ENHANCING SECURITY PROTOCOL FOR VOIP COMMUNICATION USING MODIFIED VECTOR QUANTIZATION

MOHD YAZID MOHD JAAFAR

UNIVERSITI SAINS MALAYSIA

2013

ENHANCING SECURITY PROTOCOL FOR VOIP COMMUNICATION USING MODIFIED VECTOR QUANTIZATION

by

MOHD YAZID MOHD JAAFAR

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

July 2013

ACKNOWLEDGEMENTS

First and foremost, I am deeply indebted to my supervisor, Associate Professor Azman Samsudin, for sharing his passion, knowledge and guidance during my research at School of Computer Sciences, Universiti Sains Malaysia. His enthusiasm in work has always motivate me, and his effort is so much to be admired. Thank you so much.

I want to thanks all of the staff in School of Computer Sciences, Universiti Sains Malaysia, particularly Mrs. Azlina Yusof and Mr. Redzuan Asmi. With their help, I manage to get through all my candidature matters with ease.

I would like to take this oppurtunity to thank my beloved parents, Hj. Mohd Jaafar Hj. Abdul Gani and Hjh. Fulanatin Hj. Mukri, as well as my sisters Umi Mahmudah and Siti Maisarah. Their unlimited support and true love had keep my faith and spirit on top everytime I face a tough time. They stand by me, raised me, supported me, tought me, and love me. To them I dedicate this thesis.

I would like to thank the person who share my happiness and saddes. The person who supported me and provided me with a caring environment and unforgettable moments. Thank you Danya Ayesya Abdull Razak. To my friends, Alfin Syafalni, Iqmal Rahiman and Hafiz awang, thank you for coloring my life. Hope we can achieve together whatever things that we always imagined.

Mohd Yazid Mohd Jaafar

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AES Advance Encryption Standard

CA Certificate Authority

DES Data Encryption Standard

DH Diffie-Hellman

DLP Discrete Logarithm Problem

DTLS Datagram Transport Layer Security

ECC Elliptic Curve Cryptography

ECDH Elliptic Curve Diffie-Hellman

ECDLP Elliptic Curve Discrete Logarithm Problem

IGF Image metric Generation Function

IV Initial Value

MAC Message Authentication Code

MIKEY Multimedia Internet Keying

MITM Man In The Middle

MVQ Modified Vector Quantization

NP-hard Non-deterministic Polynomial-hard

PKC Public Key Cryptography

PKI Public Key Infrastructure

PRNG Pseudo-random Number Generator

QoS Quality of Service

RA Registration Authority

RSA Rivest-Shamir-Adleman

RTCP Real-time Transport Control Protocol

RTP Real-time Transport Protocol

S/MIME Secure/Multipurpose Internet Mail Extension

SAS Short Authentication String

SDES Security Description

SDP Session Description Protocol

SIP Session Initiation Protocol

SRTP Secure RTP

TCP Transmission Control Protocol

TLS Transport Layer Security

TTP Trusted Third Party

UDP User Datagram Protocol

VA Validation Authority

VoIP Voice over Internet Protocol

VQ Vector Quantization

ZID ZRTP ID

ZRTP Zimmerman's RTP

MEMPERBAHARUI PROTOKOL KESELAMATAN UNTUK KOMUNIKASI VOIP MENGGUNAKAN VEKTOR KUANTISASI TERUBAH

ABSTRAK

VoIP telah menerbitkan cabaran baru yang tidak pernah didengar ketika talian telefon tetap masih digunakan. Protokol Permulaan Sesi (SIP) biasa digunakan sebagai protokol utama dalam VoIP. Namun, tiadanya kemudahan sekuriti telah mendedahkan SIP kepada banyak ancaman rangkaian. Infrastruktur Kekunci Awam (PKI) digunakan sebagai lapisan pengesahan tetapi ia memerlukan kos. Pengesahan melalui percakapan adalah protokol keselamatan yang khas untuk VoIP. Malangnya, pengguna perlu melaksanakan protokol ini secara manual. Kajian ini mencadangkan protokol keselamatan untuk panggilan video dalam VoIP tanpa bergantung kepada PKI. Ia menggunakan metrik imej bagi melindungi kunci awam. Protokol ini tidak bergantung kepada sijil dan menggunakan kepandaian manusia untuk mengesahkan pemanggil. Pengguna tidak perlu bercakap dan membandingkan kod pengesahan secara manual lagi. Vektor Kuantisasi Terubah (MVQ) dan Fungsi Generasi Metrik Imej (IGF) turut diperkenalkan bagi meningkatkan kebolehpercayaan protokol ini terhadap isu-isu rangkaian. Keputusan eksperimen menunjukkan protokol ini adalah kukuh, boleh dipercayai dan praktikal untuk panggilan video dalam VoIP. Analisis keselamatan juga telah membuktikan bahawa protokol ini mampu menahan serangan ke atas VoIP.

ENHANCING SECURITY PROTOCOL FOR VOIP COMMUNICATION USING MODIFIED VECTOR QUANTIZATION

ABSTRACT

VoIP has introduced a new set of challenge that practically unheard off when a landline phone was used. Session Initiation Protocol (SIP) is often used as the main signaling protocol in VoIP. However, the lack of security feature has exposed SIP to a number of network threats. Public Key Infrastructure (PKI) is widely used to provide the authentication layer, but incurs maintenance cost. Verbal authentication is a security protocol specifically developed for VoIP. Unfortunately, user needs to manually perform the authentication steps over the phone. This study proposed a security protocol for video call in VoIP that does not relies on PKI. It uses image metric to secure the public key. The protocol is a certificate-less and uses human intelligence in authenticating the caller. User does not need to speak and compare the authentication code manually. Modified Vector Quantization (MVQ) and Image Metric Generation Function (IGF) are also introduced to help increase the reliability of the security protocol against network issues. The experimental results have demonstrated that the proposed protocol is robust, reliable and practical for video call in VoIP communication. The security analysis also has proved that the proposed protocol can resist known attacks against VoIP communication.

CHAPTER 1

INTRODUCTION

People have always been fascinated by the new technologies that allow them to be more connected and feel so close to somebody they care. It is even more satisfying if all of the technologies are just a click away, and Voice over Internet Protocol (VoIP) is one of them. Most new technologies come with their own set of inperfection and VoIP is no exception. Currently, session privacy and information confidentiality are major concerns among VoIP's user which need to be addressed.

This chapter presents the overview of VoIP infrastructure and the security concern surrounding its implementation. Section 1.2 describes the research problem studied in this research. Research motivation and the scope of this research are explained in Section 1.3 and Section 1.4 respectively. Section 1.5 explains the goal and the objectives of this study whereas Section 1.6 presents the research methodology used to achieve the objectives. The contribution of this research is described in Section 1.7, followed by the organization of this thesis in Section 1.8.

1.1 Voice over Internet Protocol (VoIP)

VoIP technology is recognized as low cost, highly scalable as well as flexibile. In contrast to the phone line, VoIP does not have any geographical restrictions since the

system uses a single user ID which can uniquely identify the user across the globe. In terms of personal usage, VoIP gives an extra mobility to the user who often on the move. Since audio and video data are digitized in the form of network packet, the aggregation of VoIP communication and existing IP network will help in reducing the overall operating cost.

A VoIP system consists of three major components namely Session Initiation Protocol (SIP) Server, SIP Proxy and User Agent (Zhang et al., 2010). Figure 1.1 ilustrates some of the major component in VoIP system. SIP Server manages user information, session tracking and database interaction. It usually contains SIP Registrar, a logical entity that handles participant registration and links the Uniform Resource Identifier (URI) address with a given IP address. As its name implies, SIP Proxy serves as the intermediate server that sit closer to the respective endpoint and forward the client request. A User Agent is the actual endpoint in VoIP communication that interacts directly with the user. Such endpoint can be in the form of softphone installed on the client's terminal or a physical phone that connected to VoIP gateway via PSTN or PBX network.

Despite the significant benefits offered, VoIP has introduced a new set of problems which are practically unheard off when a conventional phone line was used. The use of circuit-switched network in the phone line system has secured the communication signal until the physical layer. Replicating the same feat in VoIP is a formidable task since the whole system are normally deployed on an existing IP network. As a result, VoIP has inherited all vulnerabilities faced by such network such as denial of service, eavedropping and impersonation(Butcher et al., 2007).

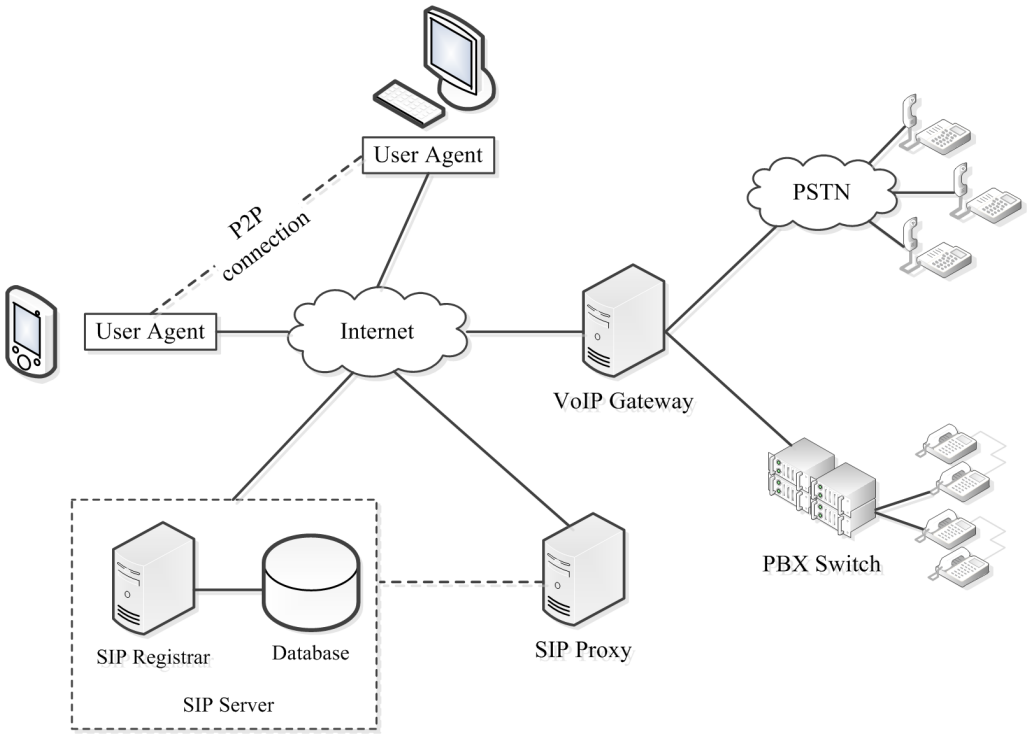

Figure 1.1: Major components in VoIP system (Dantu et al., 2009)

VoIP has two important aspects that need to be preserved, namely information confidentiality and session integrity (Butcher et al., 2007; Dantu et al., 2009). Information confidentiality means the prevention of information disclosure from the unauthorized individuals or systems whereas session integrity concerns on the protection of the identity of caller, receiver and the message. Preserving these two aspects would require a formalize security protocol to authenticate the shared key as well as to verify the user identity.

With the explosive growth of social network, communication privacy has become a major concern among VoIP user. It is more often than not where VoIP system is used not only for people to talk about general events in life, but also to convey secret information. Therefore, it is important for the service provider to give the security assurance to the user by integrating information confidentiality and session integrity aspects in their service implementation.

VoIP system is usually implemented on top of existing IP network. Such network is often under constant attacks by the adversary which causes VoIP to inherit its vulnerabilities (Bradbury, 2007). This makes VoIP system prone to active and passive attacks through the underlying network. A passive attack is one in which the intruder eavesdrops but does not modify the message stream in any way. An active attack is one in which the intruder may transmit messages, replay old messages, modify messages in transit, or delete selected messages from the session. A typical active attack is one in which an intruder impersonates one end of the conversation, or acts as a man-in-the-middle. In comparison, conventional phone system utilizes a circuit-switched network in routing the communication line. This guarantees a secure communication down to the physical layer. Tapping the physical communication line is the only way to breach its security.

Authentication, eavesdropping and impersonation are three issues that contribute to the security risk of VoIP (Keromytis, 2010). Without a reliable authentication, an adversary can masquerade as someone else or silently listen into the conversation. To avert this situation, public key exchange was employed to establish the shared key prior to the beginning of a session lifetime (Yang et al., 2005). The generated shared key will be used for encryption and decryption process on the subsequent VoIP's data. However, such method is not fully foolproof and some loopholes have been identified. Detailed on this matter and other existing approaches will be further explained in Chapter 2.

1.2 Research Problem

Eavesdropping and impersonation are two major issues in VoIP communication. However, as the primary protocol in VoIP, SIP does not provide a security mechanism to create a secure communication channel. Instead, SIP relies on other security protocols to achieve such task, particularly Secure RTP (SRTP). Nevertheless, the design flaw caused by false assumption between its components at different layers of VoIP stack has rendered SRTP unreliable. Public Key Infrastructure (PKI) is widely accepted in providing the security layer in VoIP communication. Since PKI requires maintenance cost particularly to subscribe and renew the certificate, replacing PKI with other viable method seems logical. Verbal authentication provides the alternative solution. The method uses human intelligence to authenticate the user. However, user needs to read and verify the authentication string manually over the phone which is time consuming. Therefore, there is a need to develop a formalized security protocol specifically for VoIP communication that can eliminate the manual verification process and at the same time does not rely on the PKI

To address these issues, this study proposed an alternative security protocol for video call in VoIP based on image metric derived from feature descriptor from the media stream. Image metric act as a biometric key that is unique in each session to secure the public key transmission. The nature of real-time media streaming in VoIP and the human intelligence is utilized in the decision making process. Human involvement is very minimal and they do not have to undergo the manual authentication process.

1.3 Research Motivation

VoIP session often involves video call. Due to its nature, every image frame in the video presents a very large entropy size due to the massive number of pixels. Therefore, these pixels can be used to generate a biometric key by using a feature extraction method. This ensures that the generated key is unique. Generating key from the video before and after the transmission give VoIP application the ability to perform key agreement and create the shared key without relying on PKI or human interaction. Hence, a large portion of time and processing overhead can be saved.

In addition, human brain has unparalleled neural processing and decision making capability. There are three aspects that human intelligence can play a role in VoIP security. Human can notice any difference in voice tone and able to combine visual and hearing perceptive ability to spot any abnormality between the physical action and the movement of lips. Doing the same process on the computer would require an intense image processing and complex neural network. Exploiting the human intelligence as a final decision maker instead of neural network can save memory and space consumption in the software implementation.

These two ideas are the motivation behind this study. There is a need for an alternative security protocol for VoIP without relying on the PKI or verbal authentication. The nature of VoIP stream and human intelligence can be utilized for developing the solution. Given the combination of image based key generation and human neural capacity, such solution is the most effective way in providing secure VoIP session among known user without degrading the security strength.

1.4 Research Scope

The security of SIP signaling outside of the session lifetime is highly dependent on the underlying IP network. Given the circumstances of the network and the number of existing security solutions, such domain is not covered in this study. This study emphasizes on protecting information confidentiality and the session integrity.

A single cryptography scheme may solve one issue but not the other. For example, key exchange scheme handle the key agreement without exposing the secret key but cannot verify the key sender. Whereas PKI can verify the key sender but incur extra cost. Therefore, a collective approach consists of multiple cryptography components is needed. Figure 1.2 illustrates some of the cryptography area that has been utilized to improve the security level in VoIP communication. The focus of this research are SIP signaling, public key exchange, encryption and verbal authentication.

The proposed work in this study is based on verbal authentication. Therefore, the assumptions made in verbal authentication would also apply. First, the sender (caller) and the receiver (callee) have known each other prior to the session under the premise that people only add a friend that they know on their VoIP's friend list. The familiarity should allow them to visually identify their identity. Second, the session has to take place in real-time.

The research scope in this study is limited to video call in the sense that the user can determine the identity much better if they can visually see the caller and therefore, impersonation is highly impossible. Here are the main talking points to summarize the research scope:

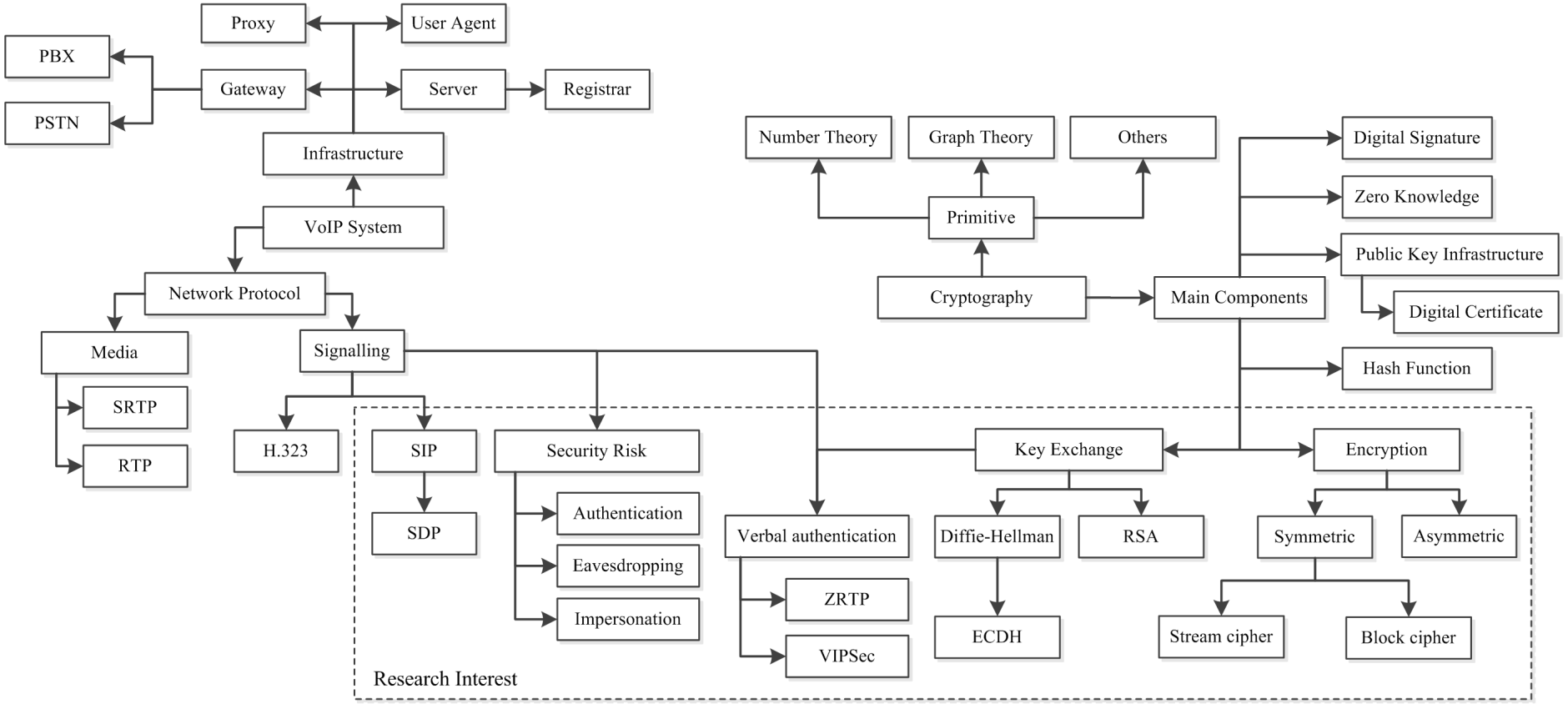

Figure 1.2: The application of cryptography to secure the VoIP communication

1. Emphasis on the session lifetime; a data streaming phase between SIP's ACK and BYE signal.
2. Aim to fulfill Confidentiality and Integrity aspects within the session lifetime.
3. Focus on SIP signaling, public key exchange, encryption and verbal authentication.
4. VoIP session involves only known peers and take place in real-time.
5. The proposed work is specifically developed for video call in VoIP.

1.5 Research Objective

The goal of this research is to develop an alternative security protocol for video call in VoIP communication based on image metric from the media stream. The proposed protocol takes the advantage of real-time video transmission in digesting the image metric independently and secures the public key during the key exchange. The objectives of this research are:

1. To design a security protocol for video call in VoIP without relying on the PKI.
2. To develop a formalized method in generating a unique biometric key from real-time video transmission in VoIP by using the feature extraction technique on multiple image frames.
3. To eliminate the need for user to manually read and compare the authentication string over the phone.

1.6 Research Methodology

A large portion of the research works are spent during the process of designing the solution. Figure 1.3 illustrates the steps involved in the research. The works are done based on four major components, namely key agreement, encryption, hash function and image metric generation. In each component, several options are considered based on the given requirement.

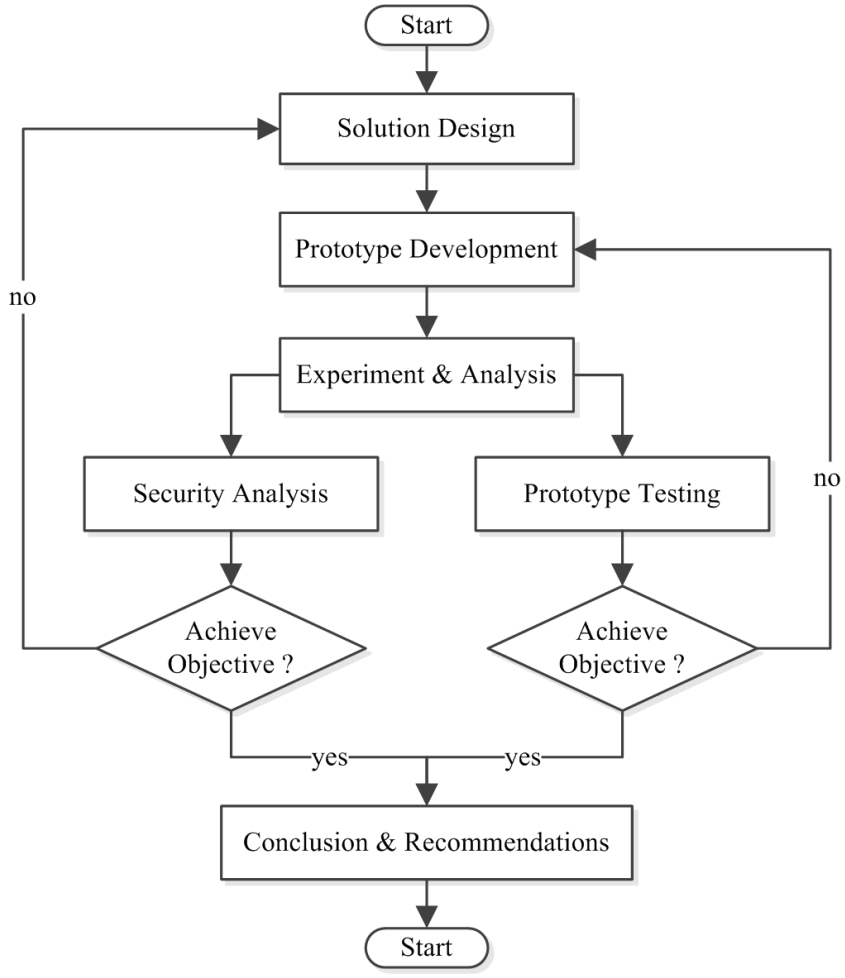

Figure 1.3: Steps involved in the research works

The proposed protocol utilized Elliptic Curve Diffie Hellman (ECDH) key exchange scheme to negotiate the key between the participants. ECDH is based on point multiplication on the elliptic curve graph, known as Elliptic Curve Discrete Logarithm

Problem (ECDLP). It produces a smaller key size but retain the same security strength as the DH. This reduce the time and processing overhead of overall solution. Advance Encryption Standard (AES) is used to encrypt the public key from ECDH scheme with the generated image metric. Sosemanuk algorithm is selected in encryption and decryption of real-time stream for secure channel since the algorithm is one of the fastest stream cipher available.

Then, an approach called Modified Vector Quantization (MVQ) and the corresponding Image metric Generation Function (IGF) are designed. In general, vector quantization involves dividing a still image into multiple subspaces. Each subspace constitutes a different feature descriptor which can produce a complete image metric when combined. A pre-shared key derived from the image metric is used as a symmetric key to encrypt the public key during its transmission.

The works on this study continues with the prototype development. This stage focuses on integrating all components and cryptography modules with the VoIP stack and the SIP server. A VoIP platform is needed in order to run the application in actual environment. At first, a complete open-sourced application called Jitsi is considered. After a thorough evaluation, Jitsi is discarded from the prototype development due to high complexity of the internal codes and a lot of unnecessary components which affect the performance of the proposed protocol. Hence, a new VoIP platform is developed specifically for this research. VoIP stack from JAINSip is selected for cross platform integration. A siphone is developed as a proof of concept. Siphone is a SIP's user agent that serves as the endpoint application and can be installed on desktop.

After prototype development is finished, the proposed security protocol is tested in the actual environment. Numbers of experiment are performed in order to measure the effect of external factors on the protocol such as color variation and packet drop. A thorough security analysis also been done against known attacks on VoIP communication. This is to ensure the reliability and applicability of the proposed security protocol. Finally, the works on this research are finished after all of the research objectives are achieved.

1.7 Research Contribution

The purpose of the protocol is to answer the question: How to exchange public key in open network without relying on digital certificate? Verbal authentication is simple but requires user to read and compare the authentication string manually over the phone. Hence, the contribution of this study is threefold: First, an alternative security protocol that does not rely on PKI. The shared secret is negotiated without relying on digital certificate. Second, a Modified Vector Quantization (MVQ), an improved approach that allow a consistent generation of key string from the image frame under certain packet lost. Image metric Generation Function (IGF), a new sequential step in generating the image metric is also proposed. Third, the proposed security protocol eliminate the need for user to manually read and compare the authentication string over the phone as in verbal authentication.

1.8 Thesis Organization

The work conducted in this thesis is presented in three chapters. This chapter provides a brief explanation of the relative concepts in VoIP communication, research domain, research motivation, problem statement, research objective and the contribution of this study.

Chapter 2 describes the security concern, a number of threats against VoIP system and their classification. Then, some of the general cryptographic algorithms used in VoIP application are presented, including public key exchange, public key cryptography (PKC) and encryption. The discussion of current security standard used in VoIP, verbal authentication and other non-standard security protocol specifically developed for VoIP is also covered in this chapter, followed by a review on image metric generation process.

Chapter 3 presents the proposed security protocol. A detailed description of protocol attributes, assumptions, complete signaling handshake and the key continuity feature is included. The steps needed to generate the image metric using IGF based on MVQ approach are also explained, followed by a concept visualization of MVQ and how it differs from the old approach.

Chapter 4 demonstrates the proof of concept of the research work. A simple VoIP application is developed to integrate with the proposed security protocol and tested in a controlled environment. Packet drop is used to signify the network issues such as high latency, low bandwidth, and packet lost. Then, the result of the experiments is discussed from the perspective of protocol's robustness, efficiency and computational

complexity. Next, the proposed security protocol is analyzed against a number of known threats against VoIP.

Finally, the work of this study is concluded in Chapter 5. The proposed security protocol, method used, security analysis and the contributions of the study are summarized and the future work is presented.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Understanding the underlying protocols involved in VoIP communication is very important in order to design and develop a reliable security protocol and integrate it with the existing infrastructure. This includes Session Initiation Protocol (SIP), Session Description Protocol (SDP) and Real-time Transport Protocol (RTP). Each protocol is developed for a specific task. For instance, SIP initiates, manages, and terminates a session while RTP handles a real-time media transmission peer-to-peer basis.

This chapter continues with a review on the threats currently faces by VoIP. A number of existing works on VoIP security are also discussed. It begins with the cryptography fundamentals such as public key exchange, Public Key Infrastructure (PKI), encryption and decryption. Issues around the use of Trusted Third Party (TTP) and the digital certificate are also highlighted. A review on image metric is also presented at the end of this chapter.

The current security standards employed in VoIP communication are also discussed in this chapter, particularly Secure RTP (SRTP) and its components, followed by other non-standard protocols that have been proposed for VoIP. The existing works done in the area of VoIP security are grouped in five groups, namely key exchange, encryption

and decryption, signaling and media transport, standard security protocol, verbal authentication and other non-standard security protocol. Figure 2.1 provides the overall representation of the literature review conducted in this study.

2.2 Signaling and Media Transport Protocol

It is important to understand the VoIP stack and some of the major protocols involved in VoIP communication before designing the security protocol for VoIP. VoIP stack is a collection of network protocol that involved in VoIP communication. The working domain of such protocols defines their position in the stack.

As illustrated in Figure 2.2, VoIP stack consist of five layers. In the transport layer, VoIP usually uses User Datagram Protocol (UDP) instead of Transmission Control Protocol (TCP) to maintain the Quality of Service (QoS). TCP is connection oriented and has to complete the three-way handshake. If using the TCP, VoIP has to wait for packet re-transmission in the event of packet lost, thus defeat the purpose of having the real-time communication.

In signaling layer, VoIP comes with two flavors namely H.323 protocol and Session Initiation Protocol (SIP). Both protocols are designed to handle the media session but has a specific approach to cater the different needs (Glasmann et al., 2003). H.323 is specifically developed to handle real-time audio and video data transmission using VoIP-compatible terminal. Due to the proprietary signaling and media formatting, H.323 is very good in an interfacing VoIP system with supplementary services like PSTN and PBX network.

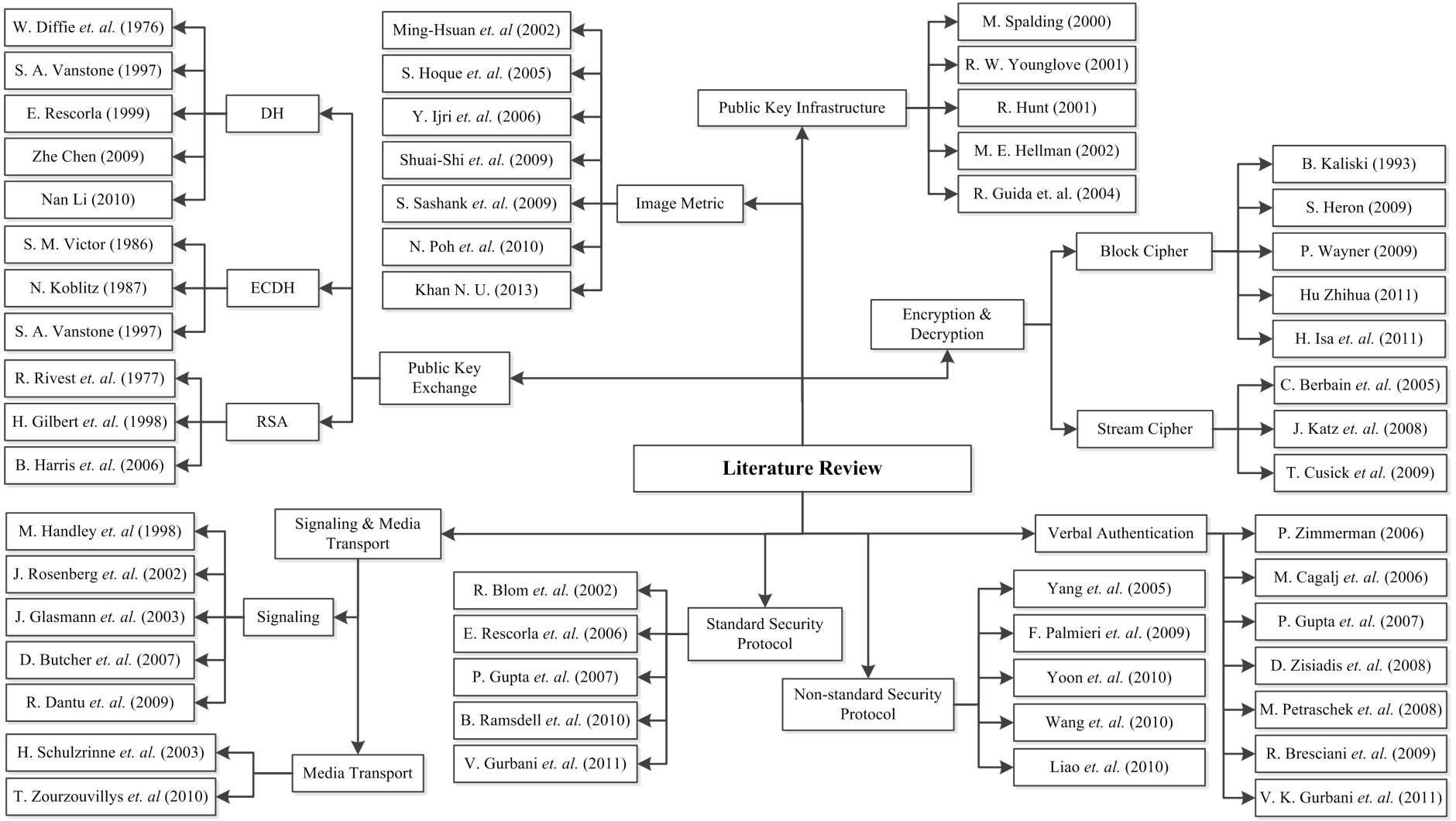

Figure 2.1: Overview of the literature review

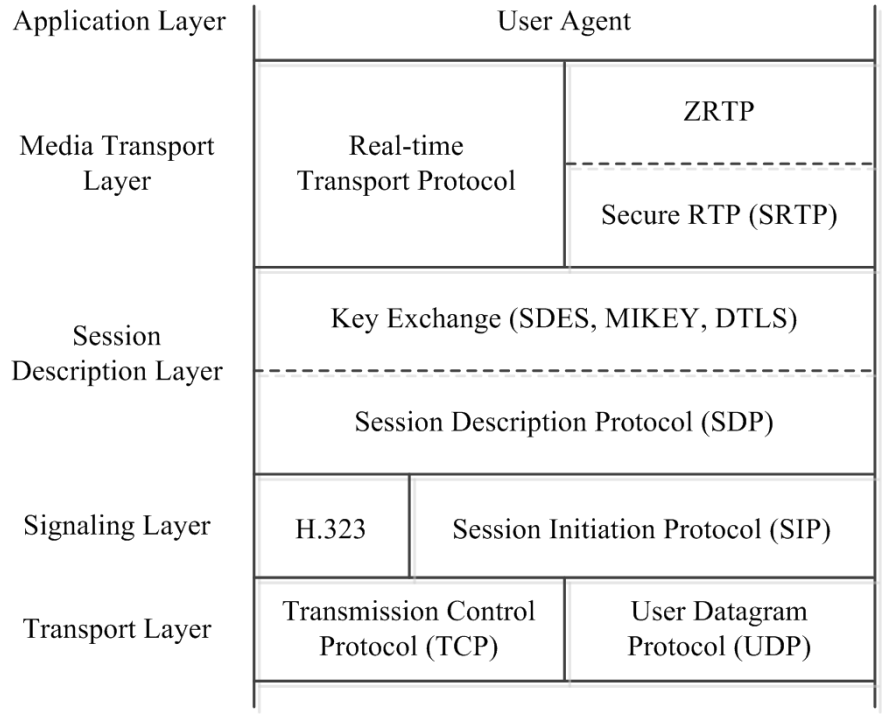

Figure 2.2: VoIP stack, adapted from (Gupta and Shmatikov, 2007)

In contrast, SIP is designed to be simpler, flexible and less complex than H.323. SIP signaling does not specify explicitly the terminal requirement. SIP gives more effective mechanism in interacting with non-VoIP compliant terminal. SIP provides a wider range of VoIP application, particularly in general session management that may not necessarily involve audio and video live streaming. As a result, SIP has been widely used as the main protocol in VoIP. Therefore, the work done in this thesis is focused on the security issue for SIP-based VoIP system.

SIP initiates, manages and terminates the session while Session Description Protocol (SDP) describe the format for media transfer. After the session is established, RTP takes over the session and begin transmitting the media stream. SIP and SDP are text-based protocol and did not provide any security layer.

2.2.1 Session Initiation Protocol (SIP)

Session Initiation Protocol (SIP) is a signaling protocol for multimedia communication which includes VoIP, multi-conference, and IP telephony. It was developed by Internet Engineering Task Force (IETF) in 1996 and designed to be independent from the underlying network. RFC 3261 describes a detailed specification of the signaling data and the required stack (Rosenberg et al., 2002). It has three primary functions which are service invitation, parameter synchronization and service termination. Instead of using numerical addresses to identify participant, SIP uses an email-like address that is easy to remember and unique across the globe.

SIP has two modes of communication, namely *Stateful* and *Stateless*. *Stateful* retain session ID and persisted until the end of the session, keeping all challenge and response handshake in a single dialog. In *Stateless* mode, every SIP signal is a new and independent signal, thus allowing a simple handshake. SIP is a text-based protocol and functions according to challenge and response mechanism. This keeps the VoIP implementation simple and very flexible. Below is the example of a SIP INVITE message sent from Alice to Bob:

```
To:  Bob <sip:bob@example.com>
From:  Alice <sip:alice@example.com>;
tag= 0gh4d
Via:  SIP/2.0/UDP a.example.com;
branch= z5kH3bKshEQ
CSeq:  76298 INVITE
Call-ID: 74622011@example.com
Content-type:  application/sdp
```

From chronology perspective, VoIP session can be divided into three phases namely call setup, data streaming, and session tear down. Figure 2.3 illustrates the SIP signal-

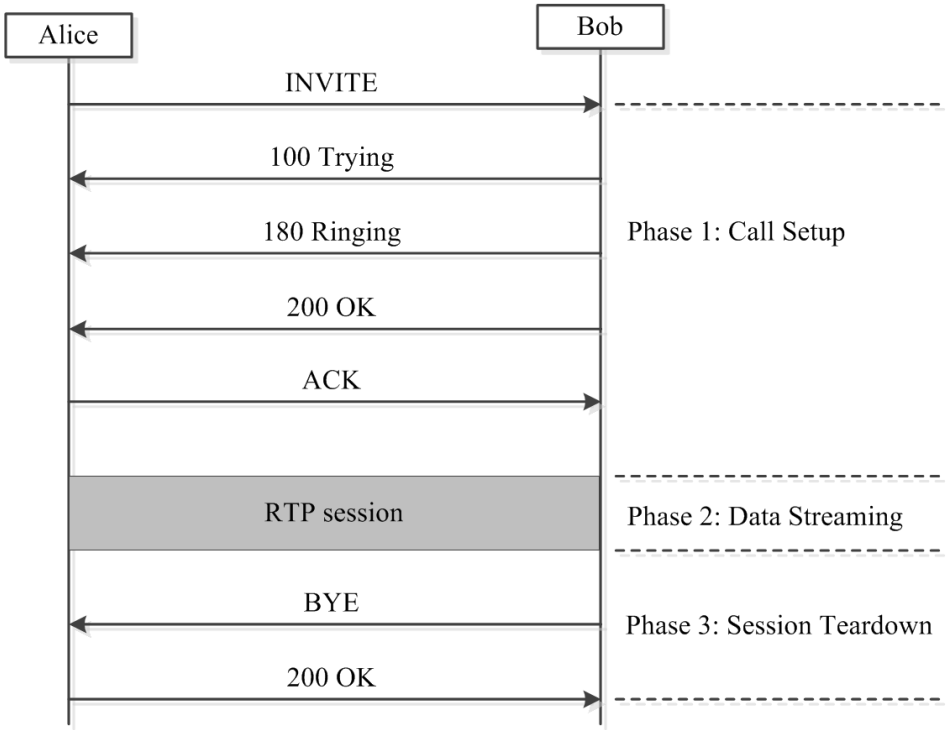

Figure 2.3: SIP protocol handshake between Alice and Bob (Wang and Liu, 2010)

ing between Alice and Bob before and after the session ended. During call setup, Alice initiates the calling process by sending INVITE signal and wait for a 200 OK signal from Bob. The INVITE message contains the SDP payload to synchronize the parameter and format needed for audio and video streaming between her and Bob.

VoIP session is established once Alice receives the response signal and replies with the ACK signal to Bob to complete the handshake. The data streaming phase begins once Alice and Bob became a VoIP participant. RTP takes over the session by initiating real-time audio and video data transmission process on the specified port. During session tear down phase, Bob sends BYE signal to Alice and she replies with 200 OK signal to end the VoIP session. At this stage, all UDP connection is terminated and VoIP stack is reset.

The fact that SIP is a text-based protocol has made it seriously exposed to various network threats (Dantu et al., 2009). For example, adversary can exploit the content of the SIP message for malicious purposes, particularly eavesdropping and impersonation. For instance, INVITE message contain call data such as ID of participant, session ID as well as the next proxy server. The adversary can intercept the message, replaces the originator's ID and makes the receiver response to his terminal.

Given the vulnerability of the underlying network, communicating secret information through VoIP is dangerous. User authentication is needed to mitigate such situation. Unfortunately, SIP does not provide any authentication layer to validate the originality and integrity of a message (Butcher et al., 2007). It has to rely on other protocol to create the secure channel and preserve the session privacy.

2.2.2 Session Description Protocol (SDP)

Session Description Protocol (SDP) is a protocol for describing the format of the streamed media for the purpose of session initiation and parameter negotiation. As illustrated in Figure 2.2, SDP works on session description layer of VoIP stack. The protocol is described in detail in RFC 2327 (Handley and Jacobson, 1998).

Both communicating parties may not have the same computer specification, operating system, input hardware and quality of network connection. For instance, Alice and Bob will have a different user agent with a different set of specification and network strength. Although Alice has a higher bandwidth and can support a full high definition video, Bob may need to settle with a lower resolution due to his poor Internet connection.

Due to these circumstances, Alice and Bob need to agree on a same set of format for audio and video that can satisfy both needs. In such situation, Alice and Bob will exchange SDP containing the expected session profile and begin the RTP session once both parties synchronized. SDP is included in the INVITE message sent by the caller. Similar to SIP, SDP does not provide any cryptography mechanism to make a secure communication. However, it can serve as a host for carrying a relevant key materials for the security handshake (Gupta and Shmatikov, 2007). Table 2.1 shows some of the SDP options and the example of the corresponding attributes.

Table 2.1: SDP options and its usage

Options	**Usage**	**Exmple**
v	The version of protocol	0
o	Source and session identifier	alice 5624825461 5624825461 IN IP4 a.example.com
s	The name of session	SDP Seminar
c	The information about the connection	IN IP4 192.0.2.101
t	Time of the session is active	0 0
m	Media description	audio 49172 RTP/AVP 0
a	More attributes	rtpmap:0 PCMU/8000

2.2.3 Real-time Transport Protocol (RTP)

Real-time Transport Protocol (RTP) is a standard protocol to handle multimedia data in real-time transmission either unicast or multicast. This protocol is specified in RFC 3550 (Schulzrinne et al., 2003). The protocol is commonly used in Internet telephony application such as VoIP.

Real time audio and video streaming require every packet to arrive at the intended destination in a timely manner. RTP is developed based on UDP and able to tolerate certain degree of packet lost. In order to keep things synchronized, RTP does not wait for the sender to resend the lost packet (Zourzouvillys and Rescorla, 2010). Instead, it will skip to the next received packet and re-order accordingly based on the sequence number. The error correction algorithm will try to make the packet lost unnoticeable to the user. If RTP is designed based on TCP, the protocol will have to send a re-transmit signal to the sender and wait for the packet to arrive. This will cause a lot of unnecessary delay and does not signify the purpose of having the real-time streaming.

RTP works in tandem with SIP to create a full duplex communication channel between the endpoints (Zourzouvillys and Rescorla, 2010). It uses the sister protocol, Real-time Transport Control Protocol (RTCP) to control and monitor the data transmission. Every endpoint will exchange RTCP packet periodically to monitor the media quality. This allows RTP to detect any packet loss and compensate the packet delay.

RTP and RTCP are independent from the underlying network and transport protocol. However, RTCP packet is exchanged separately from the RTP packet using two different ports. The adversary can exploit the RTCP packet if the protocol is not secured properly.

RTP is not designed for secure communication. However, the protocol is very flexible in the sense that the input and output stream can be modified before the transmission. Security in RTP is achieved by performing encryption and decryption using the cipher algorithm on the input and output stream. Input data is XORed with the

continuous random bits from Pseudo Random Number Generator (PRNG) to become cipher text and placed into the payload. Once the packet is received by the intended receiver, the payload will be converted back to the plain text using the same key.

2.3 Threats Against VoIP

Each threat faced by VoIP is categorized based on their effect on *Confidentiality*, *Integrity* and *Availability* (Butcher et al., 2007). Essentially, *Confidentiality* threats breach the session privacy and expose the content of the conversation to the adversary. *Integrity* threats jeopardize the accountability of the caller, the message and the recipient while *Availability* threats mean the inability of VoIP user to make and receive the call. *Availability* threats was not part of the research focus since the attacks are mainly caused by the underlying IP network which lead to the Denial of Service attack (DOS). The solution for this issue is very similar to the DOS attack on the network infrastructure (Butcher et al., 2007). As mentioned by (2008), the only way to secure VoIP is by encrypting its media content. Hence, this study focuses on the threats against *Confidentiality* and *Integrity* aspect of the VoIP session which is not protected eventhough the network infrastructure is well secured.

Eavesdrop and impersonation are two terms that often be associated with VoIP. Eavesdrop is when the adversary silently listens to the victim's conversation without their consent (Butcher et al., 2007). The adversary could retrieve the meaningful information and use it to gain control over user's credentials. Session privacy is loss as conversation is exposed plainly to the adversary. Survey has shown that the eavesdropping constitutes 20 percent of VoIP vulnerabilities (Keromytis, 2010).