International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:1, No:2, 2007

A Security Model of Voice Eavesdropping Protection over Digital Networks

Supachai Tangwongsan, and Sathaporn Kassuvan

***Abstract*—**The purpose of this research is to develop a security model for voice eavesdropping protection over digital networks. The proposed model provides an encryption scheme and a personal secret key exchange between communicating parties, a so-called voice data transformation system, resulting in a real-privacy conversation.

The operation of this system comprises two main steps as follows: The first one is the personal secret key exchange for using the keys in the data encryption process during conversation. The key owner could freely make his/her choice in key selection, so it is recommended that one should exchange a different key for a different conversational party, and record the key for each case into the memory provided in the client device. The next step is to set and record another personal option of encryption, either taking all frames or just partial frames, so-called the figure of 1:M. Using different personal secret keys and different sets of 1:M to different parties without the intervention of the service operator, would result in posing quite a big problem for any eavesdroppers who attempt to discover the key used during the conversation, especially in a short period of time. Thus, it is quite safe and effective to protect the case of voice eavesdropping.

The results of the implementation indicate that the system can perform its function accurately as designed. In this regard, the proposed system is suitable for effective use in voice eavesdropping protection over digital networks, without any requirements to change presently existing network systems, mobile phone network and VoIP, for instance.

***Keywords*—**Computer Security, Encryption, Key Exchange, Security Model, Voice Eavesdropping.

I. Introduction

THIS research presents a model of voice eavesdropping protection system over digital networks with a highly personalized security. In particular, the two parties of conversation could set their own encryption keys and rules, other than solely from their network service providers, in which we could expect a better level of performance in security.

In the second millennium, digital networking system is one of key infrastructures in the knowledge based society as it provides information and knowledge exchanges in forms of voice, data, image or video. Voice on the digital network is sliced up into small packets. The packets travel individually to their destination, not necessarily over the same route, and are reassembled in proper sequence when they arrive.

Security is required for reliable digital voice development. Unfortunately, securing voice over digital networks such as VoIP (Voice over Internet Protocol) is difficult as it is vulnerable to worms, viruses, denial of service and eavesdropping if not properly handled.

The present work is concerned with the problem of voice eavesdropping in particular and how to develop a model of protection system with a highly personalized security. The following, a working principle of the proposed model is described. Briefly speaking, the speaker's voice data are encoded according to the standard coding scheme, and then encrypted by his/her personal key before sending over the network. And a receiving device will decrypt the arrival packets with another personal key accordingly, and decode the resultant packets back into its original voice. As for a key used in the encryption, and another key used in the decryption processes are determined by the two conversation parties. Furthermore, the speakers in the conversation parties could set their encryption and decryption rules, for example, encrypt 20% or 25% of the total voice packets so-called the figure of 1:M. By this technique, it is rather difficult for eavesdroppers to discover the speaker's voice or the secret keys by applying the so-called cryptanalysis. This is because the eavesdroppers do not know which parts are encrypted (discussed later). In addition, partial encryption would definitely take less time than full encryption.

II. Literature Review

To develop the eavesdropping protection over the mobile phones or the Internet network, the following approaches should be reviewed [1], [3]-[8].

As for the security of the Internet network, in the approach of [1], the bit streams of G.729 [2] encoded voice data are partially encrypted. And the data are divided into two segments, one to be encrypted and the other to be left unencrypted. While in model [3], encoded inter-frames of voice data are interleaved at first to solve the problem of continuous packet loss. Then, the interleaved voice data are encrypted. Furthermore, SRTP [4] is the standard used in eavesdropping protection in RTP (Real-time Transport Protocol) traffic.

Supachai Tangwongsan is an Associate Professor of the Department of Computer Science, Faculty of Science, Mahidol University, Thailand (e-mail:ccstw@mahidol.ac.th).

Sathaporn Kassuvan is with the Department of Computer Science, Faculty of Science, Mahidol University, Thailand (e-mail:kengmeta@hotmail.com).

In the security of the mobile phones, we have studied the eavesdropping protection of the Global System for Mobile communications (GSM) network [6] [7]. Since the voice data are encrypted between the MS (Mobile Station) and the network only, the GSM system is not an end-to-end confidentiality service [7]. Moreover, the key used in the encryption process is determined by the network operator. And from the above-mentioned problems, this research [8] proposes the end-to-end security for GSM users. In doing so, the data channel of GSM is used to route the voice as data to secure the unencrypted link between the BTS (Base Transceiver Station) and MSC (Mobile Switching Center).

III. Conceptual Model

A. Conceptual Design of Eavesdropping Protection

1. Voice Data Encryption

This research proposes the model of voice eavesdropping protection over digital networks. In doing so, the following items should be defined.

1. Users could set their own keys, and different pairs could be set with different keys, not necessarily a single key should be used for all.

2. Users could set their own rules of encryption (1 : M block) (discussed later).

However, the encoded voice data before sending over the network are in the form of packets. Hence, in this work, partial-frame encryption scheme is investigated to be another choice in preventing eavesdropping, assuming the partially-encrypted voice data to be packet error. So, when the data are decoded, they cannot be understood as in the case of all frame encryption. Moreover, partial frame encryption takes less time in encryption and decryption than all frame encryption.

2. Key Management

In this work, the keys are sent between the parties in a different signal channel other than the signal channel of the conversation. It is recommended to use other media such as e-mail, SMS, etc. This is to avoid the keys from being trapped in eavesdropping. From this, only the communicating parties know the keys that create the exclusive-privacy circumstance.

B. Voice Data Transformation Model

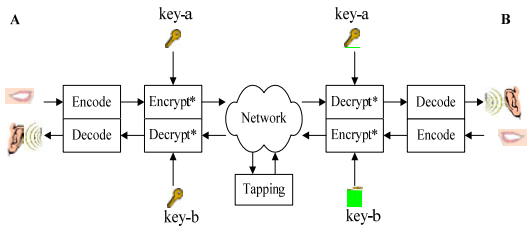

Fig. 1 Voice data transformation model

From Fig. 1, we describe how the voice data transformation model resolves the specified problems as follows. The process begins with encoding the voice data. Then, the encoded voice data are partitioned into blocks and encrypted with the key (Encrypt* means 1-to-M block encryption in this model).

Referring to the above model, the voice data are transformed into the uneavesdropped data and then sent over the network to the terminal receiver. After that, the receiver decrypts the voice data, using the key as well (Decrypt* means 1-to-M block decryption in this model). And finally, the voice data are decoded before transforming back to the voice signals that can be understood as the original voice.

1. Key Exchange Setup

In the following, we present the process of key exchange between the communicating parties: user A and user B via SMS (Short Message System). By this way, we use a sending key of 64 bits in the encryption and decryption schemes according to DES (Data Encryption Standard) algorithm as shown in Fig. 2.

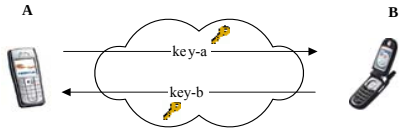

Fig. 2 Delivering a key between user A and user B

2. Encryption and Decryption

By this technique, the sender needs the same key used in the encryption process as the key used in the decryption process of the terminal receiver. And to enable standardization for widespread use, an algorithm exposure is recommended such as the DES algorithm.

As above-mentioned, the encryption and decryption start with user A's sending device by segmenting the packets of encoded voice data into blocks, 64 bits each. Then, each block of the data is encrypted with 1-to-M block (M is the number of blocks in each group of the packets). Simply mean, the first block of each group is encrypted with the key while the next block(s) (M-1 block) is (are) not encrypted as shown in Fig. 3: examples of 1-to-2 blocks and 1-to-3 blocks encryptions (the blocks shown in black color are encrypted blocks) in order.

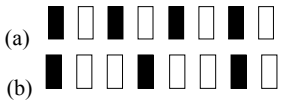

Fig. 3 Encrypted encoded voice data of 1-to-2 blocks is shown in (a) and 1-to-3 blocks is illustrated in (b)

In practice, each communicating party can choose a different ratio in the real-conversation scenarios. That is, they can specify the ratio of the encrypted and unencrypted blocks for 1-to-M block by themselves such as 1:1, 1:2, 1:3, 1:4, or 1:5. So, the eavesdroppers could not know which ones are encrypted and which ones are not which is difficult for the cryptanalysis. Moreover, by this technique, the communicating parties keep the ratio of 1-to-M block in secret for themselves.

C. Calculation of M Value

According to the proposed principle, when there is a packet error happening with the encoded voice data sent over the network, the quality of voice at the receiving terminal is lower and it cannot be understood. So, we apply this to the eavesdropping protection by comparing the encrypted encoded voice data with the packet error or the packet loss. Then, the calculation of the smallest number to be encrypted is presented in order to find the most suitable value that still could preserve the voice data security. In this case, we use a formula for the following calculation: the E-model from G.107 [10] to find the quality of voice or a MOS (Mean Opinion Score) derived from a transmission rating factor R as shown in the equations as follows:

$$\begin{aligned} &\text{For } R<0: && MOS = 1 \\ &\text{For } 0<R<100: && MOS = 1+0.035R+R(R-60)(100-R)7*10^{-6} \\ &\text{For } R>100: && MOS = 4.5 \end{aligned} \quad (1)$$

Here, R value is related to the packet loss as shown in the following equations.

$$R = R_o - I_s - I_d - I_e + A \quad (2)$$

Where R_o is Basic signal-to-noise ratio
I_s is Simultaneous impairment factor
I_d is Delay impairment factor
I_e is Effective equipment impairment factor
A is Advantage factor

$$I_e = I_m + ((95-I_m) * P_e / ((P_e/ B) + P_r)) \quad (3)$$

where I_m is Equipment impairment factor
P_e is Packet-loss probability
P_r is Packet-loss robustness factor
B is Burst ratio

From the above equations, we can determine the percentage of the packet loss of G.729 encoded voice that provides low voice quality and unknown one. From this, MOS of the voice after transformation is less than 2 that is not to be understood. And in the model, we place MOS=1.6 in an equation 1, obtaining R=29.82. And then, place R value in an equation 2, together with placing the sum of R_o and I_s variables (2 terms: intrinsic quality) that is 84.3 [11] in the equation 2. As for I_d and A variables, they are aside from the condition, so they are not used in this calculation. After that, we obtain I_e=54.47 and then place this value in an equation 3, together with I_m=10, B=1, and P_r=19 [12] [13]. Finally, the packet loss from the calculation is about 20%. And from the calculated packet loss we find that at least 20% of the encrypted voice data cannot be intercepted. Hence, M should be an integer number chosen from 1 to 5.

IV. Design and Implementation

The system consists of two main processes: Transmit and Receive processes as shown in Fig. 4: the sequence diagram of the system.

The key exchange starts with exchanging the key by the communicating parties before sending voice data over the network. At this point, the speaker sends the key used in encryption to the prospect receiver. As shown in Fig. 4, the voice data are transmitted over the voice data transformation system, while user A establishes the connection and then encryption. And next, user A sends the option of encryption, the encrypted M value, to user B. On the other hand, user B sends his/her own encrypted M value to user A. Then, the voice data are encoded and the encoded voice data are encrypted. After that, the data are sent over the network to the other party.

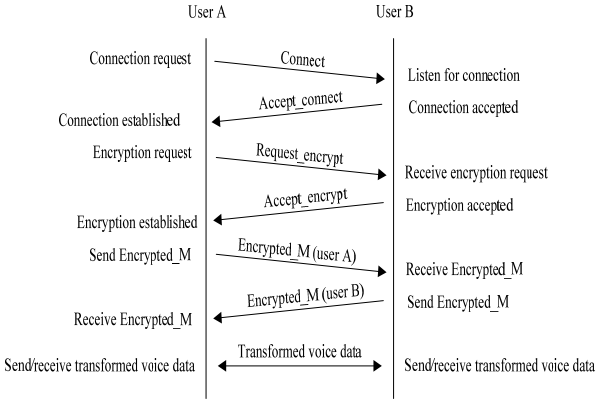

Fig. 4 Voice data transformation system process

In implementing the system, there are two main sections: key exchange and communication. By this way, we have implemented the eavesdropping simulation on a PC.

V. Experimental Results

In this experiment, we perform: *all* frame (encryption for all blocks of voice data) or *partial* frame (1-to-M block encryption; M is the number of blocks encrypted at the first block) encryption. In this regard, the time used in speech encryption and decryption has been taken into consideration. And the one that takes less time to do so could imply that it enables longer battery life, especially in mobile, portable devices such as the mobile phones that need limited battery power consumption. At the same time, we also consider the security from eavesdropping after encryption.

To learn that which encryption scheme takes less time and preserves speech security, we propose a variety of encryption schemes for G.729, G.723.1 [6.3 kb/s], G.723.1 [5.3 kb/s], G.726 [32 kb/s], and AMR [12.2 kb/s] encoded voice data.

A. Encryption of G.729 Encoded Voice Data

Here, a variety of encryptions for G.729 encoded voice data are tested as illustrated in the following three processes.

1. Preparatory Process for G.729 Speech Encoding Standard

At the beginning, speech is recorded for five minutes in the form of raw 16-bit mono PCM at a random frequency of 8,000 Hz. And after recording, the file size is 4,800,000 bytes.

Then, the file is encoded using speech encoding algorithm, the ITU-T G.729 standard at 8 kb/s, producing the new file size of 300,000 bytes to use in the next experimental process.

2. All Frame Encryption

There are three processes in speech encryption and decryption as shown in the following.

Step 1

At first, encrypt the file from the above preparatory process with the DES algorithm block by block (1 block has 8 bytes or 64 bits) until all blocks in the file have been completely encrypted. Take the time from the beginning to the end of encrypting all blocks in the file and then record the timing result in Table I.

Step 2

This step is the simulation of eavesdropping. After the encrypted file from step 1 is decoded, then we listen to this decoded file by using Sound Forge 4.5h software to open it. Listen to the speech and record the listening result in the table.

Step 3

This step is the process of normal decryption and decoding back by using the same file from step 1. At the same time, take the total time of decrypting and record the timing result in the table. Then, decode the file derived from the decryption process and open it by using the Sound Forge 4.5h software.

Step 1

In the beginning, encrypt the file from the preparatory process with the 1-to-M block encryption. Take the time from the beginning to the end of this encryption. Then, record the timing result in Table I.

Step 2

This step is the simulated eavesdropping. After the encrypted file from step 1 is decoded, then we listen to this decoded file by using the Sound Forge 4.5h software to open it. Record the listening result in the table.

Step 3

This step is the process of normal decryption and decoding by using the same file from step 1. At the same time, take the total time of decrypting and record the timing result in the table. Then, decode the file derived from the decryption process and open it by using the Sound Forge 4.5h software.

B. Encryptions of G.723.1 [6.3 kb/s], G.723.1 [5.3 kb/s], G.726 [32 kb/s], and AMR [12.2 kb/s] Encoded Voice Data

In this experiment, we study the change of the encryption periods of the remaining encoded voice data and then decode the file. Next, listen to the speech using the Sound Forge 4.5h software, as the same experimental processes as for G.729.

In summary, from Table I obviously, the more the unencrypted and the undecrypted blocks are, the less the encryption and the decryption times will be. Besides, in terms of eavesdropping simulation, it could be further concluded that the more the number of unencrypted blocks is, the less voice security will be.

TABLE I
THE COMPARISON OF LISTENING RESULTS, ENCRYPTION AND DECRYPTION TIMES, AND THE RATIO OF THE BEFORE-ENCODING AND AFTER-ENCODING FILE SIZES OF G.729, G.723.1 [6.3 kb/s], G.723.1 [5.3 kb/s], G.726 [32 kb/s], AND AMR [12.2 kb/s] SPEECH ENCODING STANDARDS

Encrypt / Decrypt	G.729 (Ratio : 16:1)			G.723.1 [6.3 kb/s] (Ratio : 20:1)			G.723.1 [5.3 kb/s] (Ratio : 24:1)			G.726 [32 kb/s] (Ratio : 2:1)			AMR [12.2 kb/s] (Ratio : 10:1)		
	Speech Quality	Encryption Time (sec.)	Decryption Time (sec.)	Speech Quality	Encryption Time (sec.)	Decryption Time (sec.)	Speech Quality	Encryption Time (sec.)	Decryption Time (sec.)	Speech Quality	Encryption Time (sec.)	Decryption Time (sec.)	Speech Quality	Encryption Time (sec.)	Decryption Time (sec.)
All blocks	1	0.11	0.10	1	0.08	0.08	1	0.07	0.07	1	0.9	0.91	1	0.16	0.15
1 block-to-2 blocks	1	0.06	0.05	1	0.05	0.04	1	0.04	0.04	1	0.48	0.49	1	0.09	0.09
1 block-to-3 blocks	1	0.05	0.05	1	0.03	0.03	1	0.03	0.03	1	0.38	0.37	1	0.07	0.07
1 block-to-6 blocks	4	0.02	0.02	2	0.02	0.02	2	0.02	0.02	1	0.26	0.27	1	0.05	0.05
1 block-to-11 blocks	4	0.02	0.02	3	0.02	0.02	3	0.01	0.01	1	0.21	0.22	2	0.04	0.04

A 4-point scale of listening: 4-fully understand, 3-fair, 2-poor, 1-not understanding at all

3. 1-to-M Block Encryption (M = 2, 3, 6, 11 in order)

In this section, let M be the block number, and we choose M value for 2, 3, 6, and 11 to perform the experiment. The details of the results are shown in the following.

VI. CONCLUSION

Finally, this proposed approach of voice eavesdropping protection over digital networks could meet the defined objective as shown in the experimental results and the system prototype. Therefore, given that a client device is equipped with those aforementioned features in the future, we are quite

confident to obtain an effective way to protect against voice eavesdropping.

REFERENCES

[1] A. Servetti, J. C. De Martin, Perception-Based Partial Encryption of Compressed Speech. IEEE Transactions on Speech and Audio Processing, Vol. 10, No. 8, November 2002.
[2] ITU-T G.729. Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP), March 1996.
[3] J.-I. Guo, J.-C. Yen, H.-F. Pai, New Voice over Internet Protocol Technique with Hierarchical Data Security Protection. IEE Proceedings-Vision, Image and Signal Processing, Vol. 149, No. 4, August 2002.
[4] M. Baugher, et al, The Secure Real-time Transport Protocol (SRTP). IETF, RFC 3711, March 2004.
[5] R. Blom, E. Carrara, F. Lindholm, K. Norrman, M. Näslund, Conversational IP Multimedia Security. IEEE Mobile and Wireless Communications Network, 2002.
[6] A. Mehrotra, L. S. Golding, Mobility and Security Management in the GSM System and Some Proposed Future Improvements. Proceedings of the IEEE, Vol. 86, No. 7, July 1998.
[7] Security Related Network Functions (Release 6), 3GPP TS 43.020 V6.1.0, December 2004.
[8] A. B. Rekha, B. Umadevi, Y. Solanke, S. R. Kolli, End-to-End Security for GSM Users. IEEE ICPWC 2005.
[9] ITU-T G.723.1. Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, March 1996.
[10] ITU-T G.107. The E-Model, a Computational Model for Use in Transmission Planning, March 2005.
[11] A. P. Markopoulou, F. A. Tobagi, M. J. Karam, Assessment of VoIP Quality over Internet Backbones. IEEE INFOCOM, 2002.
[12] ITU-T G.113. Transmission Impairments due to Speech Processing, February 2001.
[13] ITU-T G.113 Appendix I. Appendix I: Provisional Planning Values for the Equipment Impairment Factor *Ie* and Packet-Loss Robustness Factor *Bpl*, May 2002.
[14] ITU-T G.726. 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM), December 1990.
[15] Adaptive Multi-Rate (AMR) Speech Codec Transcoding Functions (Release 6), 3GPP TS 26.090 V6.0.0, December 2004.