An information - theoretical model for streaming media based stegosystems

An Information - Theoretical Model for Streaming Media Based Stegosystems

YongFeng Huang [1], Shanyu Tang [2], Wanxia Yang [1]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, P. R. China. Email: yfhuang@tsinghua.edu.cn

[2] Corresponding author: School of Computer Science, China University of Geosciences, Wuhan, Hubei Province 430074, P. R. China. Tel: +86 27 6784 8563, Email: shanyu.tang@gmail.com

Abstract

Steganography in streaming media differs from steganography in images or audio files because of the continuous embedding process and the necessary synchronization of sender and receiver due to packet loss in streaming media. The conventional theoretical model for image steganography is not appropriate for explaining the security scenarios for streaming media based stegosystems. In this paper, we propose a new information-theoretical model with two pseudo-random sequences imitating the continuous embedding and synchronization characteristics of streaming media based stegosystems. We also discuss the statistical properties of Voice over Internet Protocol (VoIP) speech streams through theoretical analysis and experimental testing. The experimental results show the bit stream consisting of fixed codebook parameters in speech frames is similar in statistical characteristics to a white-noise sequence. The relative entropy between the VoIP speech

stream and the embedded secret message has been found to be zero. This leads us to conclude that the proposed streaming media based stegosystem is secure against statistical detection; in other words, the statistical measures cannot detect the existence of the secret message embedded in VoIP speech streams.

Keywords: steganography, steganographic model, VoIP

1. Introduction

Modern steganography seeks to provide a covert communication channel between two communicating parties over a public network. With the upsurge of multimedia applications on the Internet, steganography in streaming media has become the focus of attention in the research field of information security. Multimedia objects have been found to be perfect candidates for use as cover objects. This is largely due to the fact that multimedia objects often have a highly redundant representation, which usually permits the addition of significantly large amount of stego-data by means of simple and subtle modifications that preserve the perceptual content of the underlying cover object.

Shannon [1] first considered secrecy systems from the viewpoint of information theory, and identified three types of secret communications which he described as (i) concealment systems, (ii) privacy systems, and (iii) cryptographic systems. On concealment systems, i.e. steganography,

Shannon stated that such ‘systems are primarily a psychological problem’ and did not consider them further.

Over the past ten years, several theoretical models for information hiding have been proposed; for example, the modelling of the security of steganographic systems suggested by Federrath et al [2]. Most models [3][4][5] focused on protecting the signal sources of communication for steganographic systems based on watermarking and fingerprinting. However, little effort has been made to model the secret communication process.

Cachin examined secret communications from an information theoretical viewpoint. Cachin’s model [6] is the first to explicitly require that the stego-text distribution is indistinguishable from the cover-text distribution to an adversary. He assumed the adversary’s decision is based on statistical hypothesis testing as to whether the stego-object is drawn from the distribution of the cover object. He defined conditions for both a perfectly secure and an ε-secure stegosystems, under which the relative probability is either zero or negligible that the adversary cannot detect the existence of a covert communication.

However, Cachin’s model is not suitable for the real-time covert communications over public networks where streaming media play an essential role due to the following reasons:

a) Streaming media are dynamic data chunks of a series of IP packets, and secret messages can be embedded dynamically in some packets of media streams. Information hiding in streaming media

is a continuous embedding process, and it requires both the communicating parties to continually negotiate a series of encrypted-keys. As Cachin's model assumes the stego object is encrypted by a simple key only, this conventional model does not meet the requirements of managing the series of keys.

b) In a real-time covert communication, the synchronization between embedding and extracting is one of the major problems since packet loss occurs in streaming media. Large message data will likely be split across multiple packets. The message data split across multiple packets may arrive out of order, and one or more parts of the data split across multiple packets may not arrive at all. Therefore, the receiver should know where to extract the secret message from the cover object (*i.e.* which packets of media streams). But Cachin's model falls short of synchronization mechanisms.

c) Under the same constraints on security, a covert communication over streaming media needs larger embedding capacities and higher embedding rates in comparison with watermarking and fingerprinting. Cachin's model failed to address how to enhance the embedding capacity for steganography in streaming media.

In an attempt to overcome the problems of the above-mentioned model for image steganography, a new information theoretical model for steganography in streaming media is proposed in this study. Our proposed model introduces two pseudo-random sequences imitating the continuous embedding process and the necessary synchronization of sender and receiver due to packet loss in streaming media. For verification purposes, the proposed model was applied to a VoIP based

stegosystem, i.e. a real-time covert communication over the Internet. According to our model, streaming media based stegosystems have been proved to be ε-secure against a passive adversary. Experimental results for the passive adversary detection are also included in this paper.

The rest of this paper is organised as follows. In Section 2 Cachin's steganographic model is analysed. Section 3 describes our information theoretical model for steganography in streaming media. The application of the proposed model to VoIP based stegosystems and a number of statistical experiments on VoIP speech streams are detailed in Section 4. Section 5 presents passive adversary tests using μ detection on VoIP based stegosystems. Finally, Section 6 concludes with a summary and directions for future work.

2. Analysis of Cachin's Model

Following the approach of information theory, Cachin viewed the adversary's task of distinguishing between a normal cover object (like an image) and a stego object containing a secret message as a problem of hypothesis testing. Cachin presented a mathematical description of the statistical distance between the cover object and the stego object from the viewpoint of the relative entropy [6]. He also defined conditions for both a perfectly secure and an ε-secure stegosystems. The stegosystem suggested by Cachin is shown in Figure 1.

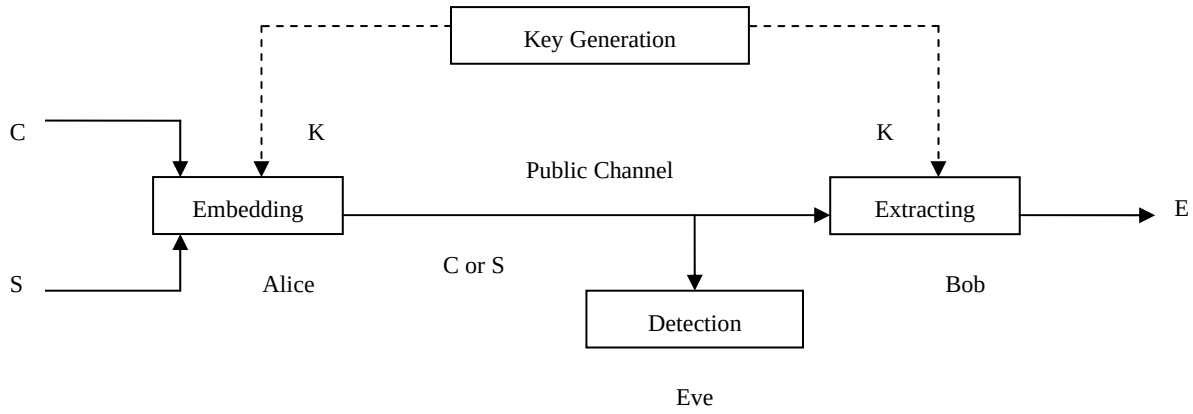

Figure 1: Cachin's steganographic model

Definition 1. A stegosystem with covertext C and stegotext S (Figure 1) is called ε-secure against a passive adversary if the relative entropy $D(P_C \,||\, P_S) \le \varepsilon$. If $\varepsilon = 0$, the stegosystem is regarded as perfectly secure.

In the above definition, $D(P_C \,||\, P_S) = \sum_{q \in Q} P_C(q) \log \frac{P_C}{P_S}$, where P_C and P_S are the probability distributions of the covertext C and the stegotext S, respectively, and Q is the space of possible measurements. Cachin considered Eve's decision performance using the theory of hypothesis testing; the two errors that can be made in a decision are called a type I error and a type II error. If Eve fails to detect the stegotext S, she makes a type II error, and its probability is denoted by β. The opposite error is that Eve decides that Alice sent stegotext although it was a legitimate cover text, and its probability is denoted by α. And the binary relative entropy is given by

$$d(\alpha, \beta) = \alpha \log \frac{\alpha}{1-\beta} + (1-\alpha) \log \frac{1-\alpha}{\beta}.$$

Lemma 1. Let P_{Q_0} and P_{Q_1} be two plausible probability distributions over the space of possible measurements Q. For any function $f: Q \to \mathrm{T}$, if $T_0 = f(Q_0)$ and $T_1 = f(Q_1)$, then $D(P_{T_0} \| P_{T_1}) \le D(P_{Q_0} \| P_{Q_1})$.

Theorem 1. In a stegosystem that is ε-secure against a passive adversary, the probability β that the adversary does not detect the presence of the hidden message and the probability α that the adversary falsely announces the presence of the hidden message satisfy $d(\alpha, \beta) \le \varepsilon$. In particular, if $\alpha = 0$, then $\beta \ge 2^{-\varepsilon}$.

The above analysis can lead to the following inferences:

a) Given a cover text C, Alice constructs the embedding function from a binary partition of the covertext space C' such that both parts are assigned approximately the same probability under P_C. In other words, the mathematical description of ε is given by

$$\varepsilon = \min_{C' \subseteq C} \left| \sum_{c \in C'} P_C(c) - \sum_{c \notin C'} P_C(c) \right| \tag{1}$$

The sender designs an embedding algorithm in such a way that the two parties of the stegosystem should have the same probability distribution.

b) The physical meaning of ε depicts the statistical distance between the cover object and the secret message. It has an intrinsic characteristic, which can be used to quantify the security of the stegosystem. A small value of ε means the statistical distance is small and it is more difficult for the adversary to use a statistical method to detect the hidden secret message, the stegosystem is

then regarded as ε secure. If $\varepsilon = 0$ and $\beta = 1$, the adversary would never detect the secret message; in this case, the stegosystem can be called ‘perfectly secure’.

3. Our Proposed Steganographic Model

As stated in Introduction, Cachin’s stego-model is not applicable to streaming media based stegosystems. Streaming media are different from traditional cover objects, such as plaintext files, images in BMP or JPEG format, and audio files in WAV or MP3 format [7-10]. The former can be used for dynamically embedding secret messages in IP packets in a real-time manner, but the latter are static cover objects in which secret messages are embedded statically.

Because of packet loss in streaming media based stegosystems, both the communicating parties need to periodically negotiate the private key so as to continuously transmit the secret message. However, Cachin’s model is not capable of addressing the continuous negotiation of the private key in the streaming media based stegosystems. In addition, Cachin’s model falls short of synchronization mechanisms, so that the receiver would not be able to identify what packets of streaming media contain part of or a secret message.

To tackle the above problems, we propose a new theoretical model with two pseudo-random sequences for steganography in streaming media. One sequence is used as the private-key for encrypting a secret message; the other sequence is utilised to provide the necessary

synchronization of sender and receiver in streaming media. The proposed model is applicable to VoIP based stegosystems, and can then be used to analyse the security of the stegosystems. The details of the proposed model for steganography in streaming media are described below.

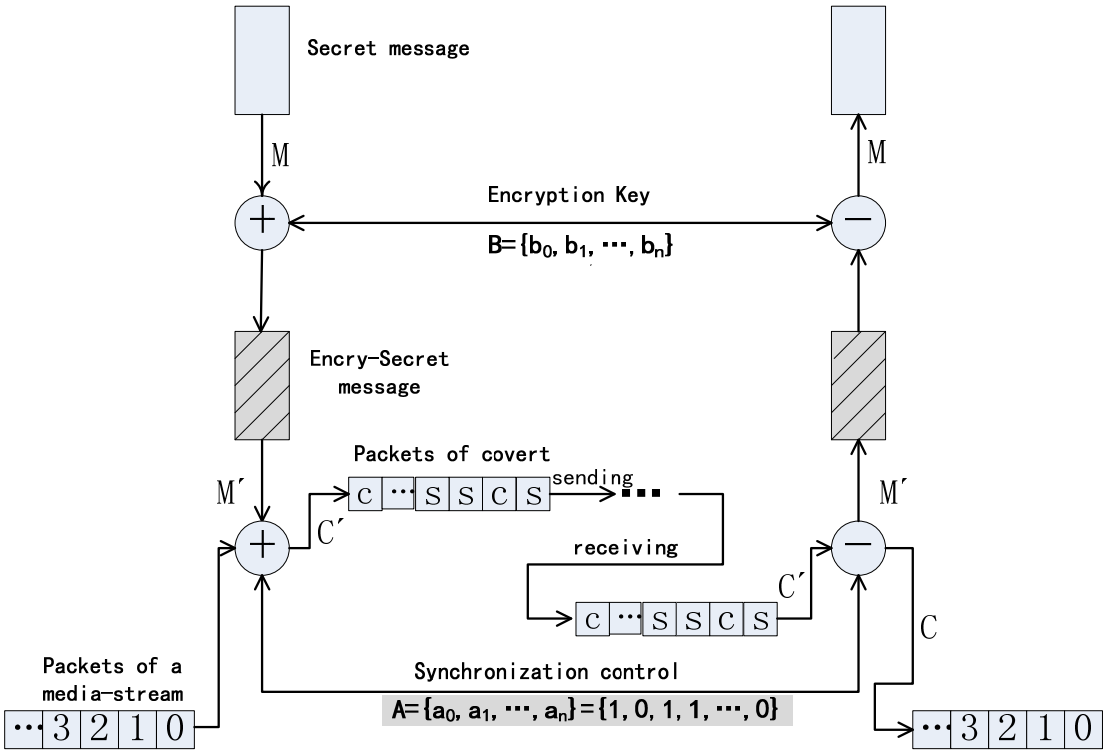

Figure 2: The proposed model for steganography in streaming media

Assuming that Alice and Bob share two key sequences (set of variables), $A = a_0 a_1 a_2 \ldots$ and $B = b_0 b_1 b_2 \ldots$, respectively. The sequence *A* is an *N*-level *M* pseudo-random sequence with the period of $(2^n - 1)$; and in one period the "1" appearance probability is $\eta = 2^{n-1} / (2^n - 1)$. The sequence *B* is the shifting sequence of the sequence *A*. The sequence *B* is used as the encryption key while the sequence *A* is used to imitate the necessary synchronization of sender and receiver in covert communications. In other words, the sequence *A* is used to decide what packets of

streaming media are used for embedding secret messages. For the *i*th packet of a media-stream, for example, if $a_i = 0$ Alice sends an 'innocent' packet *C* of the media-stream; if $a_i = 1$, Alice then sends a stego packet *S* containing a secret message. Figure 2 is the schematic description of the proposed model for steganography in streaming media.

The statistical distance between the cover object and the secret message can be expressed as $\varepsilon = \sum_{q \in Q_0} P_C(q) - \sum_{q \notin Q_0} P_C(q)$, $Q_0 \subset Q$, where P_C is the probability distribution of the cover object, Q_0 is a plausible space, and Q is the total space of possible measurements. Assuming Alice and Bob share $Q_2 \subset Q$, let $\sum_{q \in Q_2} P_C(q) = \eta$, $Q_3 = Q - Q_2$, where Q_2 and Q_3 are the observation spaces in relation to the stego object and the cover object respectively. It is often appropriate to model an information source as a stochastic process *U*. If the message sent in the *i*th packet is denoted by U_t, the stochastic process $\{U_t, t = 0,1,2, \ldots\}$ represents the stego-packet *S* or the cover-packet *C*. Then the total probability distribution of the secret message sent over the space *Q* is given by

$$P_U(q) = P(T \in Q_2)P_S(q) + P(T \in Q_3)P_C(q) = P(a_i \in Q_2)P_S(q) + P(a_i \in Q_3)P_C(q) \quad (2)$$

It becomes $P_U(q) = \eta P_S(q) + (1-\eta) P_C(q)$. As

$$P_S(q) = \begin{cases} P_C(q)/1+\varepsilon, & q \in Q_0 \\ P_C(q)/1-\varepsilon, & q \in Q_1 \end{cases} \quad (3)$$

where Q_0 and Q_1 are two plausible spaces of possible measurements, then

$$P_U(q) = \begin{cases} \eta \dfrac{P_C(q)}{1+\varepsilon} + (1-\eta)P_C(q) = P_C(q)\dfrac{1+\varepsilon(1-\eta)}{1+\varepsilon}, & q \in Q_0 \\ \eta \dfrac{P_C(q)}{1-\varepsilon} + (1-\eta)P_C(q) = P_C(q)\dfrac{1-\varepsilon(1-\eta)}{1-\varepsilon}, & q \in Q_1 \end{cases} \quad (4)$$

Following information theory, the relative entropy between the cover object and the stego object

for steganography in streaming media is given by

$$
\begin{aligned}
D(P_C \,\|\, P_U) &= \sum_{q \in Q} P_C(q) \log \frac{P_C(q)}{P_U(q)} \\
&= \sum_{q \in Q_0} P_C(q) \log \frac{1+\varepsilon}{1+\varepsilon(1-\eta)} + \sum_{q \in Q_1} P_C(q) \log \frac{1-\varepsilon}{1-\varepsilon(1-\eta)} \\
&= \frac{1+\varepsilon}{2} \log\left[\frac{1+\varepsilon}{1+\varepsilon(1-\eta)}\right] + \frac{1-\varepsilon}{2} \log\left[\frac{1-\varepsilon}{1-\varepsilon(1-\eta)}\right] \\
&\le \frac{1+\varepsilon}{2}\left[\frac{\varepsilon\eta}{1+\varepsilon(1-\eta)}\right] + \frac{1-\varepsilon}{2}\left[\frac{-\varepsilon\eta}{1-\varepsilon(1-\eta)}\right] = \frac{\varepsilon^2\eta^2}{1-\varepsilon^2(1-\eta)^2}
\end{aligned} \tag{5}
$$

For a *N*-level *M* pseudo-random sequence, $\eta = 1/2$, equation (5) becomes

$$D(P_C \,\|\, P_U) \le \varepsilon^2 / (4 - \varepsilon^2) \tag{6}$$

The relative entropy $D(P_C \,\|\, P_U)$ does satisfy the condition defined in Definition 1, i.e.

$$D(P_C \,\|\, P_U) \le \varepsilon^2 / (4 - \varepsilon^2) \le \varepsilon \tag{7}$$

Therefore, equation (7) proves the proposed model for steganography in streaming media is ε-secure against a passive adversary.

4. Application of Our Model to VoIP Based Stegosystems

This section details the application of our proposed model to a VoIP based stegosystem, and the statistical properties of VoIP speech streams which were studied using theoretical analysis and experimental testing.

4.1 Theoretical analysis of statistical characteristics of VoIP stegosystems

Suppose a VoIP stegosystem uses a speech codec of G.729a. A VoIP stream consists of a number of speech frames, which have a parameter called Fixed Codebook. A secret message can be embedded in some bits of the Fixed Codebook in speech frames. According to G.729a codec algorithms, Fixed Codebook describes the quantitative error between the raw speech and the predicted speech, and the end of the signal process model is a linear prediction error filter. The transfer function of the linear prediction error filter is defined as $A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}$, where a_i is the coefficient. The output $e(n)$ of the error filter is the difference between the raw speech $s(n)$ and the predicted speech $\hat{s}(n)$, given by

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i) \tag{8}$$

where $\hat{s}(n)$ is denoted by a linear combination of overpass samples, $s(n\text{-}1)$, $s(n\text{-}2)$, …, $s(n\text{-}p)$. Rearranging equation (8) yields

$$s(n) = e(n) + \sum_{i=1}^{p} a_i s(n-i) \tag{9}$$

Replacing n with $(n - j)$ equation (9) becomes

$$s(n-j) = e(n-j) + \sum_{i=1}^{p} a_i s(n-j-i) \tag{10}$$

Because the mean square error $E[e^2(n)]$ of the output $e(n)$ has to be minimized, the partial derivative of the coefficient a_i can be obtained from $E[e^2(n)]$ by setting the partial derivative with respect to a_j to zero, i.e. $\frac{\partial E\left[e^2(n)\right]}{\partial a_j} = 0,\ 1 \le j \le p$. It has $\frac{\partial E\left[e^2(n)\right]}{\partial a_j} = -2E[e(n)s(n-j)] = 0$. Substituting equation (10) into this formula obtains:

$$E\left[e(n)\cdot\left[e(n-j)+\sum_{i=1}^{p}a_i s(n-j-i)\right]\right]=0\Rightarrow E\left[e(n)e(n-j)+\sum_{i=1}^{p}a_i e(n)s(n-j-i)\right]=0 \quad (11)$$

Since $e(n)$ is the linear combination of $s(n)$, if P is big enough, equation (11) derives

$$\sum_{i=1}^{p}a_i e(n)s(n-j-i)=0 \quad (12)$$

Equation (12) indicates that $E[e(n)e(n-j)]=0,\ j\geq 1$ is tenable. So we can infer that the output of the error filter, $e(n)$, is of linear independence, which means the predicted quantitative error has the characteristics of a white-noise sequence. In fact, the purpose of adjusting the filter coefficient, according to the principle of the minimum error variance, is to make the output $e(n)$ to be a white-noise sequence.

4.2 Experimental simulation of statistical characteristics of VoIP stegosystems

Three statistical methods used for analysing data redundancy in this study are “0/1” probability, “0/1” jumping frequency and run-length [11].

1) “0/1” probability of speech streams

The “0/1” probability method requires the calculations of “0” and “1” numbers appearing in a VoIP speech stream as well as their proportions. The proportions of “0/1” probabilities ought to be equal if all samples are independent [13].

Table 1: The “0/1” probabilities of Fixed Codebook parameters in G.729a speech samples

	Sample1	Sample2	Sample3	Sample4
“1” probability	0.503062	0.478601	0.511638	0.495734
“0” probability	0.496938	0.521399	0.488362	0.504266

Let N be the frame length of a speech stream. Assuming the speech stream Y is divided into L frames, $Y_1Y_2Y_3 \ldots Y_L$. The probability of ‘1’ appearing in the speech stream Y is given by

$$Y_{p1} = \frac{1}{LN}\sum_{i=0}^{L-1}\sum_{j=0}^{N-1} y_{i,j} \tag{13}$$

where $y_{i,j}$ is the jth (j = 1, 2, …, N) column value in the ith (i = 1, 2, …, L) frame. Table 1 lists the “0/1” probabilities of Fixed Codebook parameters calculated using equation (13) for four different G.729a speech samples. The statistical results indicate that Fixed Codebook parameters have similar “0” and “1” probabilities, which is in accord with the distributive characteristics of a pseudo-random sequence.

2) The jumping frequency of “0/1” in speech streams

The jumping frequency of “0/1” is the number of “0” and “1” switching in the same column of the neighbouring frames in a speech stream. Assuming an initialisation $Y_{0,j}$ = 0, then $CZR(j) = \frac{1}{L}\sum_{i=1}^{L-1}\left|y_{i,j} - y_{i-1,j}\right|$. Figure 3 shows the jumping frequency of “0/1” distribution pattern with 80 bits in a frame. The results indicate the “0/1” jumping frequency of the Fixed Codebook in speech frames is high and close to 1. This is probably due to the randomicity of

Fixed Codebook. This phenomenon can also be demonstrated by the run-length of Fixed Codebook in speech streams, as detailed in the next section.

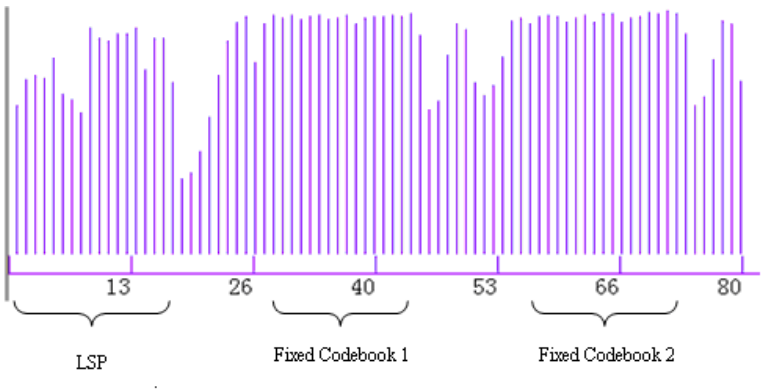

Figure 3: The jumping frequency of "0/1" in a G.729a speech stream

3) Run-length statistical method

The run-length statistical method was used to calculate the run-length of continuous "0" or "1" appearing in a bit stream. Assuming the bit stream $b_1, b_2, b_3 \cdots$ satisfies the following equation:

$$\begin{cases} b_k \neq b_{k-1} \\ b_{k+i} = b_{k+i-1}, \quad i = 1,2,\cdots,R-1 \\ b_{k+R} \neq b_{k+R-1} \end{cases} \tag{14}$$

In other words, the run-length of the bit stream is equal to the number of bits from b_k to b_{k+R-1}, where R denotes the run-length of the bit stream. Figure 4 shows the run-lengths of Fixed Codebook in four G.729a speech streams. The statistical results indicate that the run-lengths of all the four speech streams are very small. For example, up to 50% of the run-lengths of "1" equal 1. The run-lengths of "0" are quite similar to the run-lengths of "1" for all the used samples. These findings mean the experimental results are in agreement with the distributive characteristics of a

pseudo-random sequence. The results also prove the bit stream consisting of Fixed Codebook in G.729.a speech frames is similar to a white-noise sequence.

4.3 ε-Secure against a passive adversary in VoIP based stegosystems

The theoretical analysis and experimental simulation described above indicate that the linear prediction error filter can be regarded as a counter-approximate process [14], and also prove that the filter parameter $A(z)$ can be adjusted to make the output $e(n)$ analogous to a white-noise sequence. In addition, the statistical characteristics of the stego-objects containing secret messages are also similar to a white-noise sequence. According to the definition of ε in equation (1), it is reasonable to infer that the packets of VoIP streams have the same statistical characteristics as the encrypted secret messages; that is, the relative entropy that represents the statistical distance between the speech stream and the secret message is close to zero. Therefore, VoIP based stegosystems are determined to be 'perfectly secure' against statistical detection providing the secret messages are encrypted before embedding. This proves our proposed steganographic model is appropriate for describing the security of streaming media based stegosystems.

Figure 4: The run-lengths of Fixed Codebook in speech streams

5. Passive Adversary Test on VoIP Based Stegosystems

In order to prove the correlations between subsequent packets in experimental simulation (Sections 4.2 & 4.3) could not be used to break the stegosystems, the μ detection was used to further evaluate the security of VoIP based stegosystems in this study. Comparisons in probability distributions between the predictions from the proposed theoretical model and the experimental data were used to decide whether the cover objects (VoIP streams) contain a secret message.

Table 2: Estimated confidence levels for continuous embedding experiments

Continuous embedding	Data1	Data2	Data3	Data4	Data5	Data6	Data7	Data8	Data9	Data10
Pseudo-random sequence	3.070	3.333	3.626	3.258	3.674	3.248	3.551	3.712	3.094	3.286
Natural number	7.129	6.928	6.935	7.183	7.047	7.531	6.989	6.827	6.582	7.016

A series of experiments were conducted by employing STEGOTALK software to generate VoIP based stegosystems. Figure 5 illustrates the experimental environment in which two PCs with STEGOTALK being installed were deployed over the China Education and Research Network (CERNET). Alice and Bob used STEGOTALK to communicate covertly over the CERNET, and Mary, acting as a passive adversary, monitored the covert communication between Alice and Bob.

The STEGOTALK software utilises G.729a codec to encode the speech first and then uses Session Initiation Protocol (SIP) to initiate a Session, which is transmitted over VoIP. The principle of the software is referred to Figure 2 in which steganography in streaming media is carried out by employing two sets of sequences – encrypted keys and synchronization controls.

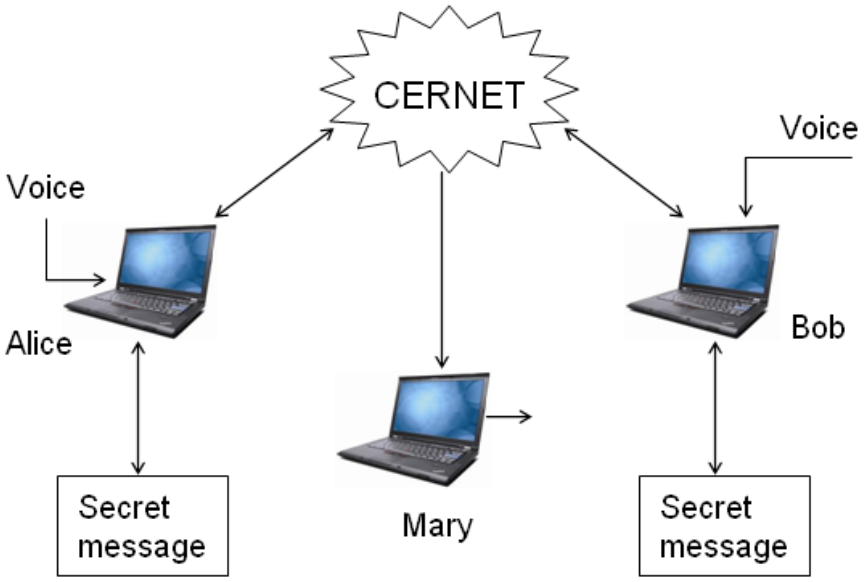

Figure 5: The experimental environment in which Alice and Bob using STEGOTALK

The speech streams of G.729a codec were used as cover objects. Ten test samples (Data1, Data2, …, Data10) were the speech streams with secret messages embedded in a continuous manner, i.e. every packet of the speech streams contained a secret message. Pseudo-random sequences and natural numbers were used as secret messages, respectively. The μ detection was used to estimate the embedding confidence level, and the results are listed in Table 2.

For comparison purposes, other specially designed experiments were conducted (Table 3). A key sequence was used in these experiments to ensure only part of VoIP packets were used to embed secret messages, in contrast with the previous experimental group (Table 2) that the embedding rates were 100%. The above-mentioned *N*-level *M* pseudo-random sequences were used to choose what VoIP packets to be used for embedding secret messages. In the experiments, half the packets were used to embed secret messages; that is, the percentage of the packets containing secret

messages is 50%. Similar to the first experimental group (Table 2), pseudo-random sequences and natural numbers were employed as secret messages, and the adversary employed μ detection to discover the secret messages. Table 3 are the detection results for the 50% embedding experiments.

Table 3: Estimated confidence levels for 50% embedding experiments

50% embedding	Data1	Data2	Data3	Data4	Data5	Data6	Data7	Data8	Data9	Data10
Pseudo-random sequence	3.216	2.950	2.847	2.784	3.018	2.918	3.174	3.658	2.958	3.085
Natural number	6.037	6.253	6.721	7.012	7.115	6.932	6.358	6.712	6.256	6.159

A detailed analysis of the results obtained from the two experimental groups (Tables 2 and 3) reveals that there are no much differences in the estimated confidence levels between the continuous embedding experiments (every packet contains a secret message) and the 50% embedding experiments (only half the packets have secret messages embedded). This means it is unlikely to use the statistical measures to detect the existence of the secret messages embedded in the G.729a speech streams. In other words, VoIP based stegosystems are perfectly secure against statistical detection as long as the encrypted secret messages are embedded in the Fixed Codebook of G.729a speech streams.

6. Conclusions

Steganography in streaming media is different from steganography in images / audio files because of the real-time embedding and synchronization characteristics of streaming media based stegosystems. The proposed information theoretical model with two pseudo-random sequences is applicable to streaming media based stegosystems. The performance of the proposed model has been evaluated by applying the model to a VoIP based stegosystem to analyse its security. Experimental results indicate the statistical characteristics of the bit stream consisting of fixed codebook parameters in speech frames are analogous to a white-noise sequence. The relative entropy between the speech stream and the encrypted secret message is found to be zero. This leads to the conclusion that it is unlikely to use the statistical measures to detect the existence of the secret messages hidden in speech streams provided that the messages are embedded in the Fixed Codebook in G.729a speech frames. Building a mathematical model to compute the embedding capacity for streaming media based stegosystems is the subject of future work.

Acknowledgements

This work was supported in part by grants from the National High Technology Research and Development Program of China (863 Program, No. 2006AA01Z444), the National Foundation Theory Research of China (973 Program, No. 2007CB310806), the National Natural Science Foundation of China (No. 60703053, and No. 60773140), and British Government (no. ktp006367). The authors would like to thank anonymous reviewers for their valuable suggestions.

REFERENCES

[1] C.E. Shannon, “Communications theory of secrecy systems,” *Bell System Technical Journal*, vol. 28, 1954, pp. 656-715.

[2] J. Zollner, H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke, and G. Wolf, “Modelling the security of steganographic systems,” in *Information Hiding, 2nd International Workshop*, vol. 1525 of Lecture Notes in Computer Science, Springer, 1998, pp. 344-354.

[3] R.J. Anderson, and F.A.P. Petitcolas, “On the limits of steganography,” *IEEE J. of Selected Areas in Communications*, vol. 16, no. 4, 1998, pp. 474-481.

[4] I.J. Cox, T. Kalker, G. Pakura, and M. Scheel, “Information Transmission and Steganography,” in *IWDW 2005*, vol. 3710 of LNCS, 2005, pp. 15-29.

[5] P. Sallee, “Model based steganography,” in *Int. Workshop on Digital Watermarking*, vol. 2939 of Lecture Notes in Computer Science, Springer, 2004, pp. 154-167.

[6] C. Cachin, “An information-theoretic model for steganography,” in *Information Hiding, 2nd International Workshop*, vol. 1525 of LNCS, Springer, 1998, pp. 306-318.

[7] F.A.P. Peticolas, R.J. Anderson, and M.G. Kuhn, “Information Hiding - A Survey,” *IEEE Trans. Proc. Thy.*, vol. 87, no. 7, July 1999, pp. 1062-1078.

[8] C. Grothoff, K. Grothoff, L. Alkhutova, R. Stutsman, and M.J. Atallah, “Translation-based steganography,” in *Proceedings of Information Hiding Workshop*, Springer-Verlag, 2005, pp. 213-233.

[9] P. Bao, and X. Ma, “MP3-resistant music steganography based on dynamic range transform,” in *IEEE International Symposium on Intelligent Signal Processing and Communication Systems* (ISPACS 2004), Nov. 2004, pp. 266-271.

[10] EasyBMP [http://easybmp.sourceforge.net/steganography.html].

[11] Y. Hu, and B. Yang, “Random-time-launching information hiding model,” *Journal of XIDIAN University*, vol. 28, no. 3, Jun. 2001, pp. 76-82.

[12] H. Farid, “Detecting hidden messages using higher-order statistical models,” in *Proceedings of IEEE International Conference on Image Processing*, Rochester, NY, 2002, pp. 34-42.

[13] J.A. O’Sullivan, and P. Moulin, “Some properties of optimal information hiding and information attacks,” in *Proc. 39th Allerton Conf.*, Monticello, IL, Oct. 2001, pp. 92-101.

[14] C. Bao, Y. Huang, and C. Zhu, “Steganalysis of Compressed Speech,” in *CESA'2006 Multiconference*, Beijing, 2006, pp. 104-113.

[15] B. Chen, and G.W. Wornell, “Quantization index modulation methods: A class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. Inform. Theory*, vol. 47, May 2001, pp. 1423-1443.

[16] R. J. Barron, B. Chen, and G.W. Wornell, “The duality between information embedding and source coding with side information and some applications,” in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001, pp. 300-305.