-3: MGCP network architecture [11]

Call Processing Server: It is used for establishing and tearing down a call session by using a signalling mechanism to control call setup traffic. In the setup process, signalling between end nodes are conducted through the processing server after which, the end nodes communicate directly without the server's mediation. Call servers also perform traffic switching between VoIP networks and traditional switched networks and normally also called IP-PBX.

Media Gateway: The power of the VoIP would have been limited if users could only communicate in a PC-to-PC fashion. The ability for VoIP to interoperate with a circuit switched network is provided by the media gateway. This component converts voice convenient for packet switched to circuit switched networks or vice versa. Other functions of the media gateway include effective bandwidth management through voice compression, silence suppression and echo cancellation. In effect, whereas the call processing server handles signalling traffic, the media gateways are in charge of the bearer traffic or the stream of the VoIP payload.

Gatekeeper: In a typical H.323 network architecture, a gatekeeper is used as central point to directly manage or control a media gateway in a call setup or control process. Other critical functions of this component include ensuring authentication, authorization for call access control, and accountability through the monitoring of the network.

Media Gateway Controller: This component, which is normally also referred to as call agent in MGCP architecture, gives the green light for media gateways to create and tear down a session. The media gateway controller establishes this client/server relation with the media gateway through the signalling protocol MGCP. MGCP basically assigns signalling control to the controller and media traffic flow to the media gateway.

3.4 Session Initiation Protocol-SIP

As already stated, SIP is a signalling protocol defined by IETF RFC 2543 [32]. It is a text-based protocol modelled on existing web's HTTP and electronic mail SMTP with aim of achieving simplicity, efficiency, extensibility. Evidence of this influence, can be seen from how SIP messages look like HTTP request and response formats whereas naming follows an e-mail or URL format. SIP, unlike H.323 was designed to be simple and to be carried over UDP preferably, or TCP. However recent SIP specifications favours TCP for call reliability. SIP provides end-to-end client-server signalling with the support for presence and mobility. To help a SIP network achieve the mentioned characteristics, it uses the following four component services:

- User Agent
- Proxy service
- Redirect service
- Location/Registrar

User agent: They are normally the end-systems devices responsible for initiating and tearing down a call. They can either act as client (UAC) as in VoIP phones, or as servers as in location servers.

Proxy service: These are servers that are used by end-system or UAs as a path or route for signalling. Proxy servers do not take part in the actual conversation during session establishment.

Redirect service: They are user agents that respond to client's request signal with a redirect message to the signalled device new location. In effect, they provide both end-systems the opportunity to establish and terminate a session by themselves.

Location/Registrar: These database servers provide the mobility characteristic of SIP by handling registration request from UACs. This service makes it easier for clients to contact each other no matter the distance or location. To locate users on different networks, SIP assigns special formatted addresses to end-systems. They look like e-mail addresses consisting of username and the domain for example *kobi@forum.net*. This naming convention provides a global scope for communication especially where DNS services can be used to resolve or track any user.

3.4.1 SIP Messages

SIP uses messages for call invitations in a connection setup and teardown. These messages are mainly text based and are divided into two parts which consists of header and body. The header gives information about the message type which is normally INVITE, the protocol type which is SIP version 2.0, the end points addresses, and the type of message body. The message body conveys a description of session multimedia capabilities the caller will like to establish. An SDP session description is made up of lines of text of the form: <type>=<value> where <type> must be exactly one case-significant character and <value> is structured text whose format depends on <type>.This is done by the ABNF (Augmented Backus-Naur Form) definitions for session description protocol defined in RFC 2327 which lists vital session information such as [33]:

- Media to use (codec, Sampling)
- Media destination(IP address and port number)
- Session name and purpose
- Time the session is active

Messages are either a server or client request or response. Message formats normally has a start line and some header fields making up the header. This is followed by the message body and illustrated in Fig. *3*-a and *3-b*. A start lines uses one of the six request methods or response codes to indicate the type of message. The codec and sampling rate associated with payload types are provided in tables 4 and 5 of RFC 3551 (RTP Profile for Audio and Video Conferences with Minimal Control) [34]. A description of the request methods and response codes follow in the next section.

3.4.2 SIP Request methods and Response codes

SIP uses the following six types (methods) of requests:

- **INVITE:** Indicates a user or service is being invited to participate in a call session.
- **ACK:** Confirms that the client has received a final response to an INVITE request.
- **BYE:** Terminates a call and can be sent by either the caller or the callee.

- **CANCEL:** Cancels any pending searches but does not terminate a call that has already been accepted.
- **OPTIONS:** Queries the capabilities of servers.
- **REGISTER:** Registers the address listed in the **To** header field with a SIP server.

The following types of responses are used by SIP and generated by the SIP Proxy Server:

- **SIP 1*xx*:** Informational Responses
- **SIP 2*xx***: Successful Responses
- **SIP 3*xx*:** Redirection Responses
- **SIP 4*xx*:** Client Failure Responses
- **SIP 5*xx*:** Server Failure Responses
- **SIP 6xx:** Global Failure Responses

Figure 3-4a: A SIP message request

Figure 3-4b: A SIP message response

3.4.3 SIP Operation Modes

In SIP network architecture, end-systems use a SIP call processing server to mediate a session establishment. The SIP server can be configured to function in two kinds of modes depending on the special needs and requirements of the network. The SIP can either assume proxy or a redirect mode.

Proxy Mode

A typical proxy mode network environment is ideal for security critical infrastructure which involves firewalls and user restrictions, access control and monitoring through authentication, authorisation and mechanism. The message flow between end-systems in this mode is illustrated in figure 3-5. In this mode, end-systems which are UACs, first send registration messages to the registrar to provide information about their location. The clients involved in a communication according to the illustration, are Tom and Alex. Tom's phone initiates a call session by sending a SIP *INVITE* message inviting Alex for a conversation. The proxy server finds the actual location of Alex through location service and proxies or forwards Tom's messages containing his phone capabilities to Alex's phone. When Alex's phone starts to ring, an informational SIP 180 ringing message is sent back to Tom through the proxy to indicate a ringing tone. When Alex picks up the phone, the phone sends a SIP *200 OK* message indicating success to Tom. The SIP *200 OK* messages convey Alex's phone media characteristics to be used in the conversation. Tom's phone then sends a SIP acknowledge message to Alex indicating connection complete. Tom and now start to communicate with each other using RTP without the assistance of the proxy server.

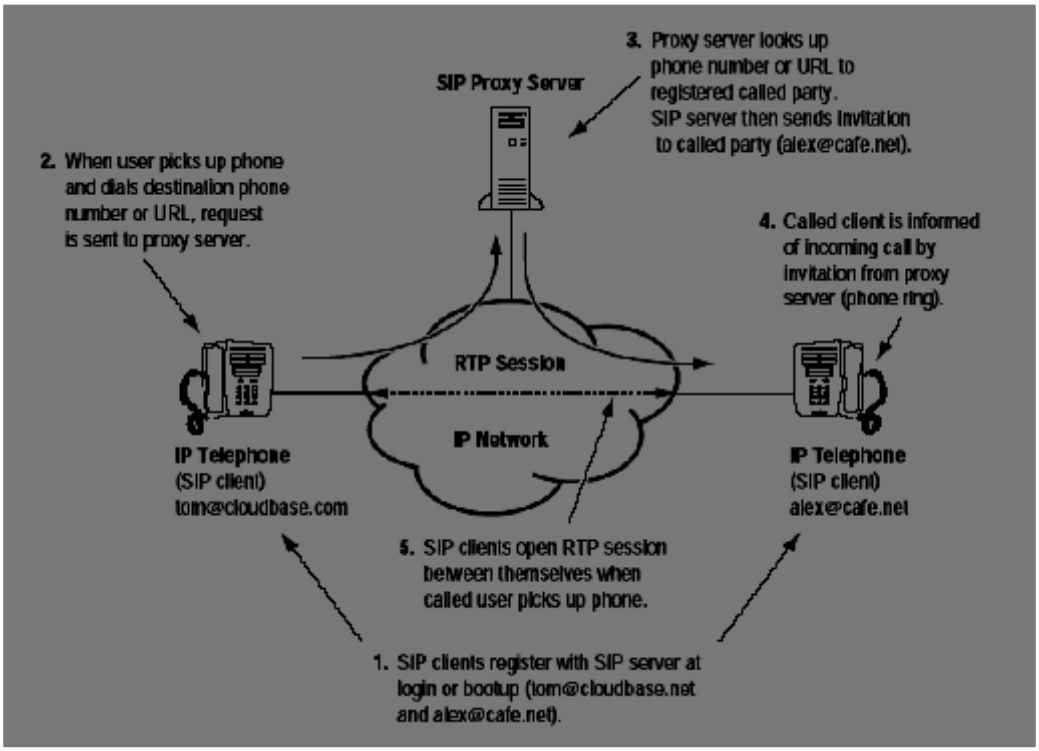


Figure 3-5: SIP in proxy mode [11]

Redirect mode

SIP operation (Figure 3-6) in the redirect mode becomes helpful when the location of the *callee* is subject to location changes or mobility. In that case, when the caller tries to initiate a call through the SIP server, the location server checks and sends the new location information of the *callee* back to the caller. The caller uses the new location to launch a direct communication with the callee. An equivalent of the proxy signalling in a redirect environment between *Tom* and *Alex* is as follows:

After both clients have registered their location to the registrar, *Tom* initiates a signal to *Alex* through the SIP server with an *INVITE* message. The server's location service performs a lookup of the actual or new destination of Alex especially in the case he has moved to a different service provider. Information on his new location is sent back to Tom, which, he uses to initiate a call directly to *Alex*. *Alex*, like in the proxy mode, responds with a SIP *200 OK* to indicate success and his readiness to communicate.

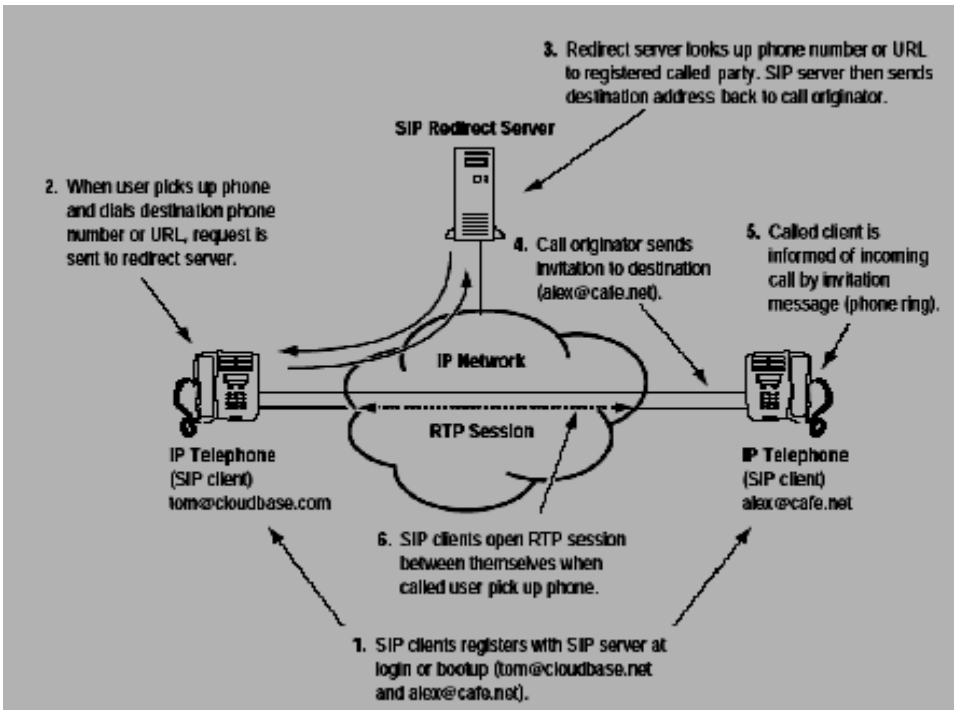


Figure 3-6: SIP in re-direct mode [11]

3.5 Overview of H.323

An extensive coverage of the ITU-T recommended packet-based signalling protocol for multimedia communication which is H.323 [42] is beyond the scope of this paper. However, an overview of it in addition to its subtle differences with IEFT's SIPS protocol will help with putting signalling protocols in better perspective. Figure 3-7 captures the complex nature of the H.323 by giving a simplified description of its constituent sub protocols. Figure 3-8 then further present some clear-cut differences between SIP and H.323.

Name	Description of protocols
H.323	Specification of the hole system
H.225.0	Call Control, Call Setup
H.235	Security protocol for authentication etc.
H.245	Capability exchange and mode switching
H.450	Supplementary services
H.246	Interoperability with circuit-switched networks
H.332	For large size conferences
H.26x	Video codecs
H.7xx	Audio codecs

Figure 3-7: Overview of H.323 [17].

	H.323	SIP
Components	Terminal/Gateway	UA
	Gatekeeper	Servers
Protocols	RAS/Q.931	SIP
	H.245	SDP
Transport protocol	Reliable or unreliable, e.g., TCP or UDP. Most H.323 entities use a reliable transport for signalling.	Reliable or unreliable, e.g., TCP or UDP. Most SIP entities use an unreliable transport for signalling.
Message Encoding	H.323 encodes messages in a compact binary format that is suitable for narrowband and broadband connections.	SIP messages are encoded in ASCII text format, suitable for humans to read.
Addressing	Flexible addressing mechanisms, including URLs and E.164 numbers.	SIP only understands URL-style addresses.
PSTN Networking	H.323 borrows from traditional PSTN protocols and is therefore well suited for PSTN integration.	SIP has no commonality with the PSTN and such signalling must be made SIP compatible.

Figure 3-8: Differences between SIP and H.323 [49].

3.6 Network design issues

3.6.1 Bandwidth

Proper or adequate bandwidth allocation to VoIP applications and services is one right way to ensure a smooth service delivery of audio, video and data. Since it is true that video streaming or conferencing certainly requires more bandwidth than voice conferencing and for that matter, data as in instant messaging, a much better way to assign bandwidth was to categorise or segment the network based on bandwidth capacity and requirements of each separate service.

This is a decision that a company and its network administrators would have to take based on business interests and priorities. An organisation adopting IP telephony to reduce cost is surely going to prioritize voice signalling and communication over web traffic (HTTP). The effective bandwidth management has been addressed using compression and silence suppression technologies such as encoding schemes in media gateways in a VoIP network.

3.6.2 Packet Loss

The issue of packet loss or dropped packet in a network has been dictated by the kind of transport layer service used and the network utilization [11]. During peak times, routers in a congested network would have no choice than to drop packets when they run out buffering resources. However, whether packet loss matter depends on whether the application requires a real-time processing like voice or non-real time like e-mail and file transfer. In the case of real-time application which prefer to UDP for transport, packet loss is critical especially when UDP promises only datagram service and no retransmission of lost packets. In voice call setup, bearer and signalling packet losses can lead to poor service quality or call disruption. A work-around this challenge is to highly prioritize packets and assigned enough bandwidth in times of congestion. On the other hand, some level of packet loss can still be tolerated as long as speech conversation makes sense. Non real-time application using TCP have the luxury of retransmission guarantees in the case where packet loss is concerned. In that case, quality of service issues is not as critical as in real-time applications.

3.6.3 Interoperability

The young and evolving nature of VoIP has come to mean that, vendors of product manufacture them using proprietary specification. The result of this individualism is a situation where networks become incompatible with each other. The only way for instance, VoIP gateways or signalling protocols can interoperate, is when they come from the same vendor. The clear solution to this problem is for VoIP to have one standard. The IETF and ITU collaboration on standardizing of IP phones with common signalling protocol (H.323/megaco) network architecture has been applauded as a step in the right direction. In other words, businesses wanting to deploy VoIP on a large scale should look out for vendors that incorporate open standards in their products. In today's multi-protocol world, interoperability is increased if products supported more than one protocol and could operate in both single and multi-protocol mode. This provision will definitely reduce cost of equipment upgrade and eliminate the need to change infrastructure in cases where communication has to be established with another network which has a different vendor.

3.6.4 Jitters

Another consequence of network congestion and for that matter, fluctuation in traffic, is the inconsistent arrival of packets at their destination. This phenomena which is known as jitters, can also be caused by packets delaying in arrival by changing their routes. Just like packet losses, jitters are inevitable and some small level of it can be tolerated. Mitigating jitters requires saving arrival packets in a media gateway's buffer jitter and reprocessing them to arrive at constant intervals. Figure 3-9 illustrates with a graph a network jitter scenario where in the first line, packet P1 to P4 all arrive as expected. In the second instance, which implies jitters, packet P2 and P4 arrive 12ms and 15ms late respectively.

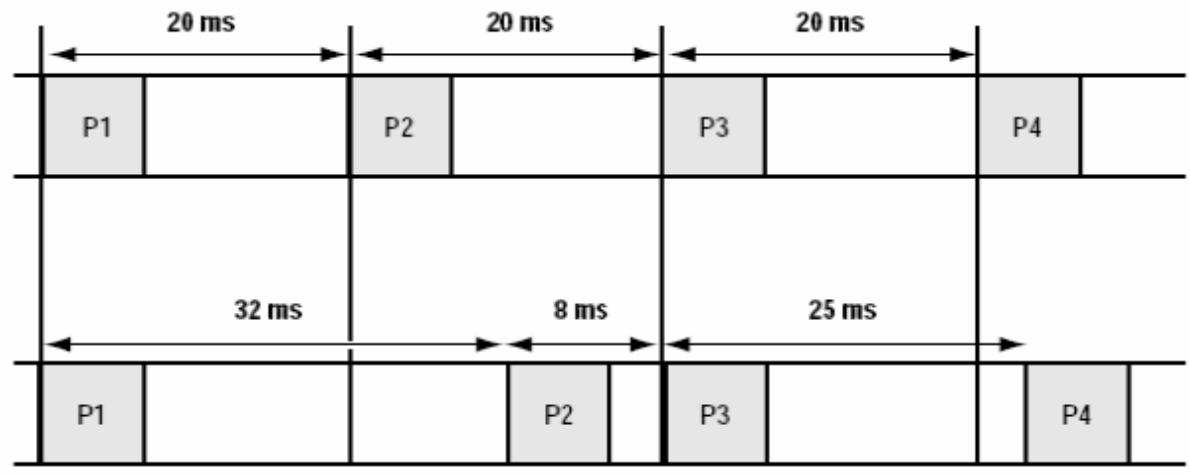


Figure 3-9: Jitters in packets [11]

3.6.5 Reliability

Maintaining an around-the-clock availability of network service has been a challenge to business that relies on them for their critical operation such as online transactions. The well too known denial of service attack and possible break down of network link has proved to be the enemies of availability. Consequently, network designers have been quick to build fault tolerant networks by adopting fail-over strategies. At the network layer, redundant paths are provided so that packet could be re-routed when a link is down. At the application level, alternative server resources are made available just in case one fails or can't handle any more service requests.

4 VoIP Security

With the advent of hackers or malicious users on the internet, the race to ensure better security has become even more critical. With the help of bug ridden software, networking capturing and network analysing tool like Ethereal, and media access control (MAC) and IP spoofing tools [55], unauthorised users have been able to by-pass security mechanisms to access, modify or delete sensitive information. In a VoIP network, ensuring adequate security is as important as traffic delivery service quality such as guaranteeing low packet delay and reliability.

This chapter present a holistic view of system security with a special bias towards network security. It touches on some critical attacks in the operation of VoIP and corresponding security measures (TLS and SRTP). Upgrading of service quality and security would be discussed in the context of switch-based LANs which supports IEEE's 802.1Q virtual LAN standard. Finally dynamic firewall techniques and the network layer security protocol IPSec are presented as mechanisms for protecting trusted networks from the hazards of the internet [45].

4.1 Branches of Security

By now, it is clear that VoIP networks are basically IP networks hence, any component such as IP-phones, media gateways or call server that plug into a network will experience similar if not the same attacks, vulnerabilities and threats that confronts IP networks. To tackle or asses the origins of these security dangers, an organisation would have to look well beyond hackers or malicious attackers. In other words, organisations would have to adopt methods and strategies for determining how to mitigate such dangers that can potentially affect the organization as a whole. One such widely adopted method for scrutinizing an organization-wide security posture is the OCTAVE (Operationally critical Threat, Asset, vulnerability Evaluation). The OCTAVE method identifies three main security areas a VoIP infrastructural management should take into account to ensure adequate protection of critical operational assets which are vital to the organisation. These three main areas reflects the principal believe of OCTAVE that, "security includes both organization and technology aspects" [20]. They are:

- Physical Security
- Information Security
- Staff Security

A physical security strategy is geared towards protecting infrastructural hardware from being tampered with, stolen, or damaged due to power cuts or natural disasters. Dealing with such challenges require a rigorous policy on physical access control to sensitive hardware components in addition to effective auditing and monitoring the system's hardware.

Information security refers to what normally comes to mind when security is made mention of. This delves into both computer and communication security. Computer security tackles vulnerabilities that expose system resource due to it faulty architecture. In this regard, sensitive information may be compromised by design weakness such as an easy-to-bypass authentication and authorization mechanisms. A weak design also leads to a problematic implementation such as buffer overflow vulnerabilities in application built with programming languages such as C and C++. Another quite obvious challenge of computer security comes from administrative over-site or incompetence in system configuration. Viega and McGraw (2001) on the issue computer security have advised building software with security from the scratch rather than applying occasional patches or fixes any time a new vulnerability is discovered. The patch approach, most often than not, introduces more security holes and bugs [18].

Communication security relates to how information can be secured when it is carried across an un-trusted network. Such information will have to be guaranteed confidentiality, integrity and authentication. The least recognised or talked about security category is the staff security which is also divided into personnel and operational security. Personnel security will be crucial in minimizing attacks or security breaches from authorised users on a system. Either less skilful users becomes victims of social engineering activities such as revealing passwords to would-be attacker or disgruntled users who try to obtain special administrative privileges in order to carry out an intended attack. The OCTAVE approach suggests staff education to combat such development. Operational security promotes and enforces security policies which are basically contingency plans and procedures that are carried out should a breach in the system occur.

4.2 Attacks

4.2.1 Denial of service

In the very competitive world of IT, denial-of-service-DoS have been used by both attackers and competitors alike as an extra tool in their marketing strategy to curb competition. The e-commerce industry is one such fertile ground for such attacks. Since the industry relies on the internet for their critical operations, rival virtual shopping malls could beat competition by sabotaging the sites so as to prevent further processing of customer requests. Indeed, the intention of DoS is to disrupt continuous operation or business continuity which is contrary to attacks which either modify or delete data. DoS attacks exist in variant forms such as ICMP's "ping of death" and SYN flooding [57]. Businesses normally reduce DoS attacks by using a server cluster or farm to handle heavy requests from customers. The wisdom in this approach is to be able to switch to another server when one is down or over-loaded.

Disruption of services in a VoIP environment is not so different from DoS in data network scenarios. Critical components such as Proxy servers become targets for denial-of-service attacks especially as they directly process requests from the internet. Mechanisms used in conducting DoS attacks include packet flooding and session tear-down. A load based DoS attack is achieved when an attacker bombards or floods a poorly designed system with overwhelming fake requests that contain spoofed IP addresses. An attacker could also use request and response messages to tear down a communication session in cases where SIP UAs do not require strong authentication. This is done when an attacker observes a signalling channel for a call and then sends a spoofed SIP "bye" or "Cancel" request message to the communicating UAs. Besides, the same result could be achieved with the use of any of the SIP 4xx, 5xx and 6xx response messages (indicating user, proxy or global failure) [57]. Fig. 4-1 illustrates this attack.

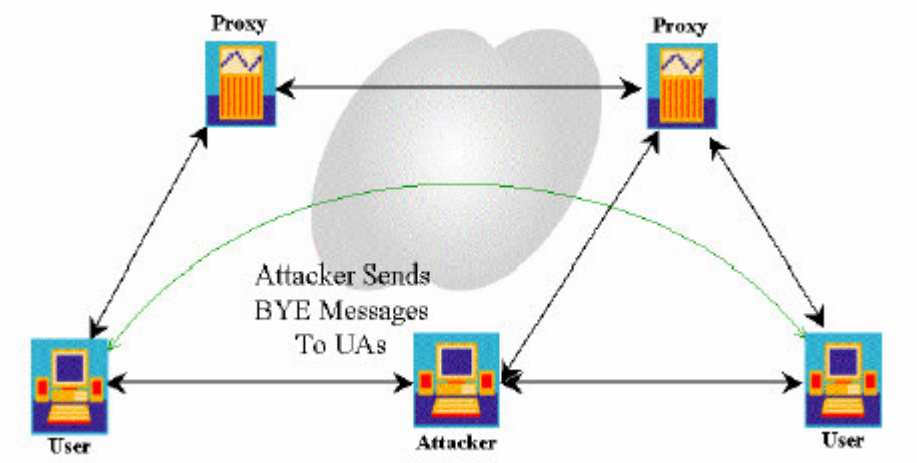


Figure 4-1: Denial-of-service attack [59]

4.2.2 Registration Hijacking

This attack occurs with the intension of impersonating a valid user agent or phone during registration to a registrar server. The attacker normally replaces the contact address in a SIP *REGISTER* message with his own IP address [60]. As a result, all incoming calls directed towards the legitimate user agent will be sent to the attacker instead. This attack which is illustrated in Fig. 4-2 is made possible when there is no cryptographic verification or protection of request message origin.

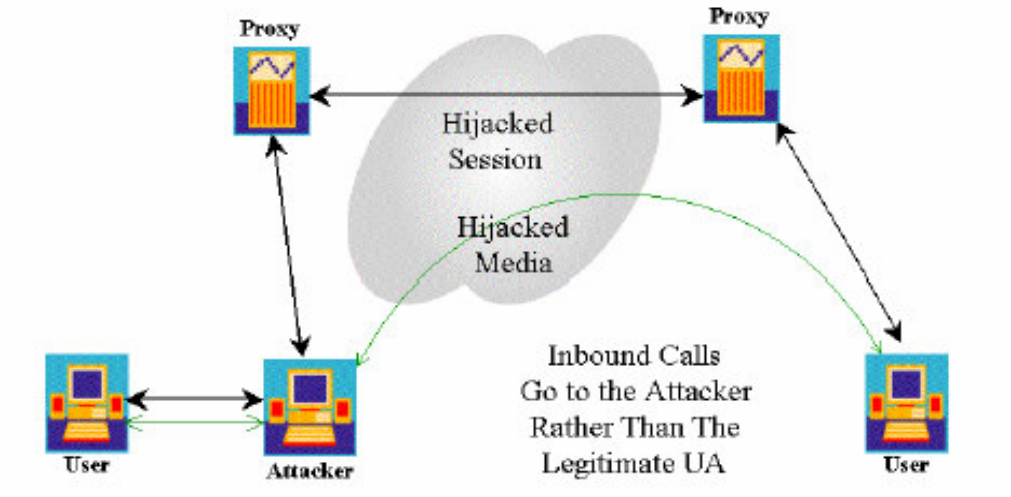


Figure 4-2: Registration hijacking [59]

4.2.3 Proxy Impersonation

This attack takes advantage of networks which don't promote mutual authentication process between entities. An attacker could insert a rogue proxy between a communication between proxies or a proxy and UAs. The rogue proxy which impersonates the real proxy would be used by the attacker to intercept request and response messages thereby taking absolute control over the entire communication [59]. Fig. 4-3 illustrates this attack which is also called masquerading attack in data networks.

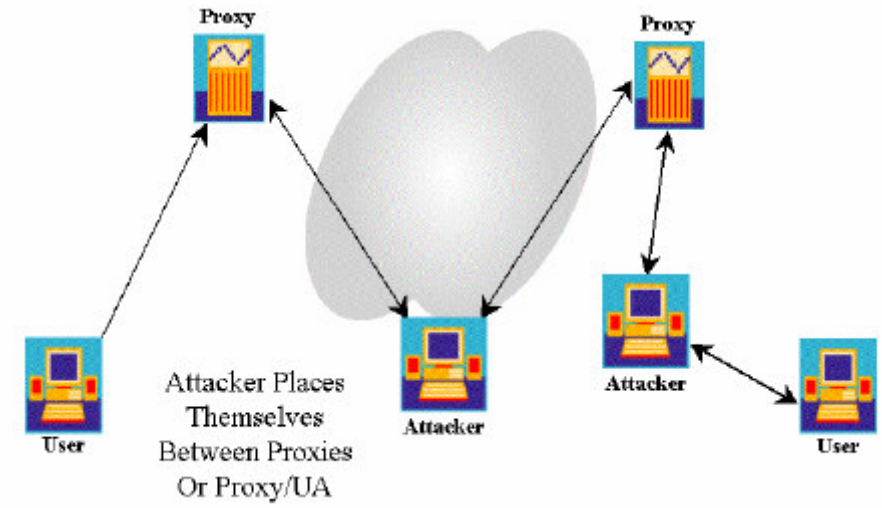


Figure 4-3: Server masquerading or impersonation [59]

4.2.4 Eavesdropping

This attack makes use of network capturing and analysis tools like Ethereal to sniff signaling messages and media streams in a conversation. The captured RTP over UDP or TCP packets are decoded and converted in audio files. A complete description of how Ethereal handles eavesdropping can be found in [59]. However, the steps taken to capture and decode voice packets include:

- First capturing and decode RTP packets by selecting **Analyze -> RTP-> Show all streams** options from the ethereal interface.
- Secondly, the session is analysed by selecting a stream to analyze and reassemble.
- The stream can now be converted into audio format through publishing. Open a file to save the audio (.au) steam that contains the captured voice.

4.2.5 Man-In-The-middle attack

Man-in-the-middle attack [57] establishes a three way communication between two parties and an attacker between them. Through out the communication session, the two parties do not notice the participation of the attacker. The attacker succeeds in directing traffic between the two parties through him. Information sent to and fro are intercepted and modified or read. A typical example of this attack would be the alternative Diffie Helman Key exchange phase in a TLS handshake call setup process. This process is vulnerable to this attack making it advisable to use public key encryption such as RSA instead for key exchange. RSA provides enough bases to eliminate this attack by the use of digital signatures and certificates for enhanced authentication.

In a VoIP environment, this attack uses a mechanism called ARP spoofing [45] to eavesdrop on communication in a switched based IP network. Since Ethernet switches restrict broadcast traffic within a network, most network designers use them to enhance security through limiting access to traffic. However with the help of a special ARP spoofing tool such as *Cain* [61], a man-in-the-middle attack can be successful in switched LANs. In ARP spoofing illustrated in Fig. 4-4, the attacker broadcast spoofed ARP messages containing fake MAC addresses to the Ethernet LAN making it possible for him to receive another station's frames. In Fig. 4-4 frames sent between user *A* and *B* are mistakenly sent to the attacker making it possible for him to sniff it. This is possible because, when user *A* wants to send frames to *B*'s IP address, a cache table which contains a mapping of IP to MAC addresses reveals that, *B*'s MAC address is actually the attacker's MAC address-resulting in the packet being sent to the attacker rather. The other way round is true when *B* sends frames to *A*.

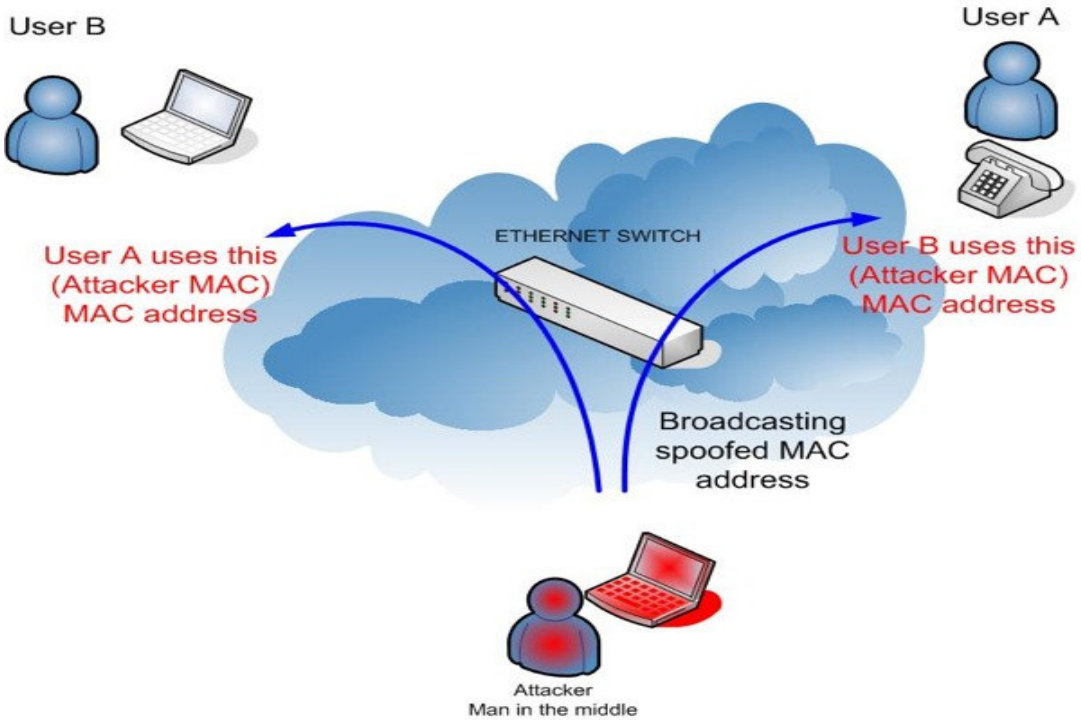


Figure 4-4: Man-in-the-Middle Attack using ARP Spoofing [60]

4.2.6 Replay Attack

An attack with the help of network packet sniffing tools can conduct replay attacks by capturing information in a communication session [57]. The information captured could be retransmitted intact or modified to achieve a purpose. Financial institutions like banks deploying VoIP could experience a situation where sensitive data such as account numbers, and credit card information can be captured by internet fraudsters. In terms of system administration, users logging in should present a perfect environment for replay attacks. In this case, the attacker captures the username and password and uses them to login as a legitimate user later. Packet integrity helps eliminate. Timstamping, a feature supported by RTP in a synchronized environment mitigates replay attacks with the help of integrity protection mechanism such as hash functions MD5 and SHA-1.

Replay Attack in a VoIP environment would be possible when an attacker combines both eavesdropping and man-in-the-middle attack (discussed above) techniques to capture and re-send vital information such as network login credentials like passwords and usernames.

4.2.7 Buffer Overflow

Buffer overflow [57] has been referred to as “the crown jewel of attacks” which is expected to “remain so for years” [10]. Viega and McGraw (2001) identified it as the cause of “more than 50% of all major bugs” found in software application in 1999 [18]. It occurs when an application attempts to store more data in a static sized buffer. The result is the spilling of excess data onto neighbouring memory location and thereby erasing sensitive information. The worst scenario is when attackers manipulate the excess data to contain a malicious to spawn an attack on the system. The root causes of buffer over-flow has been laid at the door-step of careless programming practices and less endowed security oriented programming languages specifically C and C++. A work-around for this problem is to combine defensive coding practises with security auditing tools such as open source based RATS (Rough auditing Tool for Security) [56] in codes written in C and C++ or use Java instead which is more security inclined with an effective exception handling on bounds checking.

Operating systems existing on key VoIP component such as gatekeepers, IP phones and media servers run on either UNIX or Windows platform. They are as such susceptible to buffer overflow attacks and other platform dependent malicious attacks which include Trojan horses, worms and viruses. Soft phones, which are basically software based IP phones, are more likely to suffer constant attacks or even contain spy-wares and Trojan horses than hardware phones which normally run embedded operating systems [57]. In other words VoIP components in converged networks will encounter similar attacks that exist in data networks.

4.3 Security in SIP

Signalling and media communication carried by the RTP payload traverses a VoIP network unsecured. This means that, an attacker with packet sniffing and analysing tools could capture signalling and voice traffic for attacks such as replay and eavesdropping on conversations. To arrest these security nightmares, SIP based VoIP networks use two security protocols, TLS and SRTP, to handle signalling and media streams respectively.

4.3.1 Transport Layer Security- TLS

TLS version 1.0 is a security protocol defined by RFC 2246 [30]. It is modelled on Netscape's secure socket layer (SSL) version 3.0 which is used for securing web applications. Unfortunately or surprisingly both protocols are incompatible with each other. In a SIP network, TLS provides client-server application with message confidentiality, authentication of end-nodes and message integrity. When TLS is deployed between a UAC and SIP server the privacy and integrity protection will eliminate attacks such as eavesdropping on SIP message and further secure it from being tampered with or forged. TLS accomplishes its security goals with the help of its two components. They are the TLS Handshake protocol and the Record protocol (figure 3-9). The first phase of a TLS communication is conducted by Handshake process in which the two communicating parties negotiate a set of cryptographic algorithm that both support. The RFC specification supports the following popular ciphers such as [21]:

- RSA, Diffie Helman for public key cryptography;
- DES,3DES or AES for symmetric ciphers;
- MD5, SHA for one way hash functions.

After a successful cipher negotiation, both parties are now ready to authenticate each based on exchange of X.509 certificates [21]. Normally, only the server is authenticated by the client. The other way round, which is client authentication achieves mutual authentication (Figure 4-6). The Handshake process also negotiates a 48 byte "master key" and compression algorithms to be used for encryption by the record protocol. This is done by use of either public key cryptography or the result of the less secure Diffie Helman key exchange protocol which is normally vulnerable to Man-In-the-middle-attack (section 4.2.5). The record protocol uses the master key to generate the session key for encryption and integrity check. The encryption and hash algorithms used are the symmetric and hash algorithmic functions negotiated at the handshake process.

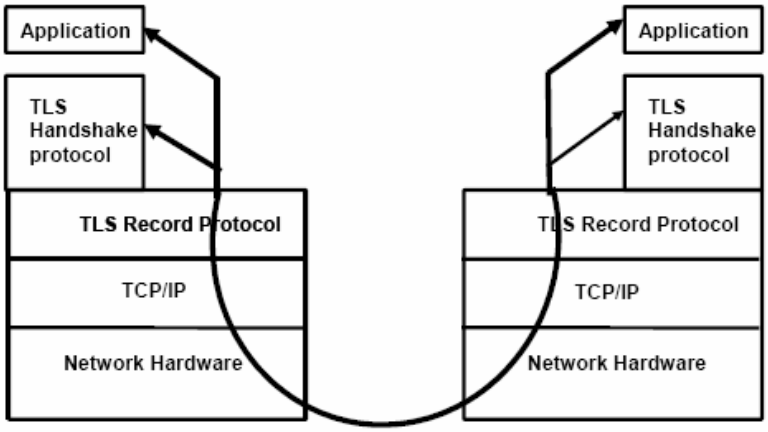


Figure 4-5: Layers of TLS [47]

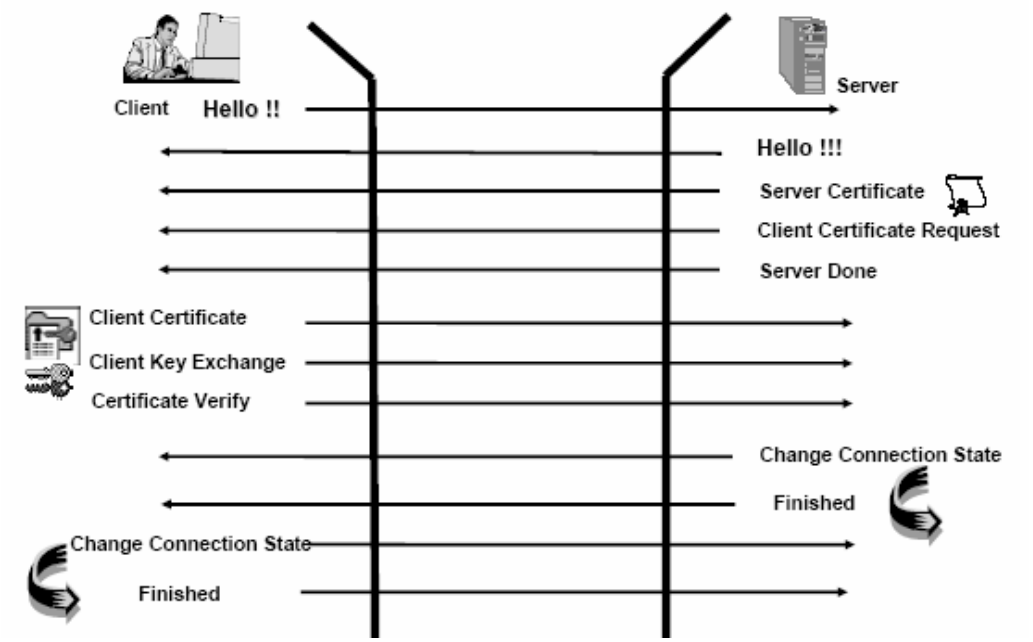


Figure 4-6: Message flow for full TLS handshake [47]

4.3.2 Secure real-time protocol-SRTP

The task of ensuring security in media communication or the real conversation is provided by the security protocol secure real-time protocol-SRTP defined by RFC 3711 [31]. SRTP is a security profile which extends the RTP profile. It provides both unicast and multicast data, audio, and video streams with confidentiality of their RTP payload, integrity of their entire RTP packet in addition to protection against replay attacks. Without these mechanisms, audio and video conversation or conferencing run the risk of being eavesdropped on, modified or captured and replayed later by would-be attackers.

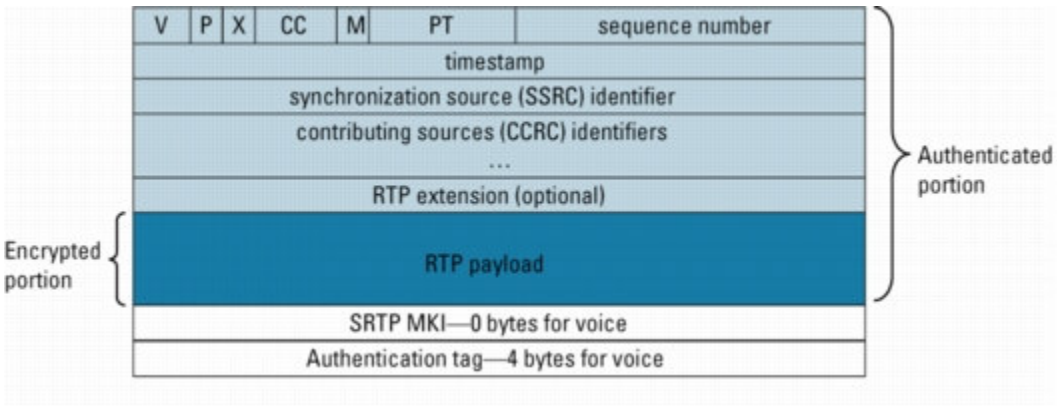


Figure 4-7: SRTP packet format [66]

SRTP resides between the RTP application and transport and referred to by specification (RFC) as a “bump in the stack” meaning that, RTP packets flowing down the stack or the sending side are converted and forwarded as secure RTP stream. On the other hand, when SRTP packets climb up the stack or receiving side, they are converted to RTP packet payloads for the consumption of the application layer. In providing confidentiality, the SRTP standard specified only one symmetric cipher which is the Advanced Encryption System-AES for encryption. AES operates in two different modes namely, counter and F8-mode [21].

In a typical encryption process, the behaviour of the modes change the block AES cipher into stream cipher. By default, AES runs in the counter mode and very suitable for RTP traffic that runs through unreliable network with the tendency of packets being dropped. The “counter” function ensures unique or non-repetitive cipher-text for a long time, making it hard for the original RTP stream to be deciphered by an attacker. The encryption process is further strengthened by a 128 bit encryption key and salt key of 112 bit.

Preserving authentication and subsequently integrity requires the use the more secure keyed-hashing message authentication HMAC-SHA-1 algorithm [21]. The keyed hashing message authentication is based on a shared secret key or the authentication key and a hash function. SHA-1 has been preferred to message digest MD5 [21] because of its strong key length of 160 bits, making it resistant to numerous attacks such as brute force.

Having provided confidentiality, authentication and integrity, it is still possible for attackers to capture RTP packets and resend them later for malicious intent. Such a scenario could arise in conversation that involves banks where sensitive information like account number, amount and personal information could land in the wrong hands. This kind of attack can be protected against by the receiver if a list of indices (sequence numbers) of previously authenticated messages is kept. The lists of indices are cross-checked against new messages index received. Only when the message has not been played before, can it be admitted. In order not to be fooled by spoofed or modified message index, integrity check must be enforced or ensured.

SRTP extends RTP by the addition of an optional MKI (Master Key Identifier) and a recommended authentication tag field. The authentication tag contains message authentication data which is the result of authentication of both the RTP header and the encrypted portion of the SRTP packet. In the absence of a defined key exchange protocol, SRTP uses the real-time efficient Multimedia Internet keying protocol (MIKEY) for key exchange. MIKEY is used to exchange keys and security parameter in a SIP environment with the help of SDP attribute fields. It also supports both public and symmetric key methods and optionally Diffie Helman (DH). Session keys are derived for encryption and integrity checks from a single “master key” which is identified by the MKI field. The derivation process also supports re-keying which basically refreshes or renews the session keys periodically to prevent attacks due to static keys.

4.4 Firewalls

Corporate network designers make it a point to regulate the flow of packets in-bound and out-bound of the network for fear of such attacks discussed prior. In effect, designers place a firewall to protect the trusted but vulnerable network from hostile public environment such as the internet. This protection strategy has normally been implemented by configuring the firewall to close all ports that are not being used and leaving just the ones that will be critical to the operational goals of the organization. An organization that heavily relies on the internet is sure to open port 80 for HTTP or web traffic.

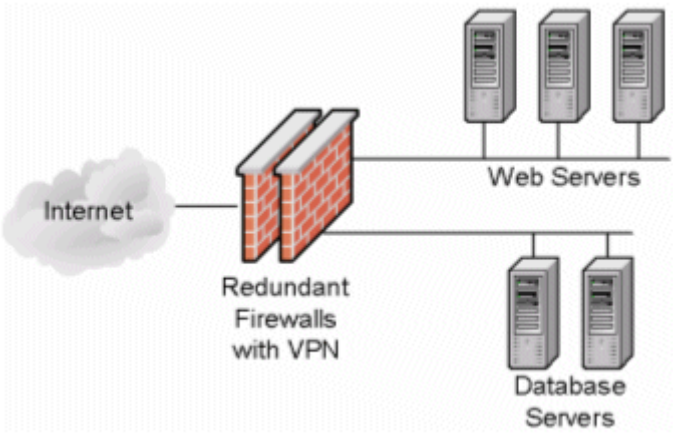


Figure 4-8: Packet filtering by a border firewall [68]

To really tackle traffic security effectively, administrators and designers would have to do more than port restrictions by using a firewall technique to inspect the content of packets based on a set of rules. These rules, when configured on routers, make them function as packet filters which become the entry point of every network. This is illustrated in Fig. 4-8. Typical filtering rules rely on IP header information to help the border router determine which packets to admit into the private network and which ones not. In other words, filtering can be seen as a check against the weaknesses in the TCP/IP stack. An example of a filtering rule can be based on the following:

- Network interface
- Source and destination IP address
- Option of the IP protocol
- The message type of ICMP message
- The ACK for TCP connections

The packet inspection process with such a rule would expect that, out-bound packets have destination addresses which are not private (such as 10.x.x.x and 192.168.x.x) since, private addresses can only be used internally or addressed only internal interfaces. Packets containing information in the option field set to source routing could be dropped to prevent a spoofed IP attack. Denial of service attacks in the form of ICMP “ping of death” where an attacker sends huge packets of size larger than 64k bytes to crash operating systems, or spoofed attacks resulting from incomplete TCP three-way handshake process, can be prevented by inspecting the ICMP messages and SYN/ACK bits in a packet.

4.4.1 Application Proxy servers

Another way for private networks to achieve increased security from public internet is to place an application proxy server between the corporate private network and the internet. The proxy servers in conjunction with a firewall server help the internal client nodes to receive from or forward traffic at the application layer. In the case of packet filters, filtering was done at transport and network layer.

In proxy oriented network illustrated in Figure 4-9, the host running the proxy is the only recipient of public IP address in addition to internal interface. The rest of the nodes in the private network assume a private address. When an internal host requests or receives traffic, a system called network address translation (NAT) modifies at the proxy server host. If the packet is out-bound, the source address in the header is substituted with the address of the proxy server whiles the destination address remains the same. In-bound traffic always has the proxy's as the destination address and by the same transformation. The packet is redirected to the host which requested the packets. The risk with such a network design is that when the server is compromised by attackers the whole network becomes exposed.

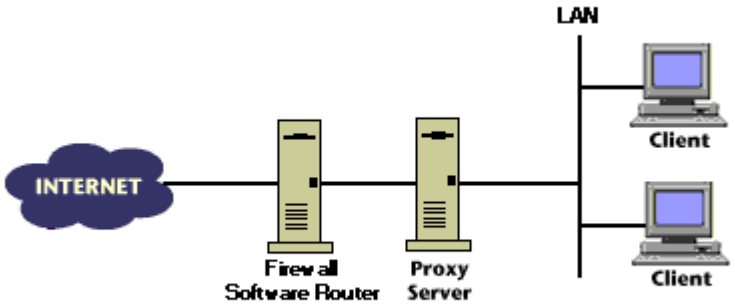


Figure 4-9: Proxy Server Implemented Behind a Screening Firewall [67]

4.4.2 Security Strategies

Securing an enterprise network has never been easy. Network administrators and designers tackle this challenge by first defining the scope and scale of the system to be protected, network capabilities, and deciding which traffic or protocol are allowed for both outbound and inbound traffic.

This information should pave way for the components of the network to be better organised with the help of a firewall security strategy such as:

- User right restriction to the highest extent
- Multiple security layers
- Little or no revelation of information about the system
- Security through obscurity

Managing or administering firewalls are critical to ensuring that the right services and ports are opened. This also means a physical access control policy to restrict the firewall administration to particular users and system accounts such as the *root* user on Linux and the *administrator* user on Window NT systems. Firewall architecture based on one host configured to protect the inner network form the outer un-trusted network stand the risk of exposure of the entire inner network when the host is compromised. One strategy to beef up internal security is to structure the network into layers with each sub layer requiring a different set of security mechanisms. The layers can be realised in the form of subnets where they are categorised into highly sensitive and low sensitive. The high network obviously receives more security than the low sensitive network which is assigned to ordinary users. This proves effective against trust-based attacks where users could be the target of social engineering techniques. Protecting the highly sensitive subnet may need two different filters with stricter rules on the second one or by alternatively using them for the same traffic. In so doing, failure of the first mechanism will not expose the entire system.

Information gathering is a critical stage in an attacker's feasibility study of a target system's compromisability. The attacker counts on some network information such as server name and version and type of operating system, to plan a successful attack. In this regard, it is wise to reveal no or as little as possible information about a system. Web browsers are normally culprits of this strategy as they tend to reveal web server type and version information. The principle of security through obscurity which believes that, internal vulnerabilities that are kept secret can not be exploited, have been the popular line towed by companies pursuing proprietary standards. Even though it is quite obvious that, the less an attacker knows about a system, the longer it takes for the system to be compromised, its better when this mechanism is used in conjunction with others. However, techniques such as reverse engineering have eroded support for this philosophy. An example of this exists in cryptography where proprietary ciphers like RC4 and A5 [21] were leaked and made public knowledge.

4.4.3 Dynamic packet filtering

In a scenario where VoIP deployment would mean opening at least two datagram (UDP) or connection oriented (TCP) transport service ports per session, the question of port restriction resurfaces. The danger is, since on a large network, the number of communication will grow, that will demand more ports to be opened for RTP streams.

This presents a perfect opportunity for attackers to exploit the network. Firewall vendors like *Cisco* have solved this problem by opening the ports dynamically during call setup and closing them after the session is over. In this context, firewalls operate through the "stateful" inspection of packets. This architecture is based on the active monitoring of outbound request packets in correspondence with inbound response packets. Only proper response packets are admitted into the internal network. Unlike static packet filtering, where inspection is done at only the header level, dynamic filtering extends inspection right to application layers.

Recent dynamic port firewalls include *Cisco*'s *PIX* firewall [9] which is capable of understanding signalling and bearer traffic protocols such as SIP and H.323. The *PIX* firewall will be suitable choice in an SIP network in a proxy mode. It also deals effectively with the eye-brows raised on performance issues such as additional delay and increase in jitters which does arise from the intense call processing.

4.5 Virtual Networks in VoIP

This section discusses two ways of improving security internally and externally in an enterprise network environment. The first way supports the segmentation of converged networks into priorities or categories to achieve higher security and better quality of service delivery by the use of switches. The second one ensures security beyond the borders of the internal network using virtual private networks (VPN) based on IPSec [58] in different modes.

4.5.1 Virtual Local Area Networks (VLANs)

The deployment of campus or corporate local area networks has become a popular medium for sharing common network resources. These resources range from confidential data to real-time applications such as VoIP. Protection through restriction of resources on the originally hub-based LANs proved to be ineffective since hubs had no mechanism for domain broadcast packet restriction. This meant that, in a LAN, any unauthorised host or IP phone plugged into a live port can access sensitive information or eavesdrop on a conversation.

This scenario raises questions as to why voice streams should suffer the same attacks that data networks are vulnerable to. In such a situation the solution that may comes to the mind of most network designers is to deploy VoIP on a dedicated network separate from data network. However, this suggestion would be counter to the main principal advantage of VoIP which is convergence [50].

Instead of dedicating a separate network for VoIP, a virtual local area network (VLAN) design could be used to group traffic into data, voice and signalling in a single infrastructure. This time round, security is enhanced by switches that support IEEE 802.1Q and 802.1p [36] standards for VLANs and quality of service (QoS) to provide the following:

- User restriction through port based or dynamic membership
- Traffic prioritization based on port and VLAN ID
- Effective bandwidth management and security

Switches optimise bandwidth usage by restricting broadcast frames within their intended VLAN ports instead of the entire network which in effect creates congestion –in the case of hubs. Besides, unicast frames only arrive at the destination ports of the target device. Communication amongst switches and between network devices are made possible by the tagging or un-tagging of switch ports. Tagged ports usually send tagged frames containing VLAN information only meant for 802.1Q compliant devices. These tagged frames link one switch to other by carrying membership information and in turn achieving network scalability. Untagged frames on the other hand, are meant to be handled by non 802.1Q devices such as PCs and IP phones whose network interface cards cannot process tagged frames.

Membership of devices on a VLAN can be configured statically or dynamically on a switch. By static means, devices belonging to a specific VLAN can be assigned a range of ports. For example, port 1-6 could be mapped to VLAN 1 and ports 7-12 to VLAN 2. Devices can then move to different VLANs by just switching ports. Dynamic membership preserves a device's VLAN assignment regardless of the port they plug into on a switch. This is done by assigning VLANs based on device IP or MAC address.

4.5.2 IPSec Virtual Private Networks (VPN)

IPSec [58] optionally extends the capabilities of IPv4 to add security to the network and upper layer protocols. IPSec is a standard architecture developed for IPv6 and defined in RFC 2401 [51]. In solving the pertinent issues in IPv4 such as message origin authentication, IP address spoofing and replay attacks, IPSec uses two different protocols. They are the authentication header (AH) [37] and the encapsulated security payload (ESP) [40]. Basically, AH and ESP ensures authentication, integrity and confidentiality in a communication.

IP datagram integrity is protected using hash message authentication codes (HMAC). The IPSec protocols use hash algorithm like MD5 [38] and SHA [39] in deriving secure hashes based on a secret (negotiated during key exchange) and the contents of the IP datagram. Confidentiality of IP datagram is provided by standard based symmetric encryption algorithms. IPSec supports DES, 3DES and AES [21].

IPSec protocols protect against denial of service (DoS) attacks through sliding windows where traffic flow is regulated using sequence numbering [45]. This means that sequenced packets are accepted only when the packet is within a set window size. In this way, replay attacks are prevented by dropping or discarding older or already accepted packets.

Protecting a communicating channel also means storage of security parameters on participating hosts. The security parameters are used to perform the security functions needed by AH and ESP. They are in turn stored in a security association database. The security parameters contained in the security association include [45]:

- Source and destination IP addresses of the resulting IPSec header. These are the IP addresses of the peers protecting the packets.
- IPSec protocol (AH or ESP)
- The algorithm and secret key used by IPSec protocol
- Security parameter index (SPI) which identifies the security association.

Secure exchange of cryptographic keys used in encryption and key hashing are ensured using the Internet Key exchange (IKE) protocol [41]. IKE first authenticates communicating devices (using certificates) and then negotiates security parameters between them. IKE uses Diffie Helman Key exchange [21] for deriving a secret symmetric key and offers a mechanism for re-keying the keys to ensure increased confidentiality.

The AH and ESP can be used in two different mode to ensure security in VPNs. They are transport and tunnel mode security association. The AH guarantees integrity protection of an IP packet and some unchangeable parts of its header such as the IP addresses. This in itself preserves message origin authenticity and eliminates spoofing attacks. ESP provides integrity in addition to confidentiality which is based on payload of the datagram. This normally does not include the IP header.

In a VoIP network environment, an IPSec VPN can be used to achieve end-to-end security between communicating devices as in the case of transport mode. A tunnel mode which is widely favoured and deployed will be suitable for establishing security associations between VoIP gateways of two private network or corporate sites over un-trusted networks such as the internet. The tunnel mode basically re-encapsulates an IP packet from an internal network with a new IP packet to reflect the gateway as the originator (source). Security is then applied to the packet and then sent to the receiving gateway for de-capsulation. Fig. 4-10 and Fig. 4-11 presents the security associations in both transport and tunnel modes.

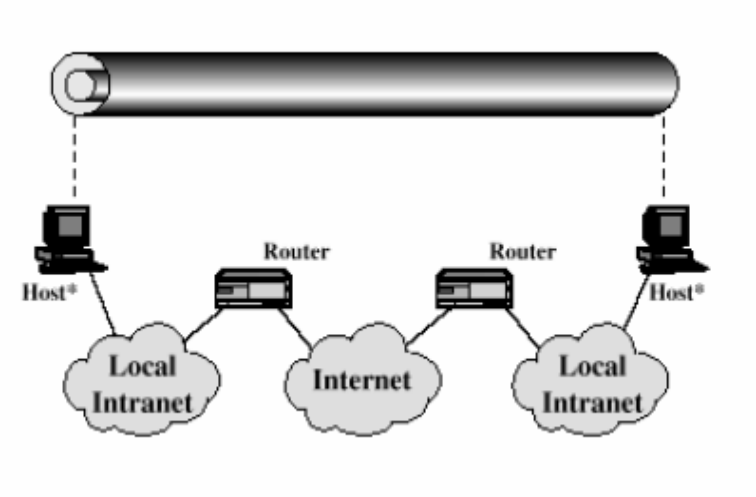


Figure 4-10 Transport mode security association [46]

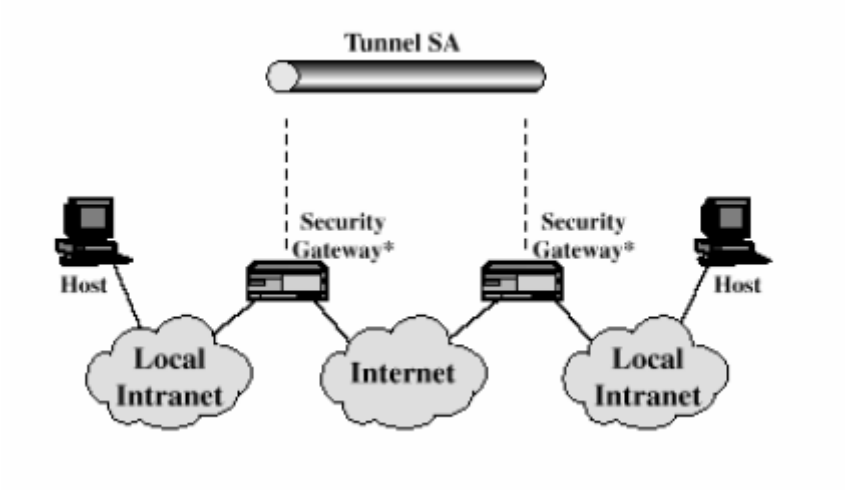


Figure 4-11 Tunnel Mode Security association [46]

4.6 IPv4 versus IPv6

Next generation IP or the popularly known IPv6 [58] defined in RFC 2401 was the answer to the security, scalability and quality of service (QoS) concern issues inherent in IPv4 networks. Though IPv6 is not widely deployed, it offers a better alternative solution to replacing IPv4 in following ways:

- Larger address space
- Address auto-configuration
- Built-in security
- Better quality of service

The concern over limited IPv4 addresses was solved in version 6 by increasing the source and destination length from 32 bits to 128 bits. With such a huge address space now available, address conservation techniques such as network address translation (NAT) [45] would no longer be needed.

Hosts in IPv4 networks can be configured manually or through a DHCP [45] server. In IPv6, none of these methods are required even though an improved version DHCPv6 [45] is supported. The improvement comes in the form of auto-configuration where, hosts on a link configure themselves automatically with link based addresses and pre-fixes assigned to them by their local routers. The IPv6 addresses are not only lengthy but formatted into hexadecimal numbers like FD04:F9:84:AB5:1C7:36:7AD:2B8 as opposed to decimal formats in the case of IPv4.The IPv6 protocol suite comes with a mandatory support for IPSec.

This equips IPv6 networks with in-built standards-based solution to security which is discussed at length in the previous section, such as [46]:

- Access control
- Connection integrity
- Message origin authentication
- Rejection of replayed packet
- Confidentiality

Packet delivery in IPv4 was mainly based on “best effort” where the network layer tries to send the packets. In addition to that, the *TOS* field for prioritizing and delivering quality of service to packets offers limited capabilities especially to real-time or time-sensitive technologies like VoIP. IPv6 inherently supports a better QoS.
The much better QoS guarantees in IPv6 may come with its trade offs since IPv6 has 40 bytes of header size to process as compared to the lighter 20 byte IPv4 header. The result is a much improved QoS in terms of reduced jitters and congestion but minimal latency assurances. An effective compression of the IPv6 header would be a work-around this issue [52].

5 Pitfalls and Strengths of VoIP

The gold rush for VoIP has certainly eclipsed the concerns of security threats coming from hackers. This initial trend which is understandable is still supported by surveys (2006) such as ZDNet research which indicate that, “the number of residential VoIP customers more than tripled to 4.2 million users in 2005, and is expected to hit 18 million by 2008” [8]. However, with the help of heightened security awareness, recent surveys seem to point to a seismic shift in the apathy towards security. Despite this change in attitude, survey reflects a reality that security concerns are not topmost. This means that, operators and end-users would rather be more enthused with the cost cutting benefits and the attractive features that the technology promises to bring.

Businesses normally gain from these benefits directly through financial rewards or indirectly through improving employee and customer satisfaction which in turn increase productivity. James Coggins (2004) [14] preferred to observe and categorise the benefits as “hard” and “soft”. His view was that, in typical hard benefit scenarios such as replacement of old traditional infrastructure, cost savings are quantifiable. The fruits of soft benefits on the other hand, are realised later in the form of work comfort, efficiency and maximised input which tend not necessarily quantifiable. Even though VoIP’s upward trend will mean a nightmare for fixed and mobile operators, quality of service (QoS) and interoperability issues of VoIP will help curb the mass defection.

This chapter discusses the rewards and challenges a VoIP deployment can bring to businesses and end-users who would wish to migrate from traditional PSTN. It closes with suggestion as to how best to improve security in different attack scenarios discussed in chapters 3 and 4.

5.1 VoIP benefits

5.1.1 Cost Savings

Like any business venture or investment, the goal is to minimise cost and maximise profit. In that regard, a business venture involving VoIP will be no exception. In addition to towing the line of this philosophy, businesses have been quick to know that, an overnight transition to VoIP is not the answer to reduced cost or a return on investment, for after all, the initial capital of deploying a viable infrastructure could be daunting which is contrary to the hype surrounding. In the end, the answer lies in Coggins suggestion to treat VoIP projects “as an investment” [14]. He proposes that, companies could spread cost of investment over a period by which time legacy equipments would gradually be phased out or used as a stand-by. Moreover, companies that are deterred by operating cost could alternatively opt for the services of equipment-leasing providers.

The fruits these practices should account partly for a Delloitte (2004) survey which suggests that, "two thirds of global business will deploy voice over IP by 2006" [15]. This is a surge which cites reduced cost as the overwhelming driver. In the US, ZEDNet survey (2005) showed "59% of new VOIP users made the switch because of the costs" [8]. Reduced cost or savings contributed by VoIP deployment can not only seen through equipments but also less network administration cost due to reduce size management staff and cost of running a single infrastructure. The contribution from these three various VoIP departments to cost savings can be broken down into percentages illustrated in a pie chart in figure 5-1.

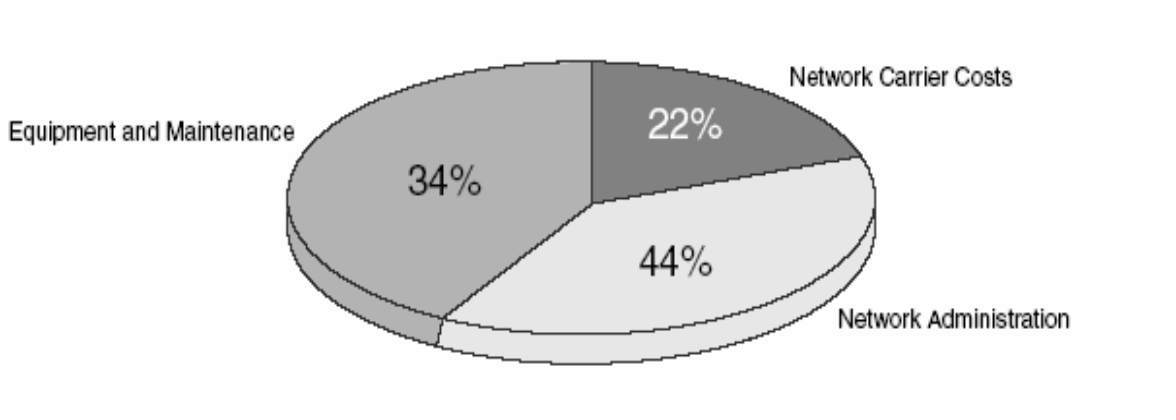


Figure 5-1: Contribution to VoIP cost cutting [14]

5.1.2 Long distance call

The inception of VoIP brought along the much hyped free call service. The move was music to the ears of businesses and telecom operators and especially third world countries who saw the technology as a last ditch effort to bridge the gap of the ever widening digital divide that exists between them and western world. Enthusiasts from both sides of the digital aisle, noticed that, by carefully facing out gradually PSTN, VoIP would be very instrumental in reducing phone bills emanating from long distance calls to or from international location.

These locations could be company branches or client sites across the world. The principle behind such a cost reduction or free calls hinges on two contributory factors. The first is that, VoIP is built on an IP network and as such, it's bound not to see distance as an issue besides, once a user has been assigned a bandwidth by an internet service provider-ISP, VoIP operation can commence. The same cannot be said about PSTN systems since they are owned by the operators and using the network comes with a fee. The second factor goes further to reveal that, by sending voice traffic over the internet instead of a PSTN switch, the cost arising from usage of PSTN can be avoided. This phenomenon is normally referred to as toll-bypass which in a sense, means by passing cost. How this network feature will help businesses especially the ones with multiple branches or sites is technically illustrated in Fig. 5-2. In Fig 5-2, toll bypass is achieved replacing the PBX to PSTN link with an IP PBX, VoIP gateway and internet connection.

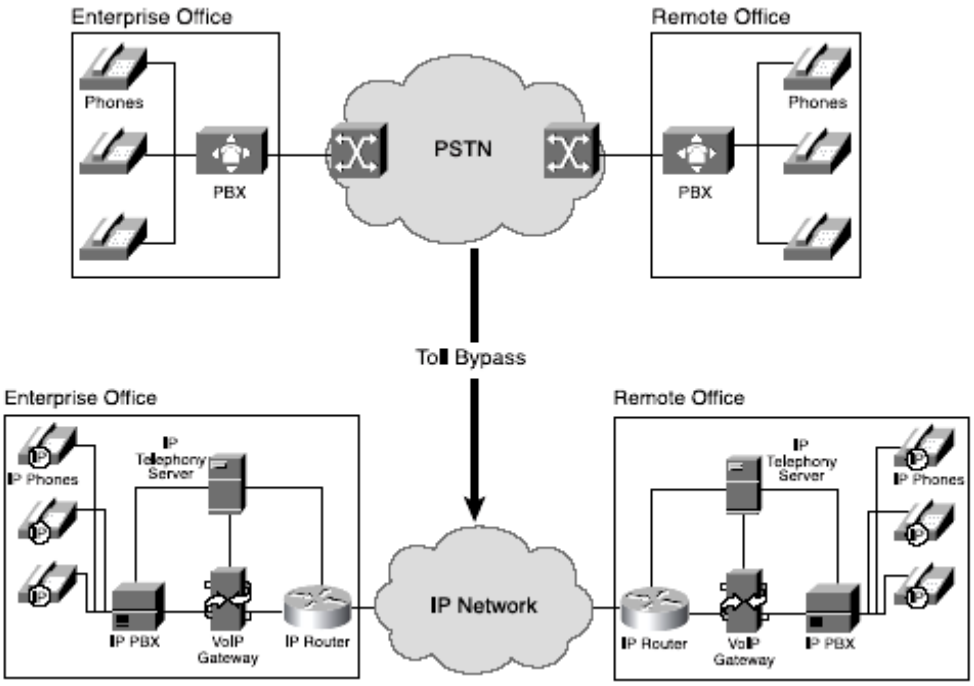


Figure 5-2: Toll by-pass [14]

5.1.3 Features

The battle between PSTN and VoIP sympathisers over undecided customers and investors will be settled overwhelmingly in favour of VoIP when designers and operators unleash the much awaited killer application and new features as promised by the telephony hype. Businesses that adopt these killer applications stand to gain by using it as part of their marketing strategy to provide satisfaction for both staff and customers. Satisfaction and comfort such as being able to work from home in the comfort of value added applications, then brings about increased productivity in addition to increased market share and company image. Ernest Nicastro (1998), a marketing strategist, puts this view into a quantifiable perspective when he observes that, “a 5% improvement in customer retention can lead to an 85% increase in profits” [16].

One such breath-taking application making the headlines is the Unified messaging-(UM). As the name implies, UM assembles all messaging media types into a central processing unit for retrieval. These media types include SMS, e-mail, fax, and voice mail. Integrated with the central unit or unified mailbox are special technologies such speech-to-text and text-to-speech which are used to provide the luxury services to employees and customers alike.

In a typical UM application users can retrieve information from the unified mailbox using any interface or device. This could be mobile terminal or a computer. So for instance, when a user receives an e-mail and he is detected to be on the phone, an SMS message is sent to alert him. If he decides to retrieve the message there and then, the text-to-speech technology is used to convert the e-mail content to voice for his hearing. Figure 5-3 captures the different components including, voice messaging, e-mail, and fax that make up a Unified Messaging system which shares a common mail-box.

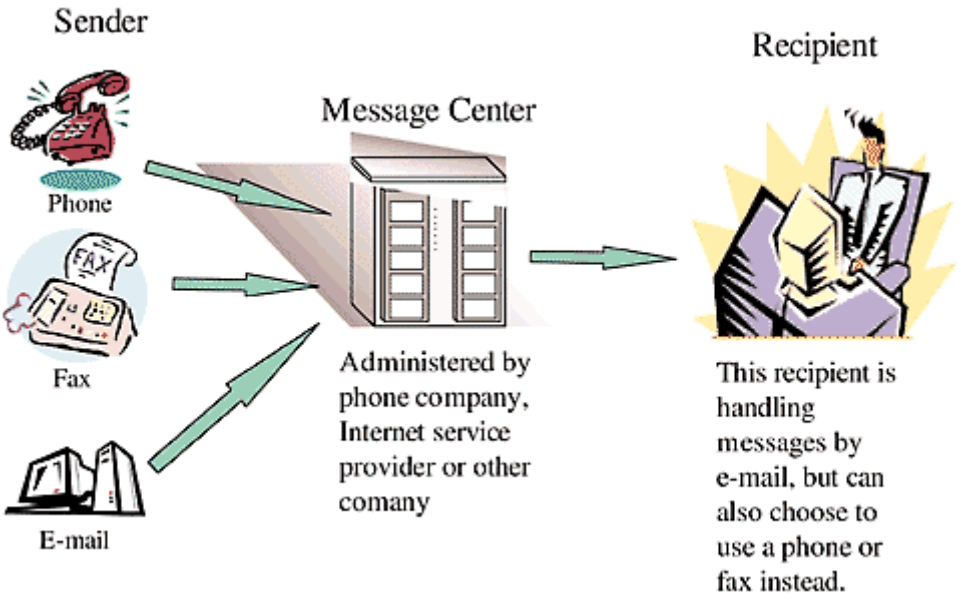


Figure 5-3: Basic concept of Unified Messaging system [69].

5.1.4 Management of Single infrastructure

Cautious investors who are not taken-in by the VoIP hype normally deploy VoIP along-side PSTN service. They tow such line in the believe VoIP will eventually mature in years to come or maintain their circuit switched networks for emergency situation particularly when an IP infrastructure brakes down. The problem is, such operational scheme of managing both data network and circuit switched may not be cost effective in the long run. This should cease to be the case when management bravely and fully adopt VoIP with minimal PSTN presence. This move paves the way for reduced cost of managing a pruned employee staff assigned to a single convergent IP infrastructure. However, this come with its trade-offs in the form of increased training of staff of both data and circuit switched networks to be conversant with both technologies.

5.1.5 Convergence

Although traditional telephony still maintain a sizable market than their IP telephony counter parts in the telecom business, there is every indication that, its only a matter of time when the VoIP status quo will change from "mind share" stage to become a dominant force in the global telecom market. This whole hearted and mind acceptance might be due to the concept of convergence. This goes to suggest that, instead of traditional telephony businesses and enterprises going through the hustle of high bandwidth requirements and system upgrade due to expanding customer base and capacity that come with it, they would prefer to adopt a single infrastructure built on an IP network which actually guarantees efficient use of bandwidth through compression and suppression technologies and hence requires minimal maintenance or upgrade. This view is well echoed by Cisco's survey (2001) in figure 5-4 which clearly shows the direction businesses are headed in terms of deployment of convergent network deployment.
In addition to cost savings that immediately comes to mind, convergence extends VoIP service capabilities to include a wireless infrastructure using IEEE 802.11 technologies which will definitely mean a remarkable reduction in spending arising from wiring and cable.

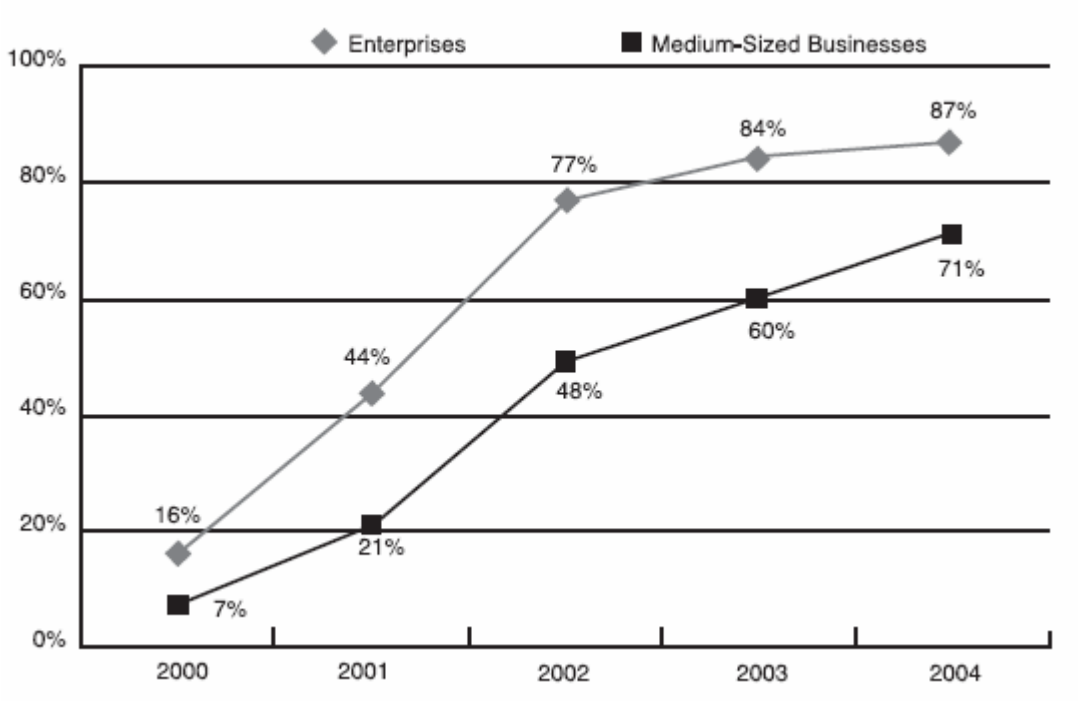


Figure 5-4: The trend of convergent network deployment [14]

5.2 Challenges of VoIP

5.2.1 Capital expenses

The migration from tried and tested dedicated PSTN networks to the still evolving VoIP has sometimes been met with mixed reactions. Whereas some businesses are quick to pounce on its cost cutting potentials, others have decided to approach this crossover with cautious optimism. This second reaction sees VoIP as a risky venture with unresolved future. At present, this assertion is true in that, businesses adopting VoIP have done so not because of business driven reasons but technology driven reasons. In effect, businesses rush into VoIP deployment with less preparation. The punishment of ill-preparation always shows up in high initial cost of equipment maintenance or upgrade and cost of staff training on a single network infrastructure. In addition to these problems, businesses would have to contend with un-standardised nature of vendor products which are not interoperable with others. The way out of these challenges is for organisation to understand and map up a business strategy that is based on technology and sound businesses practises that work hard in hand.

5.2.2 Quality of Service-QoS

After taking so much beating in terms of technological and business potential, the only way PSTN does redeem itself is through the naughty question of quality of service. The circuit-switched oriented PSTN networks are very popular for guaranteeing superior service quality when it comes to proper noise, tones and appropriate loudness levels in the circuit. They same can't be said about quality of service in an underlying IP network in a VoIP infrastructure. The faulty TCP/IP suite in itself bodes ill for prospects of increased quality of service in addition to the fact that, voice packets are preferably run over datagram service. This situation does extend to mean that security levels in PSTN will be higher than their VoIP counter-parts. This foregone conclusion is given ground reality credence by US based *ZEDNet* survey (2006) which revealed that "48% of small and medium businesses trust IP telephony security today while 76% said they trust the security of traditional service providers" [8].

To match QoS standard offered by traditional telephony, IP would have to sufficiently and effectively deal with scenarios such as packet delivery delay due to congestion or queuing, packet lose probably due to inadequate router buffer space or packet arriving out of order at their destination. Fortunately IP provides a two-pronged solution in tackling QoS. The first one intends to help routers in the network process packets according to the 8 bit type of service *–TOS* information provided in the IP header. The *TOS* field is divided into 2 hierarchies.

The first three bits are for precedence and the four out the five remaining are for type of service. The precedence bits indicate the priority of a particular datagram whiles, the type of service bits *D*, *T*, *R*, and *C* indicate either normal or high delay, through-put, reliability and cost of the datagram. In a typical real-time critical VoIP application network services should be seen guaranteeing low delay and higher priority while voice communication can't be bothered by some level of packet loss. The second part presents two philosophies as how operators and clients can agree on a suitable QoS depending on their needs and purposes. This agreement normally leads to a contract or service level agreement which tend to be differentiated service oriented (*Diffserv*) or an integrated service oriented one (*IntServ*).

Integrated Service-IntServ

In an integrated oriented service, QoS is reserved for applications by its host using reservation protocol-RSVP which is defined in RFC 2205 [35]. This means each node or router gets configured with *IntServ* in order for any application to separately request for QoS. The reservation process starts when a host requests QoS from a node or router. This prompts the service of two control mechanisms which are admission and policy control. The admission control checks whether the node has enough resources to guarantee the requested QoS whiles the policy control checks the administrative rights the user has in making the reservation. If the both checks proceed successfully, a service is provided to classify packets into QoS categories and matches or schedules them respectively according the application demands. Figure 5-5 shows how the RSVP daemon coordinates reservation between the application and the mentioned modules.

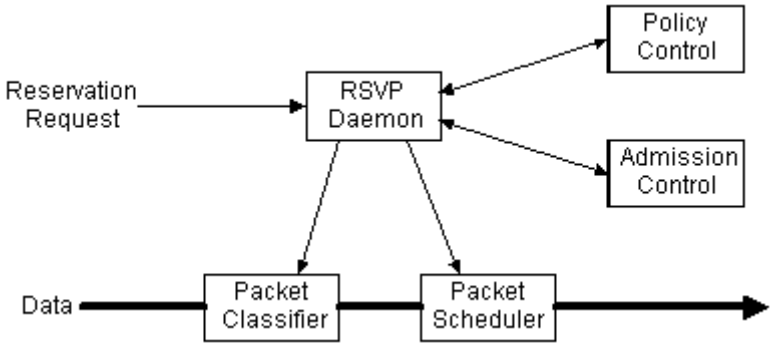


Figure 5-5: RSVP packet processing modules [65]

Differentiated Service

Unlike integrated services which guarantees QoS on per application request bases, differentiated services promotes QoS in bulk on a network scale by contracting a service that will apply to every packet in the network. Resulting contracts categorise traffic in classes and the corresponding guarantees. Classification is provided by the *TOS* information from the packet sender. In other words, the higher the value of precedence, the better the service. For example the traffic could be categorised into three classes namely *High*, *Medium* and *Low*. The *High* traffic takes precedence over *Medium* which takes precedence over *Low* traffic. Additional regulation of traffic based on their classes could include:

- Preferential forwarding, where more recent higher precedence packets are allowed to jump the queue over old lower precedence packets;

- Preferential discarding, where buffer space for higher-precedence packets is allowed to grow at the expense of lower precedence packets which are discarded.

Unlike data networks, VoIP networks experience significant reduction in service quality during congestion or fluctuation. To solve this possibility is to structure the network with methods to manage QoS levels to guarantee uninterrupted service during peak times for real-time communications. These methods include the following:

- Reserving fixed bandwidth for mission-critical voice communication applications;
- Restriction on network access and usage for defined users or user groups;
- Utilization of virtual LANs (VLANs) to separate voice and data and;
- Designation of which kinds of traffic can be dropped when congestion occurs.

5.3 Security in VoIP

Viega and McGraw (2001) have helped draw conviction from the conceptual systems theory perspective where, Ashby's law [24] of requisite variety give clues as to how best to deal with vulnerabilities within and threats from attackers in a complex system such as VoIP. This law basically suggests that, if there are multiple ways in which a system's security can be compromised, then ensuring security would also require multiple security mechanisms. In this regard, VoIP employs state-of-the-art security tools and mechanisms to prevent system compromise. The core mechanism used include compression and media encoding for effective bandwidth utility, secure and strong encryption provided by RSA, 3DES and AES for privacy and key exchange, secure hashing provided by HMAC-SHA-1 for integrity and certificates for non repudiation. These mechanisms and others discussed in chapter 4 in effect supply ample "variety to kill variety" [24].

In other words VoIP security systems have the capabilities to deal with the imminent threats from would-be attackers from multiple scenarios. With all that in mind, I made the following observations:

- VoIP is built on a faulty TCP/IP design architecture with no security options.
- VoIP is mostly secure and trusted within enterprise environments.
- Benefits of convergent networks have the tendency to overshadow security.
- VoIP still has quality of service and interoperability issues to deal with.

Tackling the issues raised by these observations has by no means been exhaustive in this thesis since, VoIP still continues to evolve. However, recommendations tightening the screws on some key security areas will be a step in the right direction. These recommendations touch on human level activity, boosting network security, secure applications and quality of service improvements.

The human element of VoIP security received comparatively less focus even though it is the weakest link to VoIP security. With an effective management system, smart cards could be used to improve user authentication to VoIP equipments and provide a solid ground to build a case for or against non-repudiation.

The war on unauthorised access should be lead by host-based intrusion detection system and application access control which also double as auditing and logging tools essential for monitoring and trouble shooting. Application security will be better served by adopting around-the-clock virus scanning service with prompt signature updates to uproot virus and Trojan-horses. A viable add-on is to subscribe to or support open source applications which also help cut cost tremendously on proprietary software spending.

In the case of any confusion as to how be best to confront VoIP attacks from multiple attack scenarios, Fig. 5-6 should be helpful in clarifying how they could be mitigated.

Attack Type	Solution Mechanisms
Denial-of-Service	Server cluster services, IPSec, IDS, VLANs Black listing rogue sites, Enough bandwidth allocation
Registration Hijacking	Message Authentication (TLS), IPSec, IDS
	Secure Registrar Server- TLS, Firewall
Proxy Impersonation	Mutual Authentication(TLS), IPSec, IDS, Firewall
Eavesdropping	VLANs, IPSec
Man-in-the-middle	VLANs, SRTP, IPSec, static ARP tables [45]
Replay Attack	VLANs, SRTP, IPSec, static ARP tables, IDS
Buffer Overflow(Platform attacks)	Virus scanning, Update virus signature, Apply patches constantly, IDS, Firewall

Figure 5-6: VoIP attacks and possible solutions

6 Conclusion

As to whether VoIP systems are adequately secure, I am convinced that the presented security mechanisms provide VoIP with a holistic sense of security. Moreover, the solutions elaborated in the previous chapters are consistent with popular security tenets favoured by Viega and McGraw (2001) like defence-in-depth and the "keep it simple stupid" philosophy which secures systems from application and network layer right down to the every day human use.

Despite the much raised fears of VoIP, I am again confident that an organisational security policy that proactively educates and promotes the discussed security solutions will surely yield any organisation the kind of rewards that VoIP promises to bring. In the nut shell state-of-the-art security mechanisms and the massive rewards of a successful implementation are enough to rescue organisations from the threats to VoIP which tend to be blown out of proportion. In this conclusion, I am joined by Ahmar Ghaffar (2004) who rightfully opines that, "VoIP security sounds like a nice idea and definitely makes the telephony environment more secure, thereby gaining end-user confidence. But it's certainly not a hurdle preventing VoIP from making it big in the telecom industry" [12].

6.1 Further studies

This paper focused entirely on securing routed protocols to the neglect of routing protocols used by both intra and inter-domain routers for path determination. Routers use either distance vector protocols such as Router Information Protocol-RIP or link state protocols such as OSPF (Open Shortest Path First) to update routing information. Both distance vector and link state update messages are vulnerable to attacks which results in modification. Integrating security in routing protocols would provide authentication, confidentiality, and integrity of routing information in an entire converged network. A further study into secure routing protocols should spell out how the mentioned security requirements can be achieved through mechanisms like public key digital signatures and its corresponding challenge in authenticating (key management) each router in large network deployment.

6.2 Challenges

In the absence of real practical VoIP experience in the telecom industry, I found myself restricted by the qualitative or the less experimental nature of this thesis. This meant that I had to rely heavily on standard documentation such as IETF's RFCs for some of the very technical issues of VoIP.

Glossary

ABNF	Augmented Backus-Naur Form
PSTN	Public Switched Telephone Network
SS7	System signalling number 7
TDM	Time Division Multiplexing
VoIP	Voice over Internet Protocol
IP	Internet Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UAC	User Agent Client
UAS	User Agent Server
TOS	Type of Service
ISP	Internet service provider
RTP	Real Time Protocol
SIP	Session Initiation Protocol
H.323	ITU multimedia signalling Protocol
RSVP	Reservation Protocol
SSL	Secure socket layer
TLS	Transport Layer Security
SRTP	Secure Real-time Protocol
HTTP	Hyper Text Transfer Protocol
SMTP	Simple Mail transport Protocol
QoS	Quality of Service
MAC	Media access control
MAC	Message authentication code
ISDN	Integrated Service of Digital Network
Q.931	Connection control protocol of ISDN
RAS	Registration Admission and Status
URL	Uniform Resource Locator
SDP	Session Description protocol
DES	Data Encryption Standard
AES	Advanced Encryption Standard
RSA	Rivest Shamir Adelman Algorithm
POTS	Plain Old Telephone Service
CSRC	Contributing Source
SSRC	Synchronizing Source
ITU	International Telecommunication Union
IETF	Internet Engineering Task Force
LAN	Local Area Network
DoS	Denial of Service
DNS	Domain name service
ISO	International standards organisation

OSI	Open systems interconnect
MTU	Maximum Transport Unit
NAT	Network Address Translation
MGCP	Media gateway control Protocol
IPSec	Internet Protocol Security
VLAN	Virtual Local Area Network
VPN	Virtual Private Networks
IEEE	Institute of Electrical and Electronics Engineers
AH	IPSec Authentication Header
ESP	IPSec Encapsulated security payload

References

Internet Pages and Whitepapers

[1] Public switched telephone network. Accessed on December 10, 2005 at URL http://www.networkdictionary.com/telecom/cs.php.

[2] Internet protocol fragmentation. Accessed on December 10, 2005 at URL http://penguin.dcs.bbk.ac.uk/academic/networks/network-layer/fragmentation/index.php

[3] SS7 Overview. Accessed on December 15, 2005 at URL http://www.techfest.com/networking/wan/ss7.htm.

[4] Public switched telephone network. Accessed on December 10, 2005 at URL http://www.imagen-interactive.com/assets/powerpoint/pstn_overview_files/frame.htm, http://en.wikipedia.org/wiki/PSTN.

[5] Pulse code modulation. Accessed on December 10, 2005 at URL http://en.wikipedia.org/wiki/Pulse-code_modulation.

[6] Time division multiplexing. Accessed on December 10, 2005 at URL http://en.wikipedia.org/wiki/Time-division_multiplexing.

[7] Voice over IP-Wikipedia, the free Encyclopaedia. Accessed on December 12, 2005 at URL:http://en.wikipedia.org/wiki/VoIP

[8] VoIP IT Facts. Accessed on February 13 2006 at URL:http://blogs.zdnet.com/ITFacts/index.php

[9] Cisco PIX firewall–Practical guide. Accessed on January 12, 2006 at URL http://www.enterastream.com/whitepapers/cisco/pix/pix-practical-guide.html

[10] Hoglund, G. & McGraw G. (2004). *Exploiting Software: How to Break Code, Chapter 7 -- Buffer Overflow*. February. Retrieved March, 2006 at URL: http://searchsecurity.techtarget.com/searchSecurity/downloads/ExploitingSoftware-Ch07.pdf

[11] 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

[12] Ghaffar, A. (2004) . *Is VoIP Secure*. Retrieved at February, 2006 at URL:http://www.tmcnet.com/voip/1104/FeatureSecurity.htm

[13] Krapf, E. (2002). *VOIP vs. Firewalls*. July 2004. Retrieved March, 2006 at URL:http://www.bcr.com/architecture/briefing/voip_v_firewalls_20020719288.htm.

[14] Coggins, J. (2004).*Building a Business case for VoIP*. Retrieved March, 2006 at URL:http://www.ciscopress.com/content/images/1587200929/samplechapter/1587200929ch2.pdf.

[15] Deloitte VoIP Survey. (2004). Retrieved March, 2006 at URL: http://blog.tmcnet.com/blog/tom-keating/voip/deloitte-voip-survey.asp.

[16] Nicastro, Earnest. (1998). *To keep customers, keep in touch regularly Puget Sound Business Journal (Seattle)*. Retrieved March, 2006 at URL: http://www.bizjournals.com/seattle/stories/1998/03/16/smallb6.html.

[17] Dalgic, I. & Fang, *H. Comparison of H.323 and SIP for IP Telephony Signalling*. Retrieved February, 2006 at URL:http://www.iptel.org/info/references/papers/misc/Dalg9909_Comparison.pdf.

Books

[18] Viega, J., & McGraw, G. (2001). *Building Secure Software: How to Avoid Security Problems the Right Way*. Addison Wesley.

[19] Bishop, M. (2003). *Computer Security. Art and Science (1st Ed)*. Addison-Wesley.

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

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

[22] Perlman, R. (1999). *Interconnections (2nd Ed)*.

[23] William, Terry. (2000). *Practical Firewalls*. Que Corporation.

[24] Yngström, Louise. (1999). *Systemic-Holistic Approach to IT Security*, DSV.

IEFT Request for Comments (RFCs)

[25] Defense Advanced Research Projects Agency. September 1981. *Internet Protocol*. Request for Comment 791, http://www.ietf.org/rfc/rfc0791.txt

[26] Defense Advanced Research Projects Agency. September 1981. *Transmission Control Protocol*. Request for Comment 793, http://www.ietf.org/rfc/rfc0793.txt

[27] J. Postel. *User Datagram Protocol*. August 1980. USC Information Sciences Institute. Request for Comment 768, http://www.ietf.org/rfc/rfc0768.txt

[28] Schulzrinne, H., Casner, S., Frederick, R. and Jacobson, V. January 1996. *RTP: A Transport Protocol for Real-Time Applications.* RFC 1889, http://www.ietf.org/rfc/rfc1889.txt.

[29] Schulzrinne, H. January 1996. *RTP Profile for Audio and Video Conferences with Minimal Control*. RFC 1890. http://www.ietf.org/rfc/rfc1890.txt

[30] Dierks, T., and Allen, C. *The TLS Protocol*, January 1999. Request for Comment 2246, http://www.ietf.org/rfc/rfc2246.txt.

[31] Baugher, M., et. al. March 2004. *The Secure Real-time Transport Protocol (SRTP).* RFC 3711,.http://www.ietf.org/rfc/rfc3551.txt

[32] Handley, M., Schulzrinne, H., Schooler, E. and Rosenberg, J. March 1999. *SIP: Session Initiation Protocol.* Request for Comment 2543, http://www.ietf.org/rfc/rfc2543.txt.

[33] Handley, M., and Jacobson, V. April 1998. *SDP: Session Description Protocol*. Request for Comment 2327, http://www.ietf.org/rfc/rfc2327.txt.

[34] Schulzrinne, H., and Casner, July 2003. S. *RTP Profile for Audio and Video Conferences with Minimal Control.* RFC 3551, http://www.ietf.org/rfc/rfc3551.txt.

[35] Braden, R., Zhang, L., Berson, S., herzog, S. And Jamin, S. *Resource Reservation Protocol (RSVP) - version 1 Functional Specification*. RFC 2205, proposed standard, September 1997. http://www.ietf.org/rfc/rfc2205.txt.

[36] IEEE Std 802.1Q, 2003 Edition, *Virtual Bridged local Area Network*.

[37] Kent, S. & Atkinson, R. (1998).*IP Authentication Header*. RFC 2402. http://www.ietf.org/rfc/rfc2402.txt

[38] Madson, c. & Glenn, R. (1998). *The Use of HMAC-MD5-96 within ESP and AH*. RFC 2403. http://www.ietf.org/rfc/rfc2403.txt.

[39] Madson, c. & Glenn, R. (1998). The *Use of HMAC-SHA-1-96 within ESP and AH*. RFC 2404. http://www.ietf.org/rfc/rfc2404.txt.

[40] Kent, S & Atkinson, R. (1998).*IP Encapsulating Security Payload (ESP).* RFC 2406. http://www.ietf.org/rfc/rfc2406.txt

[41] Harkins, D & Carrel, D. (1998). *Internet Key Exchange*. RFC 2409. http://www.ietf.org/rfc/rfc2409.txt

[42] ITU-T Recommendation H.323. (1999). *Packet-based multimedia communications systems*. http://www.itu.int/rec/T-REC-H.323-199909-S/en

[43] Groves, c., et al (2003). *Gateway Control Protocol Version 1*. RFC 3525. http://www.ietf.org/rfc/rfc3525.txt

[45] The TCP/IP Guide: *IP Security (IPSec) Protocols*. http://www.tcpipguide.com/free/t_IPSecurityIPSecProtocols.htm

[46] Muftic, S & Morogan, M. *Network Security*. (2004) http://dsv.su.se/~matei/courses/ICSS-2i1503/K1l2-IPSec.pdf

[47] Muftic, S & Morogan, M.*Wireless Network Security*. (2004). http://dsv.su.se/~matei/courses/4%20-%202i1279/K4L1.pdf

[48] *IP Private branch exchange*. Retrieved March, 2006 at http://en.wikipedia.org/IP-PBX

[49] *SIP Vs. H.323 - A Comparison*. Retrieved April, 2006 at http://microtronix.ca/sip_vs_h323.htm

[50] Webster John. (2006). *VLANs Maximize VoIP investments*. Retrieved April, 2006 at http://www.infoworld.com/article/06/01/23/73728_04FEvoipmanaged_1.html

[51] Thomson, s. & Narten, T. (1998). *IPv6 Stateless Address Autoconfiguration*. RFC 2462. http://www.ietf.org/rfc/rfc2462.txt

[52] Demir, O., Pham Q, & Whiting, *IPv4 vs IPv6: Benefits Of A New Technology*. http://web.dis.unimelb.edu.au/Ugrad/e/ewhiting/615-237%20Group%20Report/t07_0405.pdf

[53] ITU-T Recommendation X.25. (1996). Retrieved April, 2006 at http://www.itu.int/rec/T-REC-X.25/en

[54] Postel, J. (1981). *Internet Control Message Protocol*. RFC 792. http://www.ietf.org/rfc/rfc792.txt

[55] Ethereal: *A Network Protocol Analyzer*. Retrieved April, 2006 at http://www.ethereal.com

[56] *Security tools*. Retrieved March, 2006 at http://www.securityfocus.com/tools/category/95

[57] *VoIP Security Risks*. (2004).http://searchenterprisevoice.techtarget.com/searchEnterpriseVoice/downloads/VoIPsecurityChap7.pdf

[58] Kent, S. & Atkinson, R. (1998). *Security Architecture for the Internet Protocol*. RFC 2401. http://www.ietf.org/rfc/rfc2401.txt.

[59] Collier M. (2005). *Basic Vulnerability Issues for SIP Security*. Retrieved April, 2006 at http://download.securelogix.com/library/SIP_Security030105.pdf

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

[61] *Cain & Abel*. Retrieved April, 2006 at http://www.oxid.it/cain.html.

[62] *Soft Phones*. Retrieved April, 2006 at http://en.wikipedia.org/wiki/Softphone.

[63] *TCP 3-way Handshake*. Retrieved April, 2006 at http://www.inetdaemon.com/tutorials/internet/tcp/3-way_handshake.shtml

[64] Chambers, C. , Dolske, J and Iyer, J. *TCP/IP security*. Retrieved April, 2006 at http://www.linuxsecurity.com/resource_files/documentation/tcpip-security.html

[65] Csurgay,P. , Oesleboe, A. & Aagesen, F. *Teleservices and Internet application technology*. Retrieved April, 2006 at http://www.item.ntnu.no/~arneos/publications/nips.html

[66] *Security Features on Cisco Integrated Services Routers*. Retrieved April, 2006 at http://www.cisco.com/en/US/products/ps5855/products_qanda_item0900aecd80169bba.shtml

[67] *Netscape Proxy Server Deployment Guide*. Retrieved April, 2006 at http://wp.netscape.com/proxy/v3.5/using/

[68] Northrup, T. *Firewalls*. Retrieved April, 2006 at http://www.microsoft.com/technet/security/topics/networksecurity/firewall.mspx

[69] *IEC:Unified Messaging*. Retrieved April, 2006 at http://www.iec.org/online/tutorials/unified_mess/topic02.html