An M-Sequence Based Steganography Model for Voice over IP

[*]Hui Tian, [*]Ke Zhou, [†]Hong Jiang, [*]Jin Liu, [‡]Yongfeng Huang, [*]Dan Feng

[*]Wuhan National Laboratory for Optoelectronics, School of Computer, Huazhong University of Science and Technology, Wuhan, China

[†]Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588-0150, USA

[‡]Network and Communication Research Institute, Department of Electronic Engineering, Tsinghua University, Beijing, China

{huitian, jinliu}@smail.hust.edu.cn, {k.zhou, dfeng}@hust.edu.cn, jiang@cse.unl.edu, yfhuang@tsinghua.edu.cn

***Abstract*—Differing from applying steganography on storage cover media, steganography on Voice over IP (VoIP) must often delicately balance between providing adequate security and maintaining low latency for real-time services. This paper presents a novel real-time steganography model for VoIP that aims at providing good security for secret messages without sacrificing real-time performance. We achieve this goal by employing the well-known least-significant-bits (LSB) substitution approach to provide a reasonable tradeoff between the adequate information hiding requirement (good security and sufficient capacity) and the low latency requirement for VoIP. Further, we incorporate the m-sequence technique to eliminate the correlation among secret messages to resist the statistical detection based on the fact that the distribution of the LSBs in the stego-speech is not uniform and to provide a short-term security protection of secret messages. To accurately recover secret messages at the receiver side, we design a synchronization mechanism based on the RSA key agreement and the synchronized sequence transmission using techniques of the protocol steganography, which can effectively enhance the flexibility of the covert communication system and be extended to other steganography schemes based on real-time systems. We evaluate the effectiveness of our model with ITU-T G.729a as the codec of the cover speech in StegTalk, a covert communication system based on VoIP. The experimental results demonstrate that our technique provides good security and transparency for transmitting secret messages while adequately meeting the real-time requirement of VoIP.**

***Index Terms*—Covert Communication, Information Hiding, M Sequence, Steganography, VoIP**

I. INTRODUCTION

As one of the effective solutions for Internet-based secure communications, steganography has attracted increasing interest. It provides a secure protection for secret messages by embedding them into digital media and making them inconspicuous and invisible to eavesdroppers. In contrast to the traditional cryptography whose purpose is to hide the content of secret messages being exchanged between the two communicating parties, the purpose of steganography is to hide not only the content but also its very existence. Therefore, steganography can offer a better security in many ways.

Recently, techniques with steganography at the core have been employed successfully in covert exchange of information [1-2], copyright protection [3], etc. However, most of the previous studies on steganography are carried out on storage cover media [4] and by contrast the area of steganography in real-time systems is largely unexplored. However, due to their instantaneity, real-time systems can potentially offer better security for hiding secret messages. Therefore, in this study we will focus on one of the typical real-time communication systems, Voice over IP (VoIP), as a possible carrier to apply steganography to enhance security for transmitting secret messages while maintaining good performance for VoIP real-time services.

VoIP is a promising technique to enable telephone calls via a broadband Internet connection. Owing to its advantages of low cost and advanced flexible digital features, VoIP has become a popular alternative to the public-switched telephone network (PSTN), and extensive research on it has been conducted [5]. The main motivations for our VoIP-based steganography study are twofold. First, the ongoing conversation of VoIP can offer an ideal camouflage for secret messages, because the voice data is naturally assumed to be the only data carried in a given VoIP channel. Second, a typically short VoIP connection does not give eavesdroppers sufficient amount of time to detect possible abnormity due to hidden messages.

From the literature [6-8], some researchers have noticed the advantages of and carried out useful studies on steganography over VoIP. Wang et al. [6] proposed a scheme for transmitting secret speeches based on information hiding in VoIP systems. Their hiding process consists of two steps: compressing the secret speeches and then filling their binary bits directly into the LSBs of cover speech coded with G.711. Dittmann et al. [7] presented a more general scheme of steganography over VoIP, which can be used to transmit arbitrary secret messages. However, both of these implemented steganography techniques only directly replace the LSBs of the cover speech with the binary bits of secret messages, which is vulnerable to detection by the steganalysis algorithm subsequently proposed by Dittmann et al. [7]. Their steganalysis algorithm is based on the fact that the distribution of the LSBs in the stego-speech is not uniform, which can detect directly embedded messages with a success rate of approximately 98.60%. Therefore, Dittmann et al. [7, 8] suggested that messages be encrypted prior to embedding to improve security.

The work is supported in part by National High Technology Research and Development Program of China (863 Program) under Grant Number 2006AA01Z444, National Basic Research Program of China (973 Program) under Grant Number 2004CB318201, Program for New Century Excellent Talents in University (No. NCET-06-0650) and National Natural Science Funds of China under Grant No. 60773140.

978-1-4244-3435-0/09/$25.00 ©2009 IEEE

Motivated by this view, they later proposed a scheme that introduces the cryptographies (i.e. Twofish, Tiger) for embedded messages [8]. However, the encryption operation must be carried out offline before the embedding operation, because the adopted cryptographies are often time-consuming and incur delays that may in turn degrade the speech quality drastically. Therefore, this method is not well suited for the real-time exchanging of secret messages. In fact, for the real-time covert communication we must strike an acceptable balance between providing adequate security and maintaining low latency for real-time services. In addition, the authors of [8] assumed that the two communicating parties must share the same knowledge of the used key, but did not reveal how the key is distributed, which is actually a crucial component for covert communication systems.

In this paper, we present a novel model of real-time steganography based on m-sequence in VoIP. Our technique aims at providing good security of real-time covert communication without sacrificing the real-time performance that is vital for VoIP. M-sequence has been employed widely in the code division multiple access (CDMA) technique and other spread spectrum communications. In our model, m-sequence is mainly used to eliminate the correlation among secret messages to resist the statistical detection mentioned above and to provide a short-term secure protection of secret messages. Moreover, we design a secure synchronization mechanism based on the RSA key agreement and the synchronized sequence transmission using techniques of the protocol steganography. The purpose of the synchronization mechanism is to safely distribute some important parameters to the receiver to correctly restitute secret messages. In addition, it is worth noting that this mechanism can also be extended for secure online distribution of key or/and important parameters in [8] and other steganography schemes based on real-time systems. Consequently, the two communication parties can momentarily decide to construct a covert communication during the working process of real-time systems (online) without the need for parameter requests or/and key agreements before running the systems (offline). Finally, we evaluate the effectiveness of the proposed model with ITU-T G.729a as the codec of the cover speech in StegTalk. The experimental results demonstrate that our technique provides good security and transparency for transmitting secret messages while adequately meeting the real-time requirement of VoIP.

II. M SEQUENCE

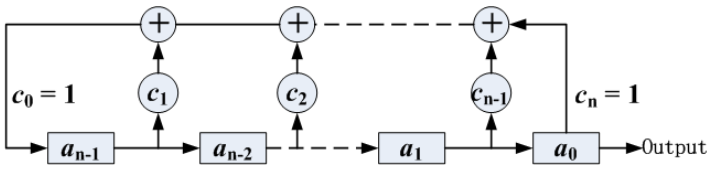

Fig. 1. The *n*-degree linear feedback shift register

M sequence is a typical pseudorandom binary sequence with the longest period. It has two properties that make it similar to zero-mean white noise: the numbers of ones and zeros are almost equal, and the autocorrelation function is as near to a delta function as we hope for [9].Therefore, it has been widely employed in the spread spectrum technique. M sequence is often generated by a linear feedback shift register (LFSR). Fig. 1 depicts an *n*-degree LFSR. In the figure, $a_i = 0$ or 1 ($i = 0, 1, \ldots, n-1$), which indicates the state of the *ith* shift register; $c_i = 0$ or 1, which indicates the state of the *ith* feedback line. If $c_i = 1$, the *ith* feedback line is connected; otherwise, it is disconnected. M sequence consists of all the output bits. In addition, the vector $\{a_0, a_1, \ldots, a_{n-1}\}$ indicates the current state of LFSR. Generally, if the current state of LFSR is $\{a_{k-n}, a_{k-n+1}, \ldots, a_{k-1}\}$, the next state is determined by the following steps: 1) outputting a_{k-n}; 2) shifting right all other bits by a position; and 3) calculating a new input of the *n*th register (a_k) by the following formula:

$$a_k = c_1 a_{k-1} \oplus c_2 a_{k-2} \oplus \cdots \oplus c_n a_{k-n} = \sum_{i=1}^{n} c_i a_{k-i} \pmod 2 \quad (1)$$

That is, the new state of LFSR depends on the following characteristic polynomial:

$$f(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n = \sum_{i=0}^{n} c_i x^i \quad (2)$$

If we choose one of the primitive polynomials as the characteristic polynomial, we can obtain the m sequence with the longest period $P = 2^n - 1$, i.e. $\{a_0, a_1, \ldots, a_{p-1}, a_0, a_1, \ldots, a_{p-1}, a_0, a_1, \ldots\}$. In addition, we can generate different m sequences with different initial states for an LFSR with a given degree. The total quantity of m sequences generated by an *n*-degree LFSR can be calculated via the following formula:

$$q(n) = (2^n - 1) \cdot \phi(2^n - 1) / n \quad (3)$$

where, $\phi(n)$ is the Euler's function that can be stated as follows:

$$\phi(n) = \begin{cases} 1, & n = 1 \\ \prod_{i=1}^{m} p_i^{e_i - 1} \cdot (p_i - 1), & n = \prod_{i=1}^{m} p_i^{e_i} \end{cases} \quad (4)$$

where, each p_i ($i = 1, 2, \ldots, m$) is a unique prime number that is different from all the others. Furthermore, if the degree of LFSR is chosen from 1 to *N*, the quantity of m sequences generated by LFSR with different degrees can be calculated by the following formula:

$$Q(N) = \sum_{n=1}^{N} q(n) \quad (5)$$

III. PROPOSED MODEL

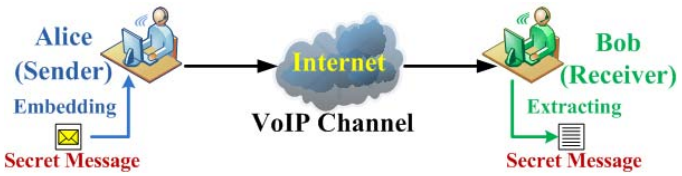

Fig. 2. Steganography in VoIP

Fig. 2 depicts a general framework of steganography in VoIP. Let us assume that Alice (the sender) wants to transmit secret messages to Bob (the receiver), while they are talking about some inconspicuous topics via the VoIP system. For that, Alice embeds

some secret messages into the voice with an embedding algorithm; after the voice is sent through the VoIP channel, Bob retrieves the secret messages with the corresponding restituting algorithm. In this paper, we construct our steganography model based on encryption with the m sequence and well-known LSB substitution that provides a reasonable tradeoff between the adequate requirement (good security and sufficient capacity) for information hiding and the low latency requirement for VoIP service.

A. Embedding and Restituting Algorithm

The purpose of the embedding algorithm is to hide secret messages behind the voice samples for security. Let us assume that Alice wants to send L bits of secret messages $M = \{m_i = 0 \text{ or } 1 \mid i = 0, 1, \ldots, L\text{-}1\}$; the n-degree LFSR is adopted to produce the m sequence $X = \{x_i = 0 \text{ or } 1 \mid i = 0, 1, \ldots, P\text{-}1, P = 2^n - 1\}$, where P is the period of the m sequence; The LSB set of each frame of the speech coded by a given codec is $B = \{b_i = 0 \text{ or } 1 \mid i = 0, 1, \ldots, S\text{-}1\}$, where S denotes the total number of the LSBs in each frame. The embedding algorithm can be formalized as follows:

$$M^* = \varphi(M, X, B) = \sum_{i=0}^{L-1}((m_i \oplus x_j) \otimes b_k) \tag{6}$$

where, $M^* = \{m_i^* = 0 \text{ or } 1 \mid i = 0, 1, \ldots, L\text{-}1\}$ is the secure form of M; $j = i \bmod P$; $k = i \bmod S$; $\oplus$ is the Exclusive OR operation; and $\otimes$ is the operation of bit substitution, i.e. $\forall\, y_1, y_2 = 0 \text{ or } 1$, $y_1 \otimes y_2 = y_1$.

The restituting algorithm, whose purpose is to extract the hidden secret messages, consists of two steps: extracting the cipher M^* and deciphering it. The process can be formalized as follows:

$$\begin{cases} M^* = \sum_{i=0}^{L-1} m_i^* = \sum_{i=0}^{L-1} b_k \\ M = \psi(M^*, X) = \sum_{i=0}^{L-1}(m_i^* \oplus x_j) \end{cases} \tag{7}$$

where, $j = i \bmod P$ and $k = i \bmod S$.

B. Synchronization Mechanism

Fig. 3. The structure of the header

The receiver often knows exactly the LSBs of voice samples and the generation algorithm of the m sequence that are adopted at the sender side. However, in order to restitute all secret messages accurately, the receiver must ascertain other three parameters: the length of secret messages (*LoM*), the degree (*D*), and the initial state (*IS*) of LFSR. Therefore, before sending the encrypted secret messages, the sender should first transmit the three parameters to the receiver. For the sake of privacy, the parameters should be encrypted with a public key cryptography. In our work, we choose the RSA algorithm that is employed widely in the key negotiation and the identification processes. Consequently, we can define a header as shown in Fig. 3. In the figure, *RSA* (*D*, *IS*, *LoM*) denotes the cipher (*C*) of the three parameters encrypted with the public key of the receiver. The parameter *LoC* indicates the total length of *C*. It is often set as a fixed size, in the interest of exact parsing. After retrieving *LoC*, the receiver can extract *C* from the LSBs of the voice samples and obtain the three parameters by deciphering *C* with his (her) private key.

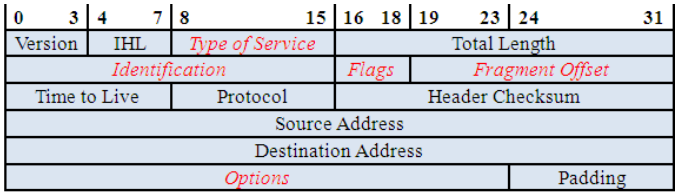

Fig. 4. The fields available for steganography in IP header

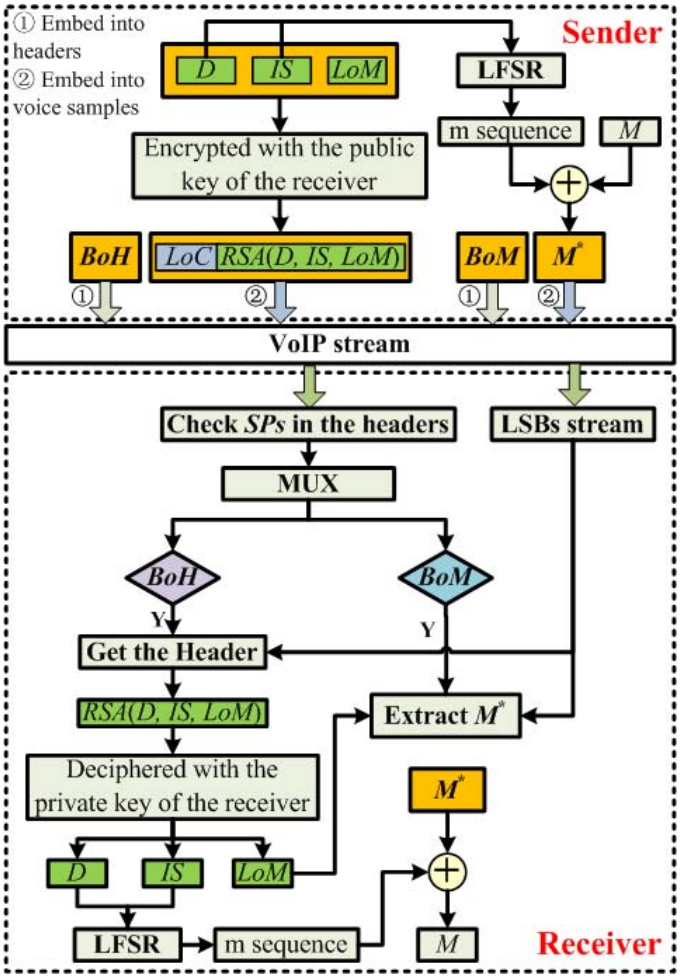

Fig. 5. The covert communication scheme based on VoIP

Furthermore, we define two synchronization patterns (*SPs*): *Beginning of Header* (*BoH*) and *Beginning of Message* (*BoM*) marking the start of the header and the start of the message respectively. Only upon detecting *SPs* can the receiver begin to extract the hidden data and carry out other relevant operations. In order to minimize the operations on voice samples, we do not embed *SPs* into the LSBs of voice samples, and instead employ the protocol steganography techniques that utilize the fact that few headers in packets are changed during transmission [10]. As described in [10], an IP header contains some fields that are

available for steganography. Those fields are marked in Fig. 4 with italics. The total capacity of the fields amounts to 64 bits per packet. There are some fields in UDP and RTP headers left to be used. Therefore, we can distribute the *SPs* among those fields in a predetermined fashion. *SPs* are often of short length (e.g. 8bits) and altered continually, so such a type of transmission is potentially hard to discover.

Using the above synchronization mechanism, we can describe the covert communication process as follows (as Fig.5 shows): the sender transmits the header (*H*) to the receiver by embedding *BoH* into the header of the first packet and *H* into the VoIP stream; after this, the sender transmits the encrypted secret message (M^*) by embedding *BoM* into the header of the first packet and M^* into the VoIP stream. At the receiver side, upon detecting *BoH* the receiver parses *D*, *IS* and *LoM* from the succeeding *H*, and waits for the forthcoming M^*. After detecting *BoM,* the receiver extracts M^* from the VoIP stream according the length indicated by *LoM*, and restitutes the original secret messages. In addition, there are other two key operations: 1) The sender must ensure that *SPs* do not appear in the header of each packet, if he (she) does not want to send secret messages; 2) Whenever not receiving any hidden information, the receiver must continuously check *SPs* so as to detect new transmissions of secret messages in a timely manner.

IV. Evaluation

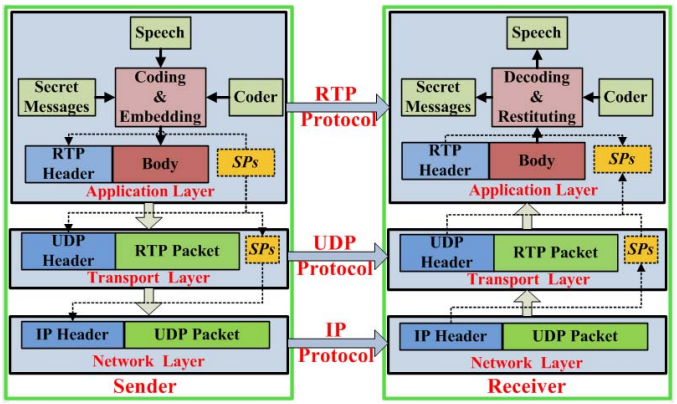

Fig. 6. Architecture of StegTalk

We evaluate the feasibility and effectiveness of our model in StegTalk, a prototypical covert communication system based on VoIP. Fig. 6 shows its architecture. In this system, the VoIP communication is constructed based on JVoIPLIB [11], a free library providing a primitive VoIP model. In addition, the system supports typical codecs, such as ITU-T G.711, G.723.1, G.729a, etc. In the tests, we choose G.729a as the codec of the cover speech. G.729a is an algorithm for the coding of speech signals at 8kbit/s using conjugate-structure algebraic-code- excited linear prediction, and operates on speech frames (80 bits per frame) of 10 ms corresponding to 80 samples at a sampling rate of 8000 samples per second [12]. Su and Huang [13] found that the parameters of fixed codebook in G.729a have the best transparency for information hiding. Motivated by this observation, we examine the impacts of these parameters, and define two covert rates, 8 bits per frame (0.8 kb/s, denoted by R_1) and 26 bits per frame (2.6 kb/s, denoted by R_2). Moreover, our system can vary the degree of LFSR from 1 to 60. In other words, according to Formula (5), the total quantity of m sequences that can be generated in our system is 1.4854×10^{34}, an astronomically large space in which the potential attacker can hardly decipher the embedded secret messages in a short time. Therefore, our system can provide sufficient short- term protection for secret messages. In the tests, we also focus on other two key issues, i.e. impacts of steganography on speech quality and the additional delay due to the proposed algorithms.

A. Speech Quality

In order to evaluate the impact of steganography upon speech quality, we contrast the spectrograms of the cover speeches without steganography and with steganography at R_1 and R_2 respectively. We choose the word "Hello" (P_1) and the phrase "It is nice to meet you" (P_2) as the cover speeches, and the text in section VI (references) as the secret message. Moreover, we randomly choose the degree and the initial state of LFSR. After encoding by G.729a, P_1 has 50 frames, and P_2 has 126 frames. Consequently, P_1 conceals 400 bits of message at R_1 and 1300 bits at R_2; P_2 conceals 1008 bits of message at R_1 and 3276 bits at R_2. Fig. 7 shows the spectrograms of the original speeches and their steganographical versions with secret messages embedded at R_1 and R_2 respectively. The spectrograms at R_1 show very slight differences from the original spectrograms, indicating that the speech quality degradation at R_1 is nearly negligible; the spectrograms at R_2 show a few visible differences from the original spectrograms, suggesting that the speech quality degradation at R_2 may be faintly perceivable, but not enough to impact the understanding of the cover speeches. Therefore, our technique can provide good transparency for steganography. These conclusions are also consistent with the results of the Mean Opinion Score (MOS) experiment that is used popularly in speech quality tests (See Table 1). In the MOS experiment, each MOS score is a mapping of perceived levels of the distortion into either the descriptive terms "excellent, good, fair, poor, unsatisfactory".

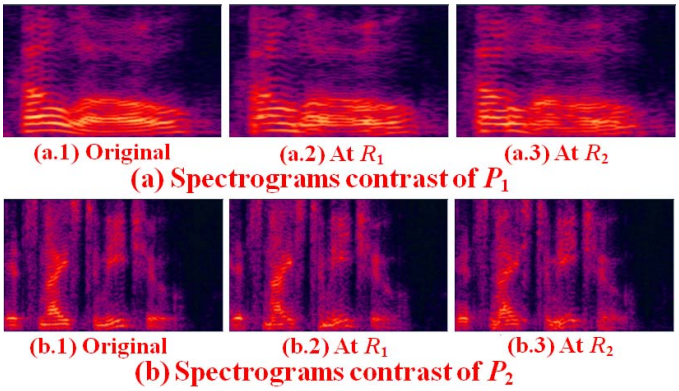

Fig. 7. Spectrograms contrast of P_1 and P_2

TABLE I Results of MOS Experiment

Speech	Original	At R_1	At R_2
MOS	4.0	3.7	3.0

B. Embedding Latency

Another crucial issue to evaluate is the performance impact of our algorithms on the real-time services of VoIP. To meet the real-time requirement of VoIP, the additional delay due to the

proposed algorithms must not noticeably impact the normal workings of VoIP. In our model, the additional delay is mostly induced by the embedding algorithm (the restituting algorithm can be carried out offline and consequently not increase the delay), so we focus on the processing time of the embedding algorithm. We test the average embedding time (*AET*) per frame using LFSR with different degrees at R_1 and R_2 respectively on Intel Celeron 2.66GHZ computers with 512M DDR2 SD RAM. From Fig. 8, we can observe the following facts: 1) *AET* increases with the degree of LFSR; 2) the maximum *AET* is approximately 5us at R_1 and 16us at R_2. That is negligible compared with the allowable coding time of 15ms for each frame. Therefore, we can safely conclude that our technique very adequately meet the real-time requirement of VoIP.

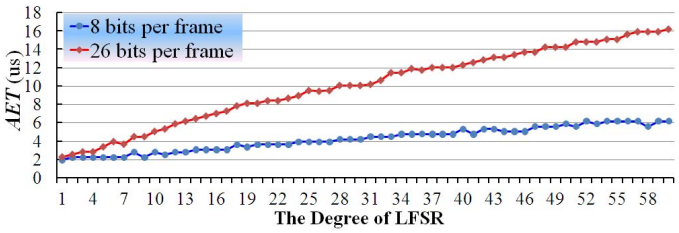

Fig. 8. The result of the test on AET per frame

In our previous work [14], we attempted to adopt a simple encryption scheme to encrypt secret messages before embedding operation. The encryption is based on an Exclusive OR operation between secret messages and a very large pseudo random number (PRN). We derived an efficient generation algorithm of PRN with arbitrary length from Mersenne Twiste [15] (a generation algorithm of PRN with 32 bits length). We ensured the privacy of our key by keeping the seed, the length and the generation algorithm secret, and encrypted the secret messages section by section. When used to transmit 1 MB secret messages, the steganography using the encryption increases the latency of VoIP system by about 4.7ms. While this latency is still acceptable considering the allowable maximum of 150ms one-way latency recommended by ITU-T G.114 [16], the m-sequence based technique proposed in this paper is significantly more efficient by encrypting the messages stream-like (bit by bit) during the coding process. Hence, we believe that the m-sequence based technique is a much better choice in the interest of maintaining the required performance of real-time services.

V. Conclusion

In this paper, we proposed a novel real-time steganography model for VoIP. The proposed model consists of three critical components: (1) a novel m-sequence encryption technique that eliminates the correlation among secret messages to resist detection by the statistical steganalysis algorithm and provides a short-term secure protection for secret messages, (2) a well-known LSB substitution that can provide an acceptable tradeoff between the adequate information hiding requirement for steganography and the low latency requirement for VoIP, and (3) a novel synchronization mechanism that ensures the accurate restitution of secret messages at the receiver side. Differing from the existing synchronization mechanisms, our mechanism introduces techniques of the protocol steganography for the transmission of synchronization patterns, which effectively minimizes the extra operations on voice samples and potentially reduces the operation time of synchronization patterns check. Moreover, this mechanism includes a method for secure online distribution of some important parameters, which makes it possible to construct the covert communication in real time without the need for offline communication requests in advance. That is a significant strategy for enhancing the flexibility of the covert communication system. In addition, it is worth noting that this mechanism can also be extensively applied in other steganography schemes based on real-time systems. We evaluated the effectiveness of our model with ITU-T G.729a as the codec of the cover speech in StegTalk. The experimental results show that our technique provides good security and transparency for transmitting secret messages while very adequately meeting the real-time requirement of VoIP service.

References

[1] N. Provos, P. Honeyman. “Hide and seek: an introduction to steganography”, *IEEE Security & Privacy Magazine*, vol. 1, Issue 3, pp. 32-44, May-June 2003.

[2] K. Bailey, K. Curran. “An evaluation of image based steganography methods”, *Multimedia Tools and Applications*, vol. 30, Issue 1, pp. 55-88, July 2006.

[3] E. T. Lin, A. M. Eskicioglu, etc. “Advances in digital video content protection”, *Proceedings of the IEEE: Special Issue on Advances in Video Coding and Delivery*, vol.93, No.1, pp.171-183, January 2005.

[4] M. Shirali-Shahreza. "A new method for real-time steganography", in *Proc. 8th Int. Conf. on Signal Processing*, vol. 4, 2006, pp. 16-20,.

[5] B. Goode. “Voice over Internet protocol (VoIP)”, *Proceedings of the IEEE*, vol. 90, Issue 9, pp. 1495-1517, Sept. 2002,.

[6] C. Wang, Q. Wu. “Information hiding in real-time VoIP streams”, in *Proc. 9th IEEE Int. Symposium on Multimedia*, pp. 255-262, 10-12 Dec. 2007.

[7] J. Dittmann, D. Hesse and R. Hillert. “Steganography and steganalysis in voice over IP scenarios: operational aspects and first experiences with a new steganalysis tool set”, in *Proceedings of SPIE, vol. 5681, Security, Steganography, and Watermarking of Multimedia Contents VII*, March 2005, pp. 607-618,.

[8] C. Kratzer, J. Dittmann, T. Vogel and R. Hillert. “Design and evaluation of steganography for voice-over-IP”, in *Proc. of 2006 IEEE Int. Symposium on Circuits and Systems*, pp. 2397-2340, 21-24 May 2006.

[9] S. Engelberg, H. Benjamin. “Pseudorandom sequences and the measurement of the frequency response”. *IEEE Instrumentation & Measurement Magazine,* vol. 8, Issue 1, pp. 54 -59, Mar. 2005.

[10] S. J. Murdoch, S. Lewis. “Embedding Covert Channels into TCP/IP”, in *Proc. of the 7th Information Hiding workshop*, June, 2005, pp. 247-262.

[11] Jori Liesenborgs. Jori's Voice over IP Libraty (JVoIPLIB), [Online]. Available: http://reserach.edm. luc.ac.be/jori/jvoiplib/jvoiplib.html.

[12] ITU-T, Recommendation G.729. “Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS- ACELP)”, Jan. 2007.

[13] Y. Su, Y. Huang and X. Li. “Steganography-Oriented Noisy Resistance Model of G.729a”, in *Proc. of Multi-conference on Computational Engineering in Systems Applications*, vol.1, 4-6 Oct. 2006, pp.11-15.

[14] H. Tian, K. Zhou, Y. Huang, etc. “A Covert Communication Model Based on Least Significant Bits Steganography in Voice over IP”, in *Proc. of the 9th Int. Conf. for Young Computer Scientists,* Nov. 2008.

[15] M. Matsumoto, T. Nishimura. "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator", *ACM Transactions on Modeling and Computer Simulations: Special Issue on Uniform Random Number Generation*, vol. 8, Issue 1, pp. 3-30, Jan.1998.

[16] ITU-T Recommendation G.114. “One-way transmission time, SERIES G: transmission systems and media, digital system and networks”, May 2003.