SELF-VERIFICATION OF PUBLIC-KEY AGREEMENT OVER VOIP USING RANDOM FUSION SCHEME

ALFIN SYAFALNI

UNIVERSITI SAINS MALAYSIA

2016

SELF-VERIFICATION OF PUBLIC-KEY AGREEMENT OVER VOIP USING RANDOM FUSION SCHEME

by

ALFIN SYAFALNI

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

August 2016

ACKNOWLEDGEMENTS

FIRST of all, I would like to express my greatest gratitude to Allah S.W.T, the Most Beneficent and the Most Merciful. Because of the blessings and the opportunities HE has given to me, I am able to complete this thesis. Then, I am sincerely thankful my supervisor, Prof. Azman Samsudin, and co-supervisor, Dr. Mohd. Adib Hj. Omar, for all their supports, knowledges, and supervisions during my research study. Besides, I would like to thank both examiners of this thesis, Prof. Suhaidi Hassan and Assoc. Prof. Aman Jantan, for giving thoughtful and useful comments.

Importantly, I forward my deepest gratefulness to my family, especially to my father, **Syafalni**, my mother, **Elly Sasmita**, and my brother, **Infall**. I am very grateful for every second of their supports, encouragements, inspirations, and loves throughout my life, regardless the condition and situation.

I would like to thank Universiti Sains Malaysia for providing any financial assistance during my research, particularly Short Term Research Grant (No. 304/PKOMP/6312091) and Graduate Assistant Scheme. Along with that, I send the best credit to the Dean of School Computer Sciences, Prof. Ahamad Tajudin Khader, my former supervisor, Prof. Rahmat Budiarto, and also my undergraduate advisor, Mr. Mohd. Azam Osman.

I thank my fellows in Information Security Lab., specially Mohd. Yazid, for the help and cooperation. And, I am grateful to all of my friends and family whom I met in Penang, Apak Syafrudin, Ibu Aci, Mega, Sukma, Hadri, Kurnia, Idris Xian, Maulana, Najihah, Ramizah, Ezzeddin, Syahmi, Ariefandi, Razif, Erpi, Zulhusni, Mas Ali, Kang Tedi, Bang Didi, Kak Ayu, Bang Hilal, Pak Supriyanto, Dr. Yose, Ust. Yasir, Pak Herpandi, Pak Arnawan, Flying Titans Volleyball buddies and others whose names could not be mentioned here one by one, which have made my life more colorful. At last, I appreciate anyone who has contributed either directly or indirectly in this matter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

ACELP Algebraic code-excited linear prediction

ADC Analog-to-digital converter

ADPCM Adaptive differential pulse-code modulation

AES Advanced Encryption Standard

AMR-NB Adaptive Multi-Rate Narrowband

AMR-WB Adaptive Multi-Rate Wideband

API Application programming interface

CA Certificate authority

CBC Cipher Block Chaining

CELP Code-excitation linear prediction

CFB Cipher Feedback

CID Caller identification

CS-ACELP Conjugate structure ACELP

CTR Counter

EC-DKCS Elliptic Curve – Dynamic Key Change Scheme

EC-KGF Elliptic Curve – Key Generation Function

ECB Electronic Codebook

ECC Elliptic Curve Cryptography

ECDH Elliptic Curve Diffie-Hellman

ECDSA Elliptic Curve Digital Signature Algorithm

DAC Digital-to-analog converter

DDoS Distributed denial-of-service

DH Diffie-Hellman

DH-SC DH Key Agreement Based on Short String Comparison

DoS Denial-of-service

DSA Digital Signature Algorithm

DTLS Datagram Transport Layer Security

GSM Global System for Mobile Communications

GSM-AMR GSM – Adaptive Multi-Rate

GSM-EFR GSM – Enhanced Full Rate

GUI Graphical user interface

HCI Human – computer interaction

HTTP Hypertext Transfer Protocol

HTTPS Hypertext Transfer Protocol Secure

IDE Integrated development environment

IDS Intrusion detection system

IEEE Institute of Electrical and Electronics Engineers

IETF Internet Engineering Task Force

IKE Internet Key Exchange

iLBC Internet Low Bitrate Codec

IMS IP Multimedia Subsystem

IP Internet Protocol

IPS Intrusion prevention system

IPSec Internet Protocol Security

IPTV IP television

iSAC Internet Speech Audio Codec

ITU International Telecommunication Union

ITU-T ITU Telecommunication Standardization Sector

JMF Java Media Framework

LAN Local area network

LCP Longest common prefixes

LD-CELP Low-delay – code excited linear prediction

LSB Least significant bits

SHA Secure Hash Algorithm

MGCP Media Gateway Control Protocol

MIKEY Multimedia Internet Keying

MIME Multipurpose Internet Mail Extensions

MITM Man-in-the-middle

MLT Modulated Lapped Transform

MOS Mean opinion score

MP-MLQ Multi-pulse Maximum Likelihood Quantization

OFB Output Feedback

OS Operating system

P2P Peer-to-peer

PBX Private branch exchange

PCBC Propagating Cipher Block Chaining

PCM Pulse-code modulation

PFS Perfect forward secrecy

PGP Pretty Good Privacy

PKC Public-Key Cryptography

PKG Private Key Generator

PKI Public-Key Infrastructure

PRNG Pseudorandom number generator

PSTN Public switched telephone network

QoS Quality of service

QR Quick Response

RA Registration authority

RAM Random-access memory

RC4 Rivest Cipher 4

RFS Random Fusion Scheme

RSA Rivest-Shamir-Adleman

RTCP RTP Control Protocol

RTP Real-time Transport Protocol

S/MIME Secure / Multipurpose Internet Mail Extensions

SAS Short Authentication String

SB-ADPCM Sub-band – ADPCM

SCTP Stream Control Transmission Protocol

SDP Session Description Protocol

SDES SDP Security Descriptions

SID Session Identifier

SIP Session Initiation Protocol

SPIT Spam over Internet telephony

SPOF Single point of failure

SRTP Secure Real-time Transport Protocol

SS7 Signalling System No. 7

TCP Transmission Control Protocol

TCP/IP Transmission Control Protocol / Internet Protocol

TLS Transport Layer Security

TSNFC Two-stage noise feedback coding

TTP Trusted third party

UA User agent

UAC UA client

UAS UA server

UDP User Datagram Protocol

VIPSec Voice Interactive Personalized Security

Vishing Voice phishing

VoIP Voice over Internet Protocol

VoIPSA VoIP Security Alliance

WoT Web of Trust

XML Extensible Markup Language

XMPP Extensible Messaging and Presence Protocol

ZRTP Zimmermann Real-time Transport Protocol

PENGESAHAN KENDIRI PERJANJIAN KUNCI-AWAM MELALUI VOIP MENGGUNAKAN SKIM RAWAK FUSION

ABSTRAK

Telefoni Internet, yang dikenali juga sebagai Suara melalui Protokol Internet (VoIP), menjadi salah satu alternatif telekomunikasi yang popular disebabkan penggunaan Internet yang semakin meluas. Internet memperkaya cara sistem telefoni digunakan, tetapi dalam masa yang sama menimbulkan pelbagai kebimbangan, terutamanya keselamatan. Tidak seperti telefon konvensional, pencuri-dengar komunikasi Internet boleh dilaksanakan secara maya tanpa memerlukan sebarang akses fizikal. Isu ini memberi peluang yang lebih besar kepada musuh untuk mengeksploitasi privasi komunikasi Internet. Kerana itu, penyulitan telah digunakan untuk melawan tindakan yang merugikan ini. Selain itu, perundingan kunci, yang merupakan asas dalam penyulitan, mesti diselamatkan untuk menghindarkan ancaman yang dikenali sebagai serangan orang-di-tengah (MITM). Namun begitu, perundingan kunci yang selamat seperti Infrastruktur Kunci-Awam (PKI) tipikalnya melibatkan pihak ketiga yang dipercayai (TTP) untuk mengesahkan kunci awam, yang menuntut kos atas perkhidmatan yang diberikannya. Tesis ini membentangkan satu mekanisme pengesahan alternatif untuk kunci awam melalui komunikasi VoIP. Alternatif ini dirancang untuk mewujudkan sebuah perjanjian kunci yang boleh dipercayai tanpa kehadiran TTP pada sesi panggilan antara dua peserta yang telah mengenali antara satu sama lain terlebih dahulu. Satu skim pengesahan baru diperkenalkan sebagai Skim Rawak Fusion (RFS) yang mengambil kelebihan dari komunikasi telefon dimana interaksi masa nyata dan kecerdasan manusia dapat dioptimumkan secara fleksible semasa sesi telefon. RFS memasukkan ‘cap jari’ (nilai cincangan) kunci awam di dalam aliran suara peserta. Satu teknik untuk mengekstrak ‘cap

jari' ditakrifkan dalam RFS yang dicadangkan menggunakan algoritma-algoritma pencarian corak dan rentetan sepadan. Sebuah kerangka kerja hibrid kemudian dicadangkan yang menggunakan RFS pada perjanjian kunci Lengkungan Eliptik Diffie-Hellman (ECDH), yang kemudian dirujuk sebagai kerangka kerja ECDH-RFS. Kerangka kerja yang dicadangkan mengesahkan kunci awam secara automatik sebagai satu kunci yang sah jika suara peserta membawa 'cap jari' yang setanding dengan kunci awam yang diterbitkan. Akhirnya, kerangka kerja ini memberikan kesukaran yang besar kepada penyerang untuk mengganggu penukaran kunci awam. Oleh kerana itu, kerangka kerja menyediakan integriti kunci awam yang mantap. Akibatnya, sebarang percubaan memalsukan kunci awam sama ada akan menyebabkan penolakan terhadap pertukaran kunci atau merosakkan komunikasi itu sendiri yang mana menggerakkan penggera kepada peserta. Keputusan uji kaji menunjukkan kaedah yang dicadangkan telah melakukan pengesahan yang munasabah dalam masa 5 hingga 60 saat perbualan dengan overhed yang marginal. Selain itu, analisis keselamatan telah membuktikan bahawa kerangka kerja yang dicadangkan dapat mengesan dan mengelakkan percubaan-percubaan dalam serangan MITM. Penyelidikan ini menghapuskan peranan orang ketiga dalam keselamatan VoIP, dengan itu membantu mengurangi kos dan kerumitan dalam pengurusan sistem VoIP. Tambahan pula, cadangan kerja ini hanya memerlukan fungsi asas komunikasi telefon yang menjadikannya mungkin dilaksanakan dalam pelbagai keadaan, dengan atau tanpa adanya sokongan visual.

SELF-VERIFICATION OF PUBLIC-KEY AGREEMENT OVER VOIP USING RANDOM FUSION SCHEME

ABSTRACT

Internet telephony, also known as Voice over Internet Protocol (VoIP), has become one of popular alternatives in telecommunication due to the widespread of the Internet usage. The Internet enriches the way of telephony system is used, but in the meantime it elevates many concerns, particularly security. Unlike the conventional telephone, tapping the Internet communication is feasibly done virtually without requiring any physical access. This issue gives a greater opportunity for the adversaries to exploit the communication privacy. Hence, encryption has been utilised to combat such adverse acts. Besides, the key negotiation, which is the cornerstone for encryptions, must be secured to avert a threat known as man-in-the-middle (MITM) attack. However, a secure key negotiation like Public-Key Infrastructure (PKI) typically entails trusted third party (TTP) for public key verification, which demands costs in its service. This thesis presents an alternative verification for public key over VoIP communication. The alternative is designed to establish a trustworthy key agreement without the presence of TTP on the call session between two participants who have known each other in advance. A new verification scheme is introduced as Random Fusion Scheme (RFS) that takes advantages of telephone communications where real-time interaction and human intelligence can flexibly be optimised during the session. RFS inserts the public key's fingerprint (hash value) within the participants' voice stream. A technique to extract the fingerprint is defined in RFS using pattern searching and string matching algorithms. A hybrid framework is proposed that employs RFS on Elliptic Curve Diffie-Hellman (ECDH) key agreement, which is then referred as ECDH-RFS framework.

The proposed framework automatically verifies a public key as authentic if the participant's voice carries a comparable fingerprint as the published public key. Eventually, this framework gives a great difficulty for an attacker to interfere with the exchange of the public key. Therefore, the framework provides a robust public key integrity. Consequently, any attempt on forging the public key will either result in rejecting the key exchange or damaging the communication itself which raises an alarm to the participants. The experimental results show the proposed framework has performed a reasonable verification within 5 to 60 seconds of conversation with marginal overhead. Moreover, the security analysis has proved that the proposed framework could detect and avert attempts in MITM attack. This research excludes third party role in VoIP security, thus helps reducing cost and complexity in managing VoIP systems. Furthermore, the proposed work only requires the basic functionality of telephone communication that makes the application is feasible under a wide range of circumstances, either with or without visual support.

CHAPTER 1

INTRODUCTION

Alternative telephone services over the Internet, known as Internet telephony, has started gaining popularity over the conventional telephone services, e.g. Skype and WhatsApp. The increasing demand of this Internet service is triggered by the remarkable benefits offered such as low-cost, portability, and enriched functionality (Karapantazis and Pavlidou, 2009). The development of this telephony service is highly promising and foreseen to overtake the majority of telephony systems in the course of time (Conti, 2005).

Internet telephony, principally known as Voice over Internet Protocol (VoIP), resembles telephone functionality in many ways, especially digital telephones. The fundamental networking technology behinds VoIP system is based on the Internet Protocol (IP) network, which is a packet switching system. As the Internet is reaching an ubiquity, convergence between the two switching systems is highly anticipated as illustrated in Figure 1.1. The interoperability can be made by utilising a networking tool that is capable for adapting telephone signals and channelling the communication data across diverse networking systems, VoIP gateway for instance. The gateway connects the Internet backbone with the public switched telephone network (PSTN) and private branch exchange (PBX).

The development of this interoperability ecosystem has attracted many attentions up to the enterprise levels. However, VoIP adaptation has been encoutering quality of service (QoS) issues, because of the nature of the underlying network. Data sent in packets through the IP network are subject to experience transmission problems such as loss and delay that occur due to many factors, including software, hardware, and network. As a

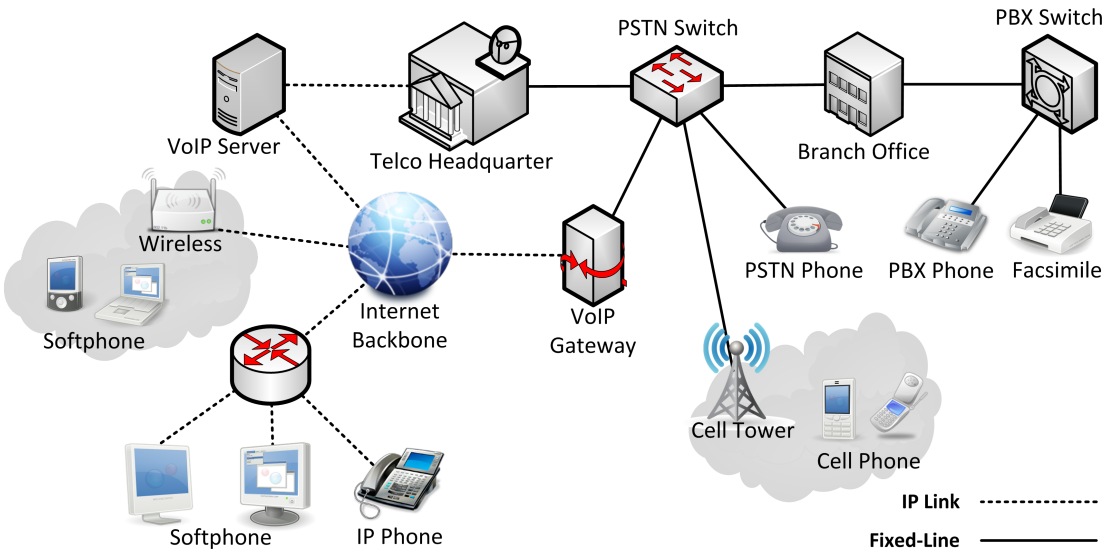

Figure 1.1: Interoperability telephone ecosystems

result, perceivable disruptions easily transpire during VoIP calls when these problems are excessive. These disruptions have become the major obstacles in VoIP adaptation and may affect the overall reliability of the ecosystem.

In recent times, the QoS of VoIP systems has shown a positive progress due to the advancement of software, hardware, and network technologies. In contrary, the security issues have become a greater challenge along the rapid growth of Internet services and its users. Many works including Butcher et al. (2007), Karapantazis and Pavlidou (2009), and Keromytis (2011) have reviewed that relying on IP networks, mainly the Internet, makes VoIP system more vulnerable to security threats.

1.1 Security Threats

Potential threats in VoIP systems are classifiable based on VoIP threat taxonomy proposed by VoIP Security Alliance (VoIPSA) (Endler et al., 2005). The classification is concentrated on the trend of research publications in VoIP security. The classification divides VoIP threats into four major types: social threat, service abuse, denial-of-service (DoS), and traffic attack. Social threat covers misrepresentation acts and unsolicited calls. This type

of threat misleads users through scam, spam, and phishing. In VoIP systems, the attempts are correspondingly known as impersonation, spam over Internet telephony (SPIT), and voice phishing (Vishing). Secondly, service abuse attempts to use VoIP service in improper manners such as committing frauds and trespassing services. Thirdly, DoS is one of the most common threats in the Internet aiming to fail or interrupt the services. Lastly, traffic attack is vulnerable to most services that communicate through public networks. The attempts of the attack include eavesdropping, interception, and modification on the communication traffic.

In telephony, traffic attack has been noted as one of prominent risks that conceivable by third parties. Traffic attack is accomplished through wiretapping as schemed on Figure 1.2. In circuit-switched networks, it is difficult to wiretap the communication without a lawful access to its physical line or device (Sicker and Lookabaugh, 2004). Thus, it minimises possibilities of the attack from the adversaries. However, risks of traffic attack on IP networks are multiplied since security penetration tools readily available on the Internet. The adversaries can perform remote observation (passive attack) and manipulation (active attack) on the IP packets by modifying the path or implanting bug (malware) on vital points in the communication channel.

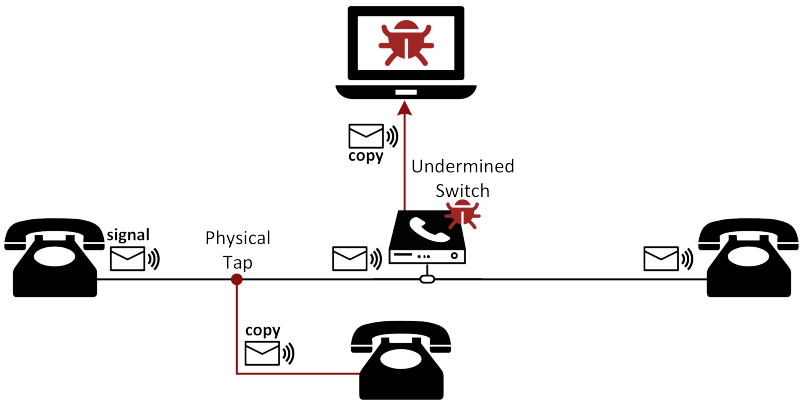

Figure 1.2: Typical wiretapping schemes

These exploitation acts are jeopardizing the primary aspects of information security, particularly confidentiality and integrity (Butcher et al., 2007). Although eradicating the acts is not likely, adding encryption helps prevent the adversaries to learn or alter the original content in the communication. Therefore, encryption has been noted as the the major solution to preserve confidentiality and integrity in VoIP communication as suggested by Dantu et al. (2009) and Son et al. (2012).

Encryption is the elementary method in cryptography to provide confidentiality or privacy of information that is exchanged through a public network. Encryptions require a single key, at least, to accomplish. Without the key, the users are neither able to encrypt nor decrypt information. Thus, the users need to share or agree on the key. The easiest method is by sending the key through the same channel beforehand. Nevertheless, this method paves an opportunity for adversaries to duplicate the key, which causes the ciphertext being compromised. Therefore, confidentially of the key is also very crucial prior to the creation of a secure transaction using an encryption.

In the regard of key confidentiality, among the notable cryptographic protocols for key agreement, such as Diffie-Hellman (DH) by Diffie and Hellman (1976), and Public-Key Cryptography (PKC), Rivest-Shamir-Adleman (RSA) by Rivest et al. (1978), are applied to resolve the matter using the concept of asymmetric cryptography. These protocols have been widely used as the foundation on many security systems, including in VoIP security such as Zimmermann et al. (2011), Zisiadis et al. (2008), and Wang and Liu (2010). However, the protocols are neglecting the public key's owner that makes the authenticity issue remaining unresolved (Trappe and Washington, 2005). Furthermore, this issue creates a possibility for adversary to hijack in the middle of secure channel unknowingly which is further identified as man-in-the-middle (MITM) attack.

MITM attack has been the greatest hindrance to accomplish a secure channel. MITM attack can impact the secure channel losing over its identity control and encryption (Zisiadis et al., 2008). MITM attack involves both passive and active traffic attacks by making the legitimate users believe that their communication is direct without any interference. In asymmetric-key cryptosystems, the adversary intercepts the public keys from the legitimate users and then presents the adversary's public key on behalf to succeed the attack. In order to prevent this attempt, the public key has to be verified and accountable.

Public-Key Infrastructure (PKI) has been widely trusted to perform public key verification on many electronic commerce services such as *e*-shopping and *e*-banking. PKI distributes digital certificates that firmly bind the public key with the identity of its owner. PKI uses digital signature to verify the authenticity of public key, thus provides accountability and integrity aspects to the negotiation. Nevertheless, unless a trusted third party (TTP) is present in PKI, a digital certificate is difficult to confirm that could lead to the occurrence of MITM attack (Aghila and Chandirasekaran, 2011).

1.2 Problem Statement

The role of TTP in PKI is very important to achieve a reliable verification. However, TTP comes with certain prices such as service and certificate management (Ellison and Schneier, 2000; Gutmann, 2002). As an alternative, approaches such as Zimmermann Real-time Transport Protocol (ZRTP) by Zimmermann et al. (2011) and Voice Interactive Personalized Security (VIPSec) by Zisiadis et al. (2008) have introduced innovative mechanisms to verify the key agreement which is dedicated for VoIP. The verification is conducted in a verbal manner by comparing the verification code throughout the call session. This verification has been shown to be able in verifying the public key manually without involving any TTP through VoIP calls. Nevertheless, these approaches consumes indecisive time in confirming

the code mutually. Furthermore, ignoring this phase gives an opportunity for the adversary to successfully execute MITM attack (Petraschek et al., 2008).

This thesis proposes an alternative public key verification approach amid the aforementioned problem of verbal verification. The proposed approach is designed to achieve a self-verification on public-key agreement throughout VoIP call sessions without requiring any TTP and manual intervention from the users. The approach utilises well-known cryptographic tools that include key agreement, hash function, and pseudorandom generation. Furthermore, the approach makes use of the common properties in telephone communications, namely the real-time interaction and human intelligence.

A new verification scheme is introduced as Random Fusion Scheme (RFS). RFS is defined as a process to fuse a string on voice stream. The string serves as the verification code in the form of public key's fingerprint, the product of a cryptographic hash function. The main goal is identical to verbal verification, which is to prevent adversaries from replacing or falsifying public key as practiced by MITM attack, hence the integrity and accountability of the public key are preserved. Afterwards, a hybrid framework is synthesised by cooperating RFS with a light-weight variant of DH key agreement, Elliptic Curve Diffie-Hellman (ECDH), which is then referred as ECDH-RFS. The final stage of the hybrid framework produces a shared session key securely from the self-verified public key, which can be used for establishing a secure session.

1.3 Objectives

The main objectives of this research are specified as follows:

1. To propose a verification scheme that exchanges verification code through voice stream.

2. To develop a hybrid framework that verifies public key without assuming PKI and TTP by implementing the proposed verification scheme on DH key agreement.

3. To evaluate the complexity and reliability of the proposed verification scheme and analyse the overall cost and security of the hybrid framework, especially against MITM attack.

1.4 Scope of the Study

The scope of this study, as presented in Figure 1.3, is within the field of applied cryptography on VoIP system. The main goal of the research is to provide a reliable public key verification that can protect the integrity of a public key. The integrity of public keys is one of the most important security aspects to avert a harmful threat like MITM attack that tries to intercept and falsify the public key during its negotiation. Hence, encryptions can be securely established by employing the verified public key.

The basic concepts in cryptography and VoIP system are studied to build a firm research foundation. This study also reviews the common security threats that endanger VoIP communication and the existing preventive mechanisms. Besides, the study covers several verification approaches typically employed to confirm authenticity of the public key including PKI, Pretty Good Privacy (PGP), identity-based cryptography, and other related methods, particularly practicable for VoIP security. Moreover, the study concentrates on verbal verification, which is dedicated for VoIP communications. The study also includes the feasibility of steganography, specifically using voice, to support verification as intended in the proposed verification scheme.

The proposed research work is motivated by the capability of verbal verification to exclude the involvement of TTP on the public key verification by using the native character-

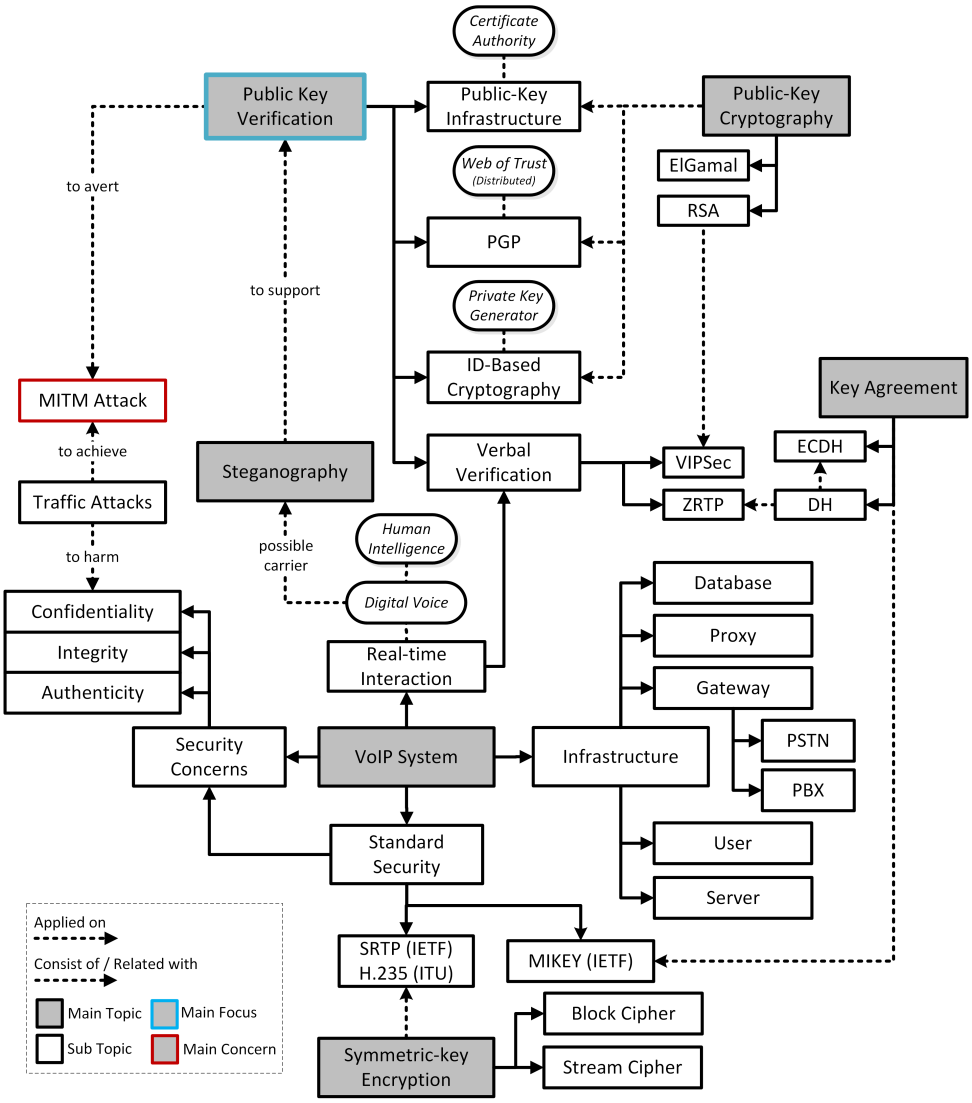

Figure 1.3: Summary of the scope in the research

istics in telephone communications, including real-time interaction and human intelligence. The verbal verification method is only effective to solve issues within the scope of the regular telephone communications where both participants, a caller (who initiates the call) and a callee (who receives the call), are able to recognise and confirm each other identity through voice. Meanwhile, ensuring the truthfulness of the participants is not technically addressed in the scope of verbal verification. Therefore, irregularity cases such as fraud and scam that occur during the communications are considered as beyond the scope of problem solving.

1.5 Contributions of the Research

The research contributes to the area of public key verification for VoIP security. The contributions are broadly divided into two modules which summarised as follows:

1. Random Fusion Scheme (RFS)
    (a) The proposed verification scheme abbreviated as RFS is based on keyless steganography that transports verification code over real-time voice.
    (b) Two core techniques are defined, namely fusing and defusing.
        i. Fusing is the technique to conceal a string on voice stream randomly.
        ii. Defusing is the technique to discover the concealed string in voice stream utilising the chosen pattern searching and string matching algorithms.
2. Elliptic Curve Diffie-Hellman (ECDH) and RFS hybridisation
    (a) A hybrid framework, ECDH-RFS, is developed for public key verification.
    (b) The framework includes some exceptional features:
        i. Self-verification, the main verbal verification feature that removes independency of the public key verification from relying the service of TTP.
        ii. Automatic integrity check, the framework eliminates any manual intervention from the users during the verification.
        iii. Low-cost, the framework only entails voice communication and the minimum computing specification for standard cryptographic systems such as pseudorandom, hashing, and key agreement.
        iv. MITM sensibility, MITM attack is sensed and immobilised by binding the public key's fingerprint with the participant's voice, thus the public key is improbable to be forged.

1.6 Research Methodology

The flow chart diagram in Figure 1.4 shows the summary activities carried out during the research study. The research consists of four chief phases: preliminary phase, modelling phase, prototyping phase, and experimental phase. Preliminary phase implicates problem definition and literature review in the research study. The literature review is crucial to support clarifying the research problem and achieving the research objectives. Upon the completion of the preliminary phase, the criteria had been defined to envision in preparing the design of the proposed scheme.

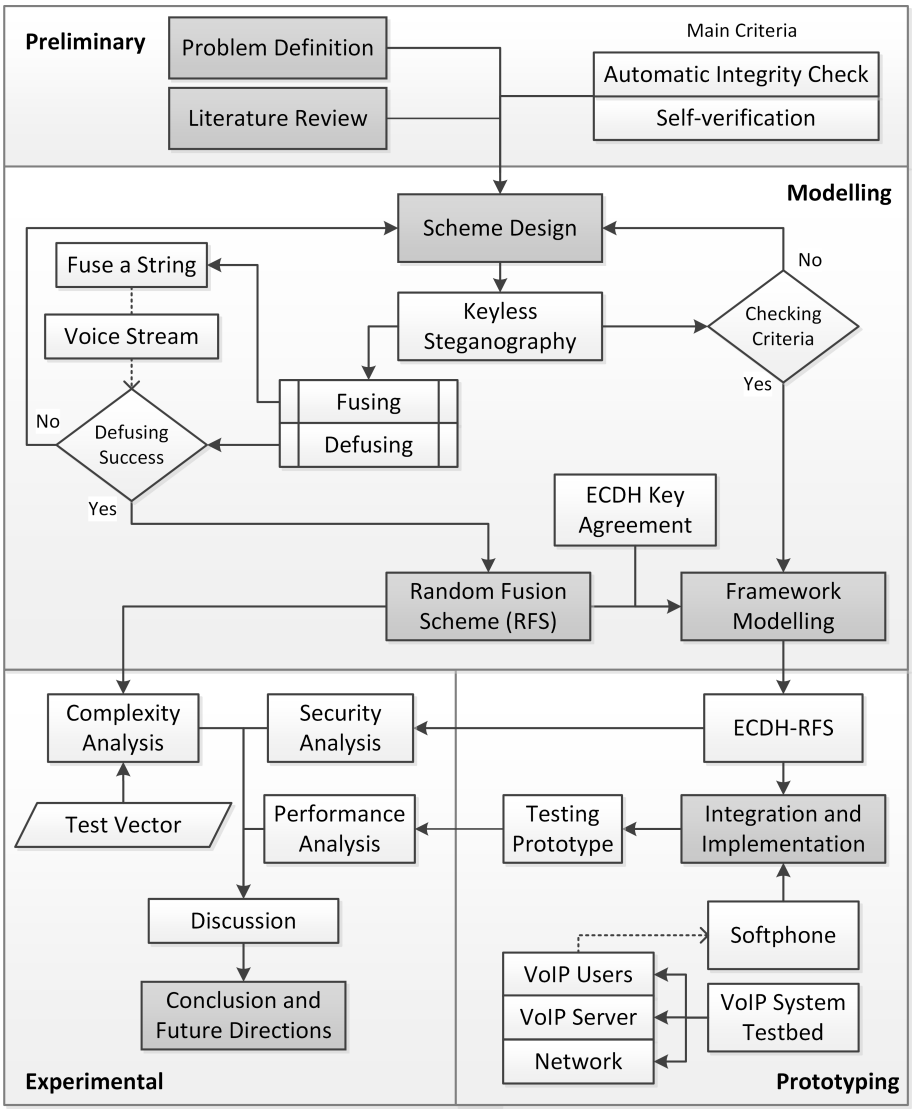

Figure 1.4: Research activities flow

Modelling phase defines the proposed verification scheme, RFS, in detail. RFS is designed based on voice steganography. In typical steganographic methods, there are two main techniques necessary for concealing the secret and revealing the secret back, which correspondingly referred as fusing and defusing in RFS. These techniques are realised and subsequently examined. The first objective is achieved in this phase if the desired outcome has been reached. Afterwards, the hybrid framework, ECDH-RFS, is modelled and prototyped to verify if the envisioned criteria are achievable.

In prototyping phase, a VoIP testbed is prepared for the implementation of ECDH-RFS. Table 1.1 lists the hardware requirements for the setup. The testbed is configured in a simple star topology network to attain a steady circumstance as illustrated in Figure 1.5. This setup is configured to allow a full control over parameters and variables, principally for the experimental purposes. Besides, a minimised setup gets rid of any foreign factors that can interfere with the experiments, thus dependable results can be expected.

Table 1.1: Hardware requirements

Hardware (Quantity)	**Function**
Computers (4)	VoIP server / VoIP user / Adversary
Recording & Playback devices (2)	Capture and play user's voice on user's machine
Ethernet Cable	Connect the devices on wired environment
Switch (1)	Link the computer devices

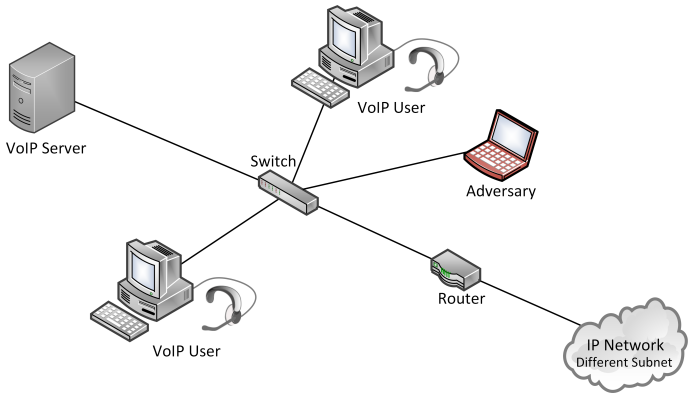

Figure 1.5: The network configuration

On software perspective, VoIP users communicate through the client application, also known as software phone (softphone). In the testbed, Session Initiation Protocol (SIP)-based softphone, called Jitsi, is adapted and installed on the users' machine. The main reasons for choosing Jitsi because Jitsi is an open source softphone with built-in security features including Secure Real-time Transport Protocol (SRTP) and ZRTP. Besides, Jitsi is running on Java platform, thus it is portable on various OS. The design specifications for the testbed are listed in Table 1.2. The main task in this phase is to develop the hybrid framework and achieve the second ojective. The framework is integrated with the configured VoIP client to serve as the testing protoype.

Table 1.2: Design specifications

Software Name	**Usage**
Linux	OS on VoIP server
Windows	OS on VoIP users
Jitsi	VoIP client
OpenSIPS	Open source SIP server
Java	Software development platform
Java Media Framework (JMF)	Library for enabling media applications on Java platform
Bouncy Castle	Java cryptography API
Eclipse	IDE for Java software development

Several experiments and analyses are conducted to evaluate the proposed verification scheme and the hybrid framework as well as to achieve the last objective. The proposed verification scheme, RFS, is analysed in terms of run-time and space complexities on its techniques. In the experiment, fixed test vectors are used and certain variables are varied in order to prove hypotheses by observing its effect towards the results. This experimental analysis confirms the applicability level of the proposed verification scheme on the deployed machine.

Moreover, security analysis is executed to verify that the hybrid framework, ECDH-RFS, is provably secure, especially against MITM attack. The analysis clarifies the defen-

sive method of the framework in dealing with the threat. Furthermore, the performance of the testing prototype is analysed and discussed. This experiment attests the reliability of the framework on an actual scenario of VoIP call by intentionally creating disruption during the call session. The analysis shows the dependability of the framework and indicates its limit to cope with such case. Eventually, the research is concluded after the objectives have been accomplished.

1.7 Thesis Organisation

The organisation of this thesis is arranged to facilitate a broad range of readers, either inside or outside the research field. The research presented involves a multidisciplinary knowledge within the domain of VoIP security and applied cryptography. In this chapter, the general overview of the research including problem statement, objectives, contributions, and research methodology is introduced. Chapter 2 reviews the fundamental research backgrounds in the domain. Besides, related public key verification approaches and VoIP steganography are studied as well. Chapter 3 defines the proposed verification scheme and its implementation on a cryptographic key agreement. In Chapter 4, the complexity and reliability of the proposed verification scheme are analysed and experimentally tested. The computational and storage cost of the hybrid framework is calculated and security analysis is discussed. Finally, the conclusion and the promising directions for the forthcoming research are revealed on Chapter 5.

CHAPTER 2

LITERATURE REVIEW

The emergence of cheaper technology and communication alternatives naturally attracts more users and malicious activities (Symantec, 2015). Hence, the landscape of potential security threats in the Internet has been continually evolving since then. As the consequence of being part of the Internet family, Voice over Internet Protocol (VoIP) systems are facing similar security risks. In order to design a reliable public key verification for VoIP security, this chapter reviews fundamental components in VoIP system and its major security concerns, especially on the communications. Besides, foundation in cryptography and well-known cryptosystems are investigated. Furthermore, several related public key verification methods and VoIP steganography are reviewed.

2.1 Voice over IP System

VoIP system provides an alternative telephony system over Internet Protocol (IP) networks such as local area network (LAN), intranet, and the Internet, which is termed specifically as Internet telephony. In general, IP networks are based on packet switching system that operates differently from conventional telephone networks. As demonstrated on Figure 2.1, a full-duplex in conventional telephones session is established using a circuit switching system. The telephone exchanges (switches) arrange a single channel every time a session is wanted. This channel is fixed and the bandwidth is fully dedicated for tranmitting voice signals throughout the session. In contrary, packet switching system is connectionless intrinsically. Each packet from a source is individually addressed to the destination over various routes based on the packet's information and the router's decision. The routes

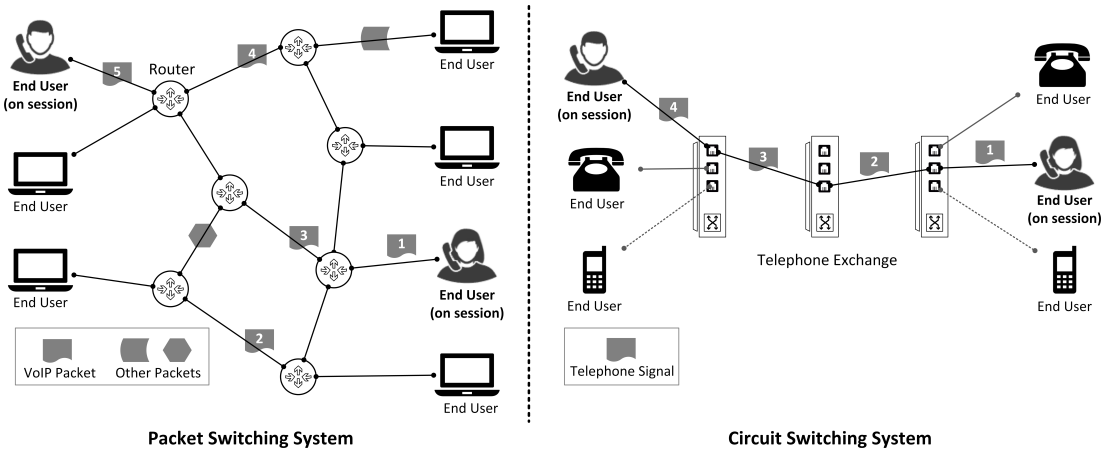

Figure 2.1: General concept of packet and circuit switching systems

are shared with several users for more than a telephone line. Hence, scenarios including packet loss, misallocated, delay, and traffic jam (bottleneck) are usual incurrences on packet switching systems.

These incurrences, especially packet loss and delay, cause quality degradations such as jitters, noises, and echoes, which can disrupt the communication (James et al., 2004). The issue of quality is one of the challenges in VoIP expansions. For many years, conventional telephones sustain a better quality of service (QoS) on both traditional analog and modern digital telephones, which has caused the lack of acceptance to VoIP technology at the beginning. Nevertheless, the recent VoIP quality has been significantly improved due to the presence of higher speed network and computing performance.

In order to construct a VoIP system, components such as server, client, protocol, and codec are commonly required as depicted in Figure 2.2. In communication systems, protocol is important to allow the communicating resources to work properly as intended. In telephony systems, there are two indispensable types of protocol to establish a call session between two participants or more as in a conference call, which include signaling and media transport. These protocols are specified in the application layer of the Transmission Control Protocol / Internet Protocol (TCP/IP) model as shown in Figure 2.3.

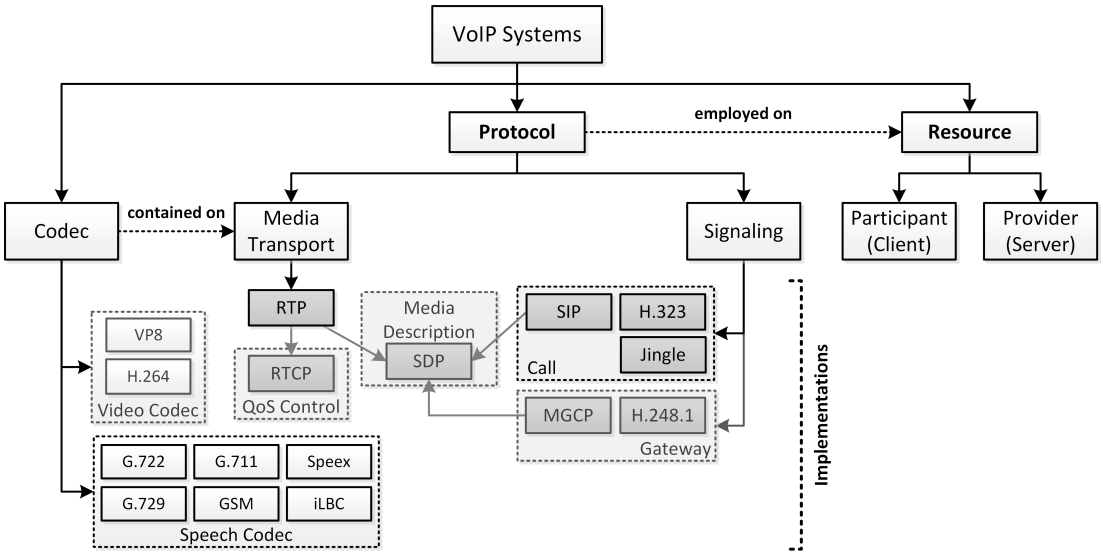

Figure 2.2: VoIP common requirements

Signaling protocol is responsible for controlling telephone sessions such as establishment, teardown, and transfer of the calls. The standards in signaling protocol are divided into two groups. The first group defines general signaling procedure for VoIP systems. Whereas, the second group, known as gateway signaling, is specifically to manage integration of VoIP systems with other telephony systems by translating signals or packets between the systems. The first group of signaling protocols serves as the core element in telephony systems for managing telephone sessions. In public switched telephone network (PSTN), signaling protocol like Signalling System No. 7 (SS7) is implemented and industrialised by telecommunications company in worldwide. In VoIP systems, standards such as Session Initiation Protocol (SIP) (Rosenberg et al., 2002), H.323 (ITU-T, 2009), and Jingle (Ludwig et al., 2016) are similarly employed to manage the signaling systems. These standards intend to handle the same case using different procedures.

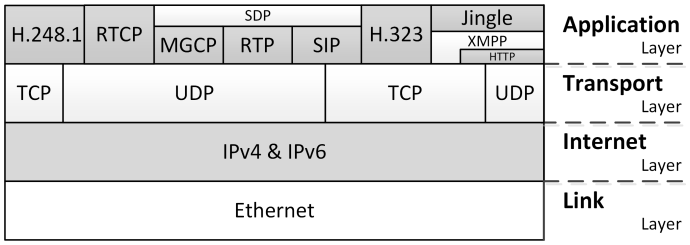

Figure 2.3: Encapsulation of VoIP protocols in TCP/IP model

Comparing the standards is rather difficult as these standards are continually revised. According to the development timeline, H.323 protocol was introduced much earlier than SIP and Jingle. H.323 was leading as the standard implementation for VoIP systems. However, H.323 is heavily based on SS7 that designated for PSTN, which makes the protocol harder to configure without a sufficient knowledge or an expert assistances (Goode, 2002). Early comparison by Rosenberg and Schulzrinne (1998) has stated that SIP provides higher extensibility with lower complexity than H.323. The rationale behind the argument is the procedures in H.323 are defined specifically to carry out the respective tasks, thus makes H.323 less flexible for the improvements (Basicevic et al., 2008). In addition, Jingle is developed as the signaling mechanism for Extensible Messaging and Presence Protocol (XMPP), formerly called as Jabber, which is successfully deployed on Google Talk (Saint-Andre, 2007). Jingle protocol is released later after H.323 and SIP. Hence, Jingle is considered at younger phase of development compared to SIP and H.323.

In the current development, SIP and Jingle are considered as the most potential signaling protocols for the future of the Internet applications. Jingle is based on Extensible Markup Language (XML) that is highly readable. Hence, Jingle is practicable to be communicated over Hypertext Transfer Protocol (HTTP)-based applications which are very familiar for the Internet users. Similarly, the elements in SIP have a comparable design as in HTTP. However, SIP works independently, instead of operating on top of another protocol like Jingle. This independency makes SIP simpler and easier to be detached from the overlying protocols. Therefore, SIP has drawn more attentions, especially for its future implementations (Liu and Mouchtaris, 2000).

2.1.1 Session Initiation Protocol (SIP)

SIP is initially designed for signaling and controlling multimedia communications on IP networks. According to Goode (2002) and Glasmann et al. (2003), SIP aims a high flexibility in the first place, thus evolutions or enhancements can be easily adapted. This attribute is indeed essential, especially due to the rapid development of the Internet systems. The most notable example is the IP Multimedia Subsystem (IMS) framework that employs SIP as the main protocol to manage and integrate several multimedia services, including VoIP system and IP television (IPTV).

SIP is operable within various transport layers. Normally, SIP performs a reliable handshake using the connection-oriented protocols such as Transmission Control Protocol (TCP) to ensure the signaling is carried out properly. The implementation is adjustable for User Datagram Protocol (UDP) for some reasons, including speeding up tasks, reducing overloads, and improving scalability (Rosenberg and Schulzrinne, 1998). Figure 2.4 shows the overall of the SIP handshakes during active SIP sessions. As illustrated, the SIP server is a middle party that bridges between two VoIP participants. The server requires client to register at the beginning in order for the clients to attain the authority in accessing the service. In the Internet services, it is usual custom for the users to move over devices and change IP address. Hence, the registration recurrently occurs to update the users' address at a time. The user is required to present the current address to initiate the registration. Subsequently, the server usually challenges the users on their authorisations. The users are able to engage in VoIP sessions whenever the registration is successful.

Typically, there are three main consecutive states in a SIP call session: session establishment, session online, and session teardown. Throughout the establishment, the caller initiates the call using an INVITE message. The caller simply appends the server's name

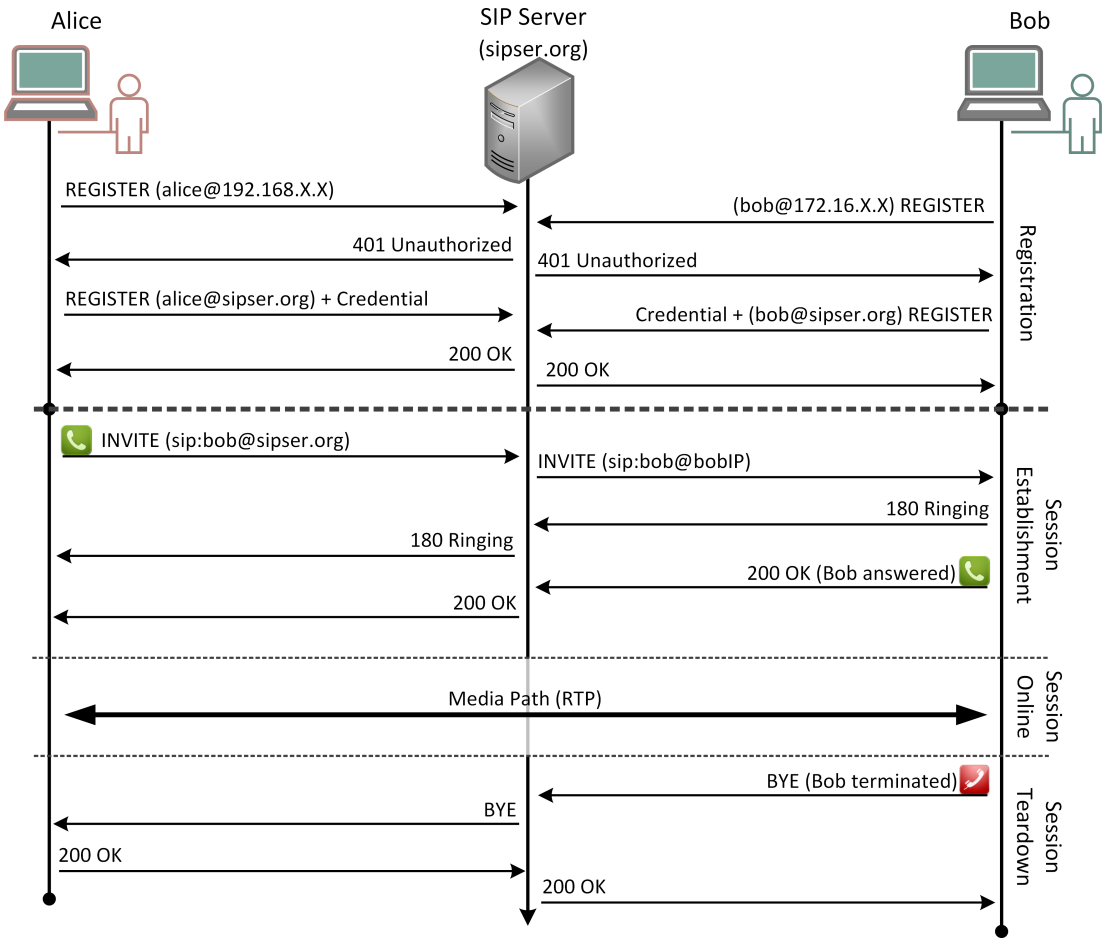

Figure 2.4: The handshakes on SIP sessions (Rosenberg et al., 2002)

on behalf of the callee's address and sends it to the server. The server assists the process of finding and connecting the caller with the callee. The server replies with either successful or failure response. In a successful request, the server forwards the INVITE message to the callee. The callee then responds with a 180 RINGING message. Finally, if the call is answered, the callee notifies the caller with a 200 OK message.

After the call is agreed (session online), the participants perform a direct real-time communication using the media transport protocol that is described in the next subsection. In order to terminate the session (session teardown), the initiator needs to send a BYE message to other participant and then close the media path. When the BYE message is received, the receiver replies a 200 OK message and close the path as well. Additionally, a session transfer is feasibly done during an online session when the users would like to

handover the session on different address. In this situation, the initiator has to inform the associated call partner using a REFER message that contains a new designated address. The associated participant then performs similar handshakes on session establishment to the new address and session teardown to the old address.

In terms of network infrastructure, SIP consists of some important elements such as user agent (UA), registrar, location service, redirect server, and proxy server. UA is a logical endpoint in SIP that is able to construct the SIP messages. In Figure 2.4, Alice and Bob act as the UA client (UAC) and `sipser.org` as the UA server (UAS). In details, registrar is a UAS that specifically handles registration process. It associates with a database known as location service to track the users' location. In order to locate the users, redirect server is a UAS that is used to enquiry the users' address from location service. Lastly, proxy server is an intermediary entity server that capable to act on behalf of both UAs. These elements can be distributed physically as long as the elements are working logically. The distributed infrastructure intends to enhance concurrencies, reduce loads, and avoid single point of failure (SPOF).

Figure 2.5 illustrates the events in Figure 2.4 and the roles of each UA. Each valid request or response message that is originated from UA may not be addressed directly. The messages may passed through proxy servers. The UAS and proxy servers are operated based on its configured behaviour, in either stateful (TCP-like) or stateless (UDP-like) modes. As illustrated, redirect server is configured in stateless mode. This configuration helps speed up in passing and responding any queries to location service without questioning ACK. Therefore, stateless mode is supportive for distributed systems, mainly to accelerate communications among the UAS that have a reliable connection.

Initially, SIP applies a client-server model, but then a considerable success of peer-to-

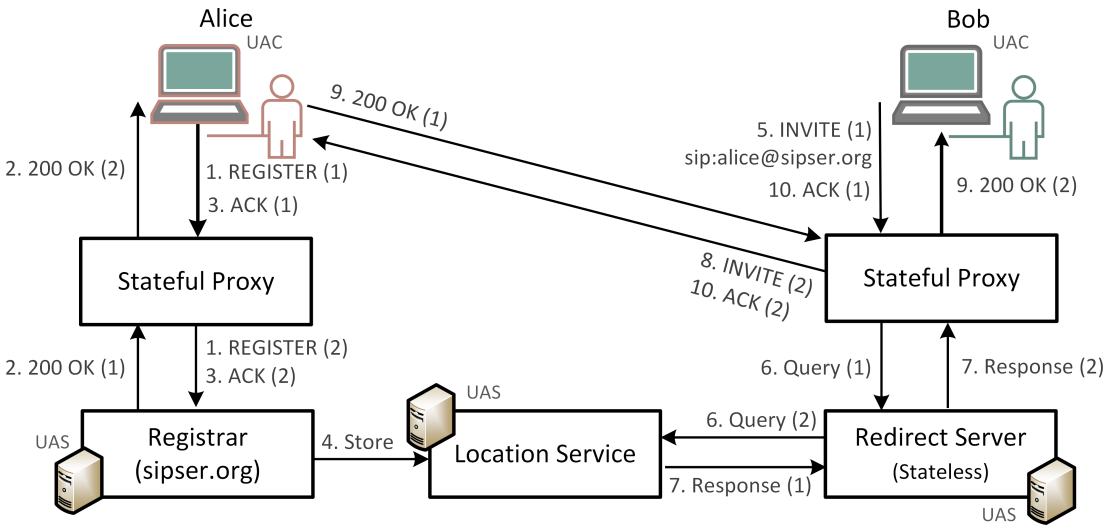

Figure 2.5: The SIP elements (Rosenberg et al. (2002))

peer (P2P) VoIP system like Skype has given an inspiration. As similarly intended, Singh and Schulzrinne (2005) has made the P2P model available for SIP as well. The model setups and distributes multiple nodes within the network in order to determine the best path for communication. The model improves reliability and scalability of the system. Despite the advantages of P2P model, the model increases security risks as the packets traverse through many nodes and routes (Seedorf, 2006).

SIP defines security mechanisms such as Secure / Multipurpose Internet Mail Extensions (S/MIME) and HTTP digest authentication. S/MIME by Ramsdell and Turner (2010) is adaptable in SIP to prevent modification on SIP messages during the transmissions. In addition, S/MIME can also be used to distribute digital certificates. S/MIME aims to achieve integrity and confidentiality of the MIME bodies or the whole body of a SIP message (Salsano et al., 2002). However, S/MIME does not provide a replay attack protection as the MIME bodies do not maintain any state (Gupta and Shmatikov, 2007).

HTTP digest authentication by Franks et al. (1999) is mainly used for authorising the user access to the SIP server. A mutual authentication between UAC and UAS takes place typically using a password-based authentication system. HTTP digest authentication

avoids the users to send a plain password. As the replacement, the user exchanges digested value of the credential and combines it with a one-time randomly generated value by the server (nonce). Nevertheless, HTTP digest authentication is not highly secured as it requires password negotiation beforehand (Salsano et al., 2002). The flaw of this mechanism can be exposed if the adversary deduces the secret during its negotiation or through comparisons of the digested values. Furthermore, this authentication is defenseless against man-in-the-middle (MITM) attack (Asokan et al., 2005).

Moreover, SIP can employ external security protocols to attain a higher security. The protocols such as Transport Layer Security (TLS) by Dierks and Rescorla (2008) and Internet Protocol Security (IPSec) by Kent and Seo (2005) are engaged to encrypt the communication channel between endpoints. The users have to signify the desire of using TLS by replacing "*sip*" with "*sips*" in the SIP messages similarly as HTTP Secure (HTTPS). Whereas, IPSec aims to establish a secure tunnel for all passing IP packets. However, IPSec requires manual pre-configuration on the desired endpoints, thus makes IPSec less convenient, especially when the users need to change to another address or device (El Sawda and Urien, 2006).

2.1.2 Real-time Transport Protocol (RTP)

Transmitting real-time speech distinguishes VoIP from other Internet services. As defined by (Schulzrinne et al., 2003), Real-time Transport Protocol (RTP) is a standard communication format for real-time media delivery in IP networks, especially speech and video. RTP is capable of transmitting data in unicast (to a single address) or multicast (to a group of addresses). RTP is encapsulated on UDP to afford a real-time continuous transmission. As consequence, RTP is neither able to ensure on-time delivery nor guarantee its quality. In order to address some quality issues, RTP optionally engages the RTCP that works

in parallel (out-of-band) to collect statistics and provide report on the quality of an RTP stream. This report can be used for discovering any transmission faults.

In terms of security, RTP shares the same liability as its underlying protocol. RTP does not comprehensively address the mechanism for its security. Instead, it relies on other implementations to resolve the matter. For instance, preserving the confidentially of the RTP packets is achievable through encryption using the proposed security protocols such as Secure RTP (SRTP), TLS, and IPSec.

The RTP payload carries the most essential data for VoIP systems and multimedia systems that employ RTP. It contains digital speech or video that has been adjusted for its transmission. The adjustment is managed through the codec. The type of codec used in the RTP sessions is defined on Payload Type (PT) field in the header. However, PT binds the codec to a session in a static way (Handley et al., 2006). Hence, description protocols such as Session Description Protocol (SDP) is employed on RTP to provide dynamic codec binding (Camarillo and Schulzrinne, 2010). In addition, SDP is also contained in signaling protocols such as SIP. This protocol allows changes on the codec or its attributes, i.e. clock rate and frame size, dynamically during the session. Furthermore, SDP is able to support joining multiple streams and synchronising codecs within a session.

2.1.3 Speech Codecs

The conversion between analog and digital often produces a heavy raw data. Thus, codec is required to transform digital data in a lighter format, mainly to ease data transmission. There are many types of codec used for VoIP call sessions. Each codec processes data differently and produces various quality and size, which can be a burden for outdated device or limited network bandwidth capacity. Hence, choosing a codec takes a serious

consideration between the needs and capabilities. Basically, codecs are a work of digital data compression that is run in the form of hardware device or software program. The compressions are common for audio, video, and text.

Codecs include process of encoding (compression) and decoding (decompression). Based on the compression results, a codec can be lossless or lossy. Lossless codecs intend to preserve the equivalent quality as the original data, which can be expensive for storage. Alternatively, lossy codecs degrade the original quality of the data to improve compression rate and produce lighter data. The quality loss in the results of lossy compressions should be imperceptible to human. Otherwise, the codec is not eligible for transparency, which is one of the ideal aspects in lossy compressions.

In VoIP systems, codec is one of the crucial aspects that determine the quality in the call sessions. Codec handles the digital data before and after the transmission of the RTP packets. Figure 2.6 demonstrates the flow of one-way speech transmission on RTP. Firstly, the speech is captured by a recording device, viz. microphone, in the format of analog signal. The analog signal is then converted into digital signal using analog-to-digital converter (ADC). This process is also known as digitisation. In analog telephones, digitisation is not required as the signals are exchanged in analog circumstance.

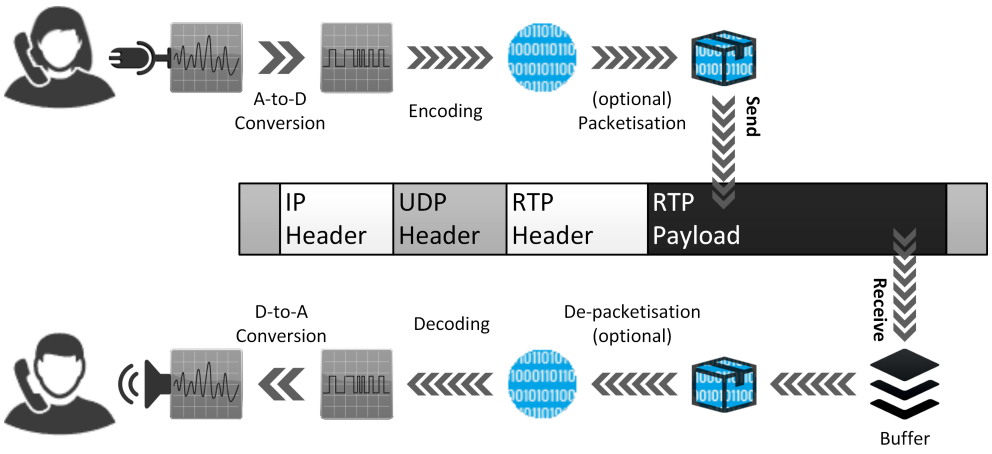

Figure 2.6: One-way speech transmission on RTP (adapted from Soares et al. (2008))