An Investigation into the Effect of Security on Reliability and Voice Recognition System in a VoIP Network

Ohnmar Nhway
University of Computer Studies, Mandalay, Myanmar
skynhway@gmail.com

Abstract— **Today, communications technology is one of the most essential roles for our life. People had used many ways of transmissions such as internet, mobile phone and so on. Moreover, VoIP (Voice over Internet Protocol) is also a communications technology. In this paper, we focus on VoIP which transmits voice over packet switched networks such as LANs (Local Area Networks) or WANs (Wide Area Networks). Adding security to a VoIP system, the quality of service and performance of the system are at risk. The contribution of this system has two main parts. Firstly, it illustrates a key from Diffie-Hellman Key Exchange Algorithm. This key is used by AES Algorithm for encryption/decryption process to secure the signaling and voice traffic within a VoIP system. Secondly, it evaluates the performance of a VoIP system to develop speech recognition at receiver. In this system, Mfccs (Mel-Frequency Cepstral Coefficients) is utilized for feature extracting from the incoming voice at receiver. Moreover, Gaussian algorithm is also used for vector quantization in feature matching MFCC features to represent raw speech signal into compact but effective representation that is more stable and discriminative than the original signal. And then the breakdown of call setup is applied by Counting Process in Stochastic Processes. So, this system had captured and analyzed registration, calls setup, voice transmission packets and finally recognized voice or speech by using Mfccs in DSP(Digital Signal Processing) technique.**

Keywords— **Voice over IP, Quality of voice, Soft-phones, SIP, RTP, and Cryptosystem.**

I. INTRODUCTION

VoIP (Voice over Internet Protocol) technology is used to transmit voice conversations using the IP (Internet Protocol) over a network using packets of data. The data network such as the Intranet or more likely the Internet has changed the strategy adopted by telecommunication managers. It is, therefore, one of the highest growth areas.

RTCP (Real-time Transport Control Protocol) provides feedback to the all members of the RTP session by a periodic transmission of control packets using the same distribution as data (e.g., multicast).

RTCP feedback reports on reception statistics on quality, (i.e., loss, delay, jitter), faults to diagnose network problems and distribution properties, (i.e., receivers of the session). RTCP facilitates the flow control & adaptive coding, but also multicast session surveillance. RTCP reports adapt to network capacities and session members.

In VoIP system, the aim of the conversion is to reduce the costs to home and business users by standardization of the network infrastructure. The popularity of VoIP is increasing rapidly due to cheap calls worldwide.

VoIP and related real-time communication applications, such as video conferencing and instant messaging, continue to attract considerable interest worldwide, with millions of active private business VoIP users today.

In this paper, we emphasized an investigation into security effect of performance in VoIP scheme, which is based upon ad-hoc network through the voice and Diffie-Hellman key exchange in security for using registration. On the other hand, registration means to check user identify correctly. And then, speech recognition means to check sender's voice at the receiver.

The remainder of this paper is organized as follows. In section 2, we describe the concept of the properties of the network. In section 3, we explain the proposed methodology and implementation. In section 4, we show experimental study and finally in section 5, we express the main points and suggest the future research.

II. PROPERTIES OF THE NETWORK

This section describes the requirements of VoIP system in detail. According to literature, VoIP is not emerged as popular without knowledge of Network. So, we study the appearing VoIP based on Network as follows:

2.1. Network Protocols

Internet communication is based on the internet protocol (IP) which is a network layer (layer 3) protocol according to the seven-layer open systems interconnection (OSI) model. The physical and data link layers reside below the network layer. On top of the network layer protocol, a transport layer (OSI layer 4) protocol is deployed for the actual data transmission. The transmission control protocol (TCP) as the transport protocol are used most internet applications. TCP is very robust since it allows for retransmission in the case that a packet has been lost or has not arrived within a specific time. However, there are obvious disadvantages of deploying this protocol for real-time, two-way communication. First and

foremost, can be a delay of very long time due to the retransmission process. Another major disadvantage of TCP is the increased traffic load due to transmission of acknowledgements and retransmitted packets. A better choice of transport layer protocol for real-time communication such as VoIP is the user datagram protocol (UDP). UDP does not implement any mechanism for retransmission of packets and it is more efficient than TCP for real-time applications [1].

2.2. Limitation of PSTN

Legacy telephony solutions are narrow-band. This property imposes several limitations on the achievable quality. In fact, in traditional telephony applications, the speech bandwidth is restricted more than the inherent limitations of narrow-band coding at an 8 kHz sampling rate. Typical telephony speech is band-limited to 300-3400 Hz. This bandwidth limitation explains why we are used to expect telephony speech to sound weak, unnatural, and lack crispness [5][7].

2.3. VoIP PBX

Voice over Internet Protocol (VoIP) is a general term for a family of transmission technologies for delivery of voice communications over IP networks such as the Internet or other packet-switched networks. Other terms frequently encountered and synonymous with VoIP are IP telephony, Internet telephony, voice over broadband (VoBB), broadband telephony, and broadband phone. VoIP systems usually interface with the traditional public switched telephone network (PSTN) to allow for transparent phone communications worldwide [9].

2.4. Pulse Code Modulation

If voice data are limited to frequencies below 4000 Hz, a conservative procedure for intelligibility, 8000 samples per second would be sufficient to characterize the voice signal completely. However, there are analog samples, called pulse amplitude modulation (PAM) samples. To convert to digital, each of these analog samples must be assigned a binary code. PCM starts with a continuous-time, continuous-amplitude (analog) signal, from which a digital signal is, produced [7].

VoIP systems employ session control protocols to control the set-up and tear-down of calls as well as audio codecs which encode speech allowing transmission over an IP network as digital audio via an audio stream. The use of Codec is varied between different implementations of VoIP (and often a range of codecs are used); some implementations rely on narrowband and compressed speech, while others support high fidelity stereo codecs.

In general, sender's voice (analog signal) is encoded such as digital signal before sender want to speak receiver at LAN in figure 1.

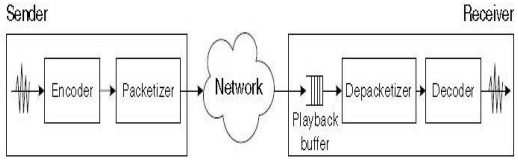

Figure 1. Components of a VoIP system

2.5. Network Characteristics

Three major factors associated with packet networks have a significant impact on perceived speech quality: delay, jitter, and packet loss. All three factors stem from the nature of a packet network, which provides no guarantee that a packet of speech data will arrive at the receiving end in time, or even that it will arrive at all. This contrasts with traditional telephony networks where data are rarely, or never, lost and the transmission delay is usually a fixed parameter that does not vary over time.

These network effects are the most important factors distinguishing speech processing for VoIP from traditional solutions. If the VoIP device cannot address network degradation in a satisfactory manner, the quality can never be acceptable. In the following sub-sections delay, jitter, and packet loss are methods to deal with these challenges are covered [8].

2.5.1. Delay

In traditional telephony, long delays are experienced only for satellite calls, other long-distance calls, and calls to mobile phones. This is not true for VoIP. The effects of excessive delay have often been overlooked in VoIP design, resulting in significant conversational quality degradation even in short-distance calls.

The International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) recommends in standard that the one way delay should be kept below 150 ms for acceptable conversation quality. Delays may be acceptable between 150 and 400 ms, but have an impact on the perceived quality of user applications. Latency larger than 400 ms is unacceptable [1].

2.5.2. Packet Loss

When a packet loss is some mechanism for filling in the missing speech must be incorporated. Such solutions are usually referred to as packet loss concealment (PLC) algorithms. For best performance, these algorithms have to accurately predict the speech signal and make a smooth transition between the previous decoded speech and inserted segment. Even if the packet losses are more spread out, the listening experience is then similar to that of having the packet losses occur in bursts [1].

2.5.3. Network Jitter

The jitter present in packet networks complicates the decoding process in the receiver device because the decoder needs to have packets of data available at the right time instants. If the data is not available, the decoder cannot produce continuous system. A jitter buffer is normally used to make sure that packets are available when it is needed [1].

III. PROPOSED METHODOLOGY AND IMPLEMENTATION

Communication is one of most important roles for all over the world to connect each other. For example, VoIP systems

(skype, gtalk and so on) are used together with user name and password account. The other authors examined to connect and speak each other based on the password. So, the cracker will be attacked easily and quickly the password using password cracker. And then this cracker pretended such as user to steal for relevant information. In this fact, we used Key Exchange Algorithm instead of user and password account for security between two parties. On the other side, we created registration using Key Exchange Algorithm such as authorized users. In this way, we presented to secure between sender and receiver and to recognize sender's voice at receiver using DSP technique.

In this system, VoIP scheme is made up of four phase approaches;

(i) registration,
(ii) call setup,
(iii) transmission of the voice information and
(iv) breakdown of the call.

VoIP Registrations and Call Setup using Diffie-Hellman Key Exchange Algorithm

If user wants to speak another person, this user must register as an authorized user for using registration function. Here, registration utilizes Diffie-Hellman Key Exchange Algorithm. This algorithm requires Central Directory. And then Central Directory, User A and User B must know the relevant formula and parameters. If a user wants to speak conversation, each user will have to request to Central Directory. Moreover, Central Directory is one of the most essential roles of this system.

TABLE 1. DIFFIE-HELLMAN KEY EXCHANGE ALGORITHM

Global Public Elements	
q	prime number
α	$\alpha < q$ and α is a primitive root of q
User A Key Generation	
Select private X_A	$X_A < q$
Calculate public Y_A	$Y_A = \alpha^{X_A} \mod q$
User B Key Generation	
Select private X_B	$X_B < q$
Calculate public Y_B	$Y_B = \alpha^{X_B} \mod q$
Calculation of Secret Key by User A and User B	
$K = (Y_B)^{X_A} \mod q = D_i$, $K = (Y_A)^{X_B} \mod q = V_i$	

In the following equation, assumes such as User A means D_i and User B also means V_i [2][3].

$$R(x) = D_i - V_i \begin{cases} \text{Accept, if } R(x) = 0 \\ \text{Reject, otherwise} \end{cases}$$

Let, R(x) = registration function
D_i = sender's secret key
V_i = receiver's secret key

Central Directory

If the central directory is trusted, then this form of communication provides both confidentiality and a degree of authentication. Because only User A and User B can determine the key, no other user can read the message (confidentiality). Recipient User B knows that only User A could have created a message using this key (authentication) [6].

Transmission of voice information using AES Algorithm (Advanced Encryption Standard Algorithm)

Encryption is one of the essential security technologies for computer data, and it will go a long way toward securing VoIP. Encryption is the process of transforming information (referred to as plaintext) using an algorithm (called cipher) to make it unreadable to anyone except those possessing special knowledge, usually referred to as a key. An encryption algorithm along with a key is used in the encryption and decryption of data.

In this system, we need an algorithm to generate keys (Gen), an encryption (Enc) algorithm and a decryption (Dec) algorithm. So, we choose AES (Advanced Encryption Standard) for encrypting and decrypting of voice data.

A triplet (Gen, Enc, Dec) of algorithms, a message space M and a key space K, is called a symmetric key encryption scheme if:

a. The key-generation algorithm: Gen is an algorithm that returns a key K using Diffie-Hellman Key Exchange Algorithm, denoted by $k \leftarrow$ Gen, such that $k \in K$.
b. The encryption algorithm: Enc is an algorithm that takes a key k and a voice-data $m \in M$, and outputs a cipher data $c \leftarrow$ Enck(m).
c. The decryption algorithm: Dec is an algorithm that takes a key k and cipher data c and outputs a voice-data m.
d. The scheme should satisfy the following property: For all $m \in M$ and $k \in K$,

[Deck(Enck(m))=m]

In this paper, we are discussing about the key generation method using Diffie-Hellman Key Exchange Algorithm in Cryptosystems.

Breakdown of the call

In this counting model, we find out two facts. The limitation call's times subtract the number of call's times. If the result is greater than or equal to the zero, there is no error. Then both sender and receiver continue to communicate and start conversation each other. So, we solve for this case using Counting Process in Stochastic processes [4].

For example, the event limits for making between "0" to "t" times. This event processes during starting "0" time to finishing "t" times ((0, t]) .

$$G(x) = t \quad - \quad \sum_{i=1}^{n} i \quad \begin{cases} \text{accept, } 0 \leq g(x) \leq t \\ \text{reject, otherwise} \end{cases}$$

where,

G(x) = counting function of ultimate destination
t = limitation time
n = number of call time, events

IV. Experimental Study

The experimental setup was carried out over a LAN environment. The two end systems were interconnected over 100 Mbps. The application of voice transmission is based on rtptools and robust audio tool. This system uses PCM(64 Kbps) data for primary data whose sampling rate is 8000 Hz for redundant data as experiment data [5].

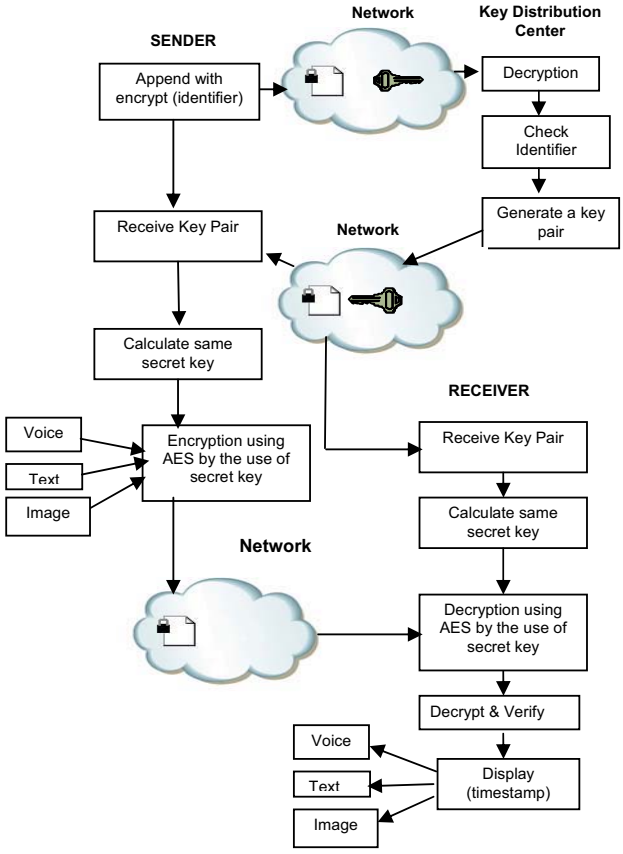

Figure 2. Overview of Connection with Sender and Receiver

Both sender and receiver were very friendly before this system was implemented by using registration and call setup. To protect the information transmitted some techniques employ encryption algorithms or apply cryptographic functions to the packet payload. And then sender transferred voice file to receiver in figure 2.

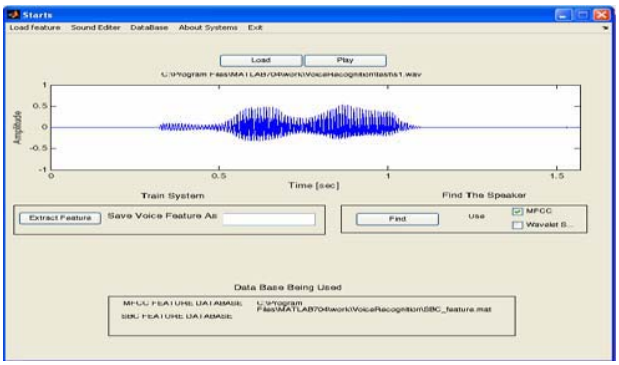

Figure 3. Speech Recognition System at Receiver

When the wave file (*.wav) arrives receiver, receiver needs to check this wave file who is speaking. So, receiver examines using Mfccs in DSP technique.

According to figure 3, we got an experimental result based on Mfccs feature extraction.

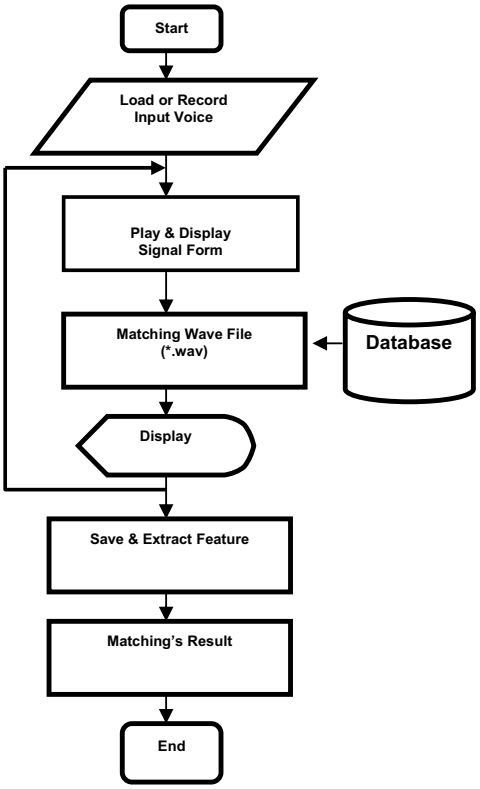

Figure 4. Speech Recognition System at Receiver

In Figure 4, this design was first step of my idea and then applied combination with java and Matlab 7.0.4 languages. Moreover, we proposed speech recognition system by using Mfccs feature extraction. This feature extraction is carried out by Gaussian and Discrete Cosine Transform at receiver [10].

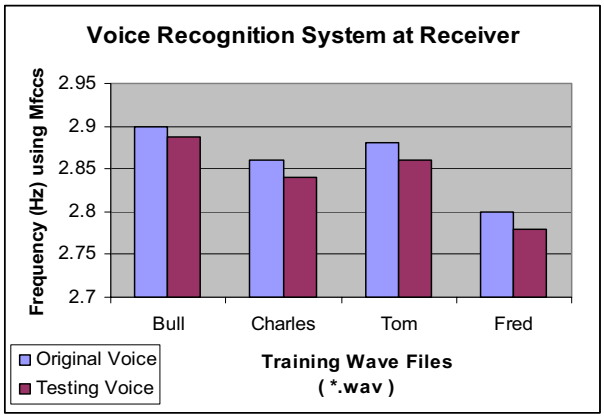

Figure 5. Experimental Result of Speech Recognition System at Receiver

This system's result is appeared as following figure 5. In this system, voice recognition will be implemented by Mfccs for testing and training voice (*.wav) files at receiver. So, it is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. This system contains training phase, testing phase and recognition phase. During these phases, features are extracted which is stored in the database and matched the template in the same database.

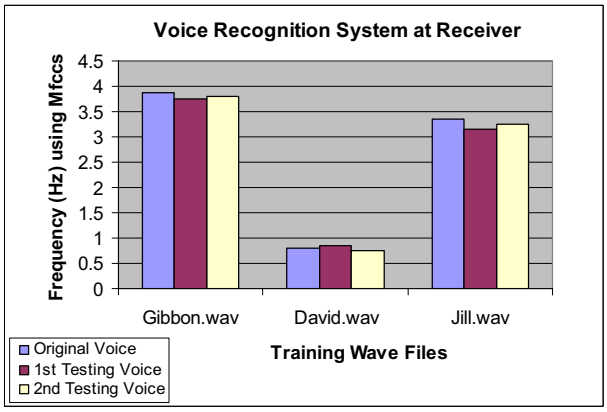

Figure 6. Results of each user at receiver

Finally, we have found that the result has a gap (little percentage %) between original voice (training voice) and testing voice according to figure 6. The main effort is to prove how to train and try a test for voice and how to apply with security. When adding security to a VoIP system, the quality of service and performance of the system are not completely result.

V. Conclusion And Future Research

This system is developed to provide security for the voice data files. The system uses registration function for the security providing medium. Generally passwords and smart cards are used for the security systems. This system uses Key Exchange Algorithm for the security system. Password can be hacked by trail and error basis. But, it is not possible to break the Key Exchange based security system. The system uses two methods for the process. The Diffie-Hellman Key Exchange Algorithm is used for the key build process and AES Algorithm is used for the encryption/decryption process to secure the signaling and voice traffic within a VoIP system. Then voice recognition system is implemented for testing using Mfccs in DSP(Digital Signal Processing) technique at receiver. The system is tested with various samples and the performance of the system is very good. The system is tested with different type of voice data.

Later, we will focus on Biometric Fingerprints for security system. The key is generated by Biometric Fingerprints. This key uses AES Algorithm for encryption/decryption to secure the signaling and voice traffic within a VoIP system. Finally, we emphasize for the performance of voice quality between traditional security system and biometric based security system in future research.

Acknowledgment

First and fore most, I would like to express my deep gratitude to Dr. Mie Mie Su Thwin for giving me the opportunity to attend the course work of Ph-D candidate and providing assistant to improve my strength.

I also wish to express my deepest gratitude to my parents and my relations for their encouragement, understanding and support throughout the period of doing my Ph-D candidate.

I am much obliged to my teacher Dr. Mie Mie Thet Thwin, Rector, and Dr. Win Aye, Pro - Rector, Dr. Than Naing Soe , Dr. Aye Thida, Dr. Mie Mie Khin, Daw Yee Yee Mon & other teachers, University of Computer Studies, Mandalay for their enthusiastic suggestion, stimulation and encouragement.

Finally, my special thanks are due to all my teachers who taught me and gave their knowledge to me from kindergartens (primary school, high school, University of Computer Studies, Yangon) to University of Computer Studies, Mandalay.

References

[1] J. Skoglund, E.Kozica, J.Linden, R.Hagen, W. B. Kleijn, "Voice over IP ; Speech Transmission over Packet Networks", Springer Handbook seof Speech Processing, Benesty, Sandhi, Huang, 2008, Pages-308, 309, 310, 311,313, 314.

[2] May Phyo Oo, "Managing Certificates of Grid Security Infrastructure", Ph.D(IT), University of Computer Studies, Yangon, Myanmar, 2007.

[3] Muhammed Tayyab Ashraf, John N.Davies and Vic Grout, "An Investigation into the Effect of Security on Performance in a VoIP Network", Centre for Applied Internet Research (CAIR) Glyndwr University, University of Wales, Wrexham, UK, 2009.

[4] SHELDON M. ROSS, "Stochastic Processes", University of California, Berkeley, 1983.

[5] Thinn Naing, Moe Pwint, "FEC-based Loss Control Mechanism for Optimizing Voice Communication Quality", Proceedings of the seventh International Conference on Computer Applications, Yangon, Myanmar, 2009, Page-129.

[6] *William Stallings,* Cryptography and Network Security Principle and Practices, Fourth Edition, Pages-298, 484, 492.

[7] *William Stallings,* Data & Computer Communications, Sixth Edition, Pages-85, 149.

[8] Available:http://www.informatik.hawhamburg.de/~schmidt, Aug. 2009.

[9] Available: http://www.theatem.com/VoIP-PBX

[10] Available:http://www.mathworks.com