Security of streaming media communications with logistic map and self-adaptive detection-based steganography

TDSC-2018-11-0493.R2

Security of Streaming Media Communications with Logistic Map and Self-Adaptive Detection-Based Steganography

Jinghui Peng, Yijing Jiang, Shanyu Tang, and Farid Meziane, *Fellow, BCS*

Abstract—Voice over IP (VoIP) is finding its way into several applications, but its security concerns still remain. This paper shows how a new self-adaptive steganographic method can ensure the security of covert VoIP communications over the Internet. In this study an Active Voice Period Detection algorithm is devised for PCM codec to detect whether a VoIP packet carries active or inactive voice data, and the data embedding location in a VoIP stream is chosen randomly according to random sequences generated from a logistic chaotic map. The initial parameters of the chaotic map and the selection of where to embed the message are negotiated between the communicating parties. Steganography experiments on active and inactive voice periods were carried out using a VoIP communications system. Performance evaluation and security analysis indicates that the proposed VoIP steganographic scheme can withstand statistical detection, and achieve secure real-time covert communications with high speech quality and negligible signal distortion.

Index Terms— Security, VoIP, streaming communications, steganography

1 Introduction

With the development of the Internet, text messaging alone is hard to meet people's demands for multimedia communications. Internet users need more direct and vivid modern ways of communication, such as audio or video communications. Voice over Internet Protocol (VoIP) is one of the most popular audio communications services on the Internet. VoIP is finding its way into several applications, and it is expected to become a service like

• *(Corresponding author: Shanyu Tang.)*

electricity or water.

The Internet enables VoIP to provide reliable, global, low-cost and/or even free services, so many users communicate with each other daily using VoIP products, leading to increasing traffic of VoIP streams transmitted over the Internet. Due to the highly redundant representation in VoIP streams, VoIP is considered to be a dynamic cover object for steganography compared with static cover objects such as text, image and audio files [1-2]. As an interesting subject in the field of information security, steganography or covert communication (channel) works by hiding messages in inconspicuous cover objects (e.g. VoIP streams) that are then sent to the intended recipient [1]. Steganography can provide an additional layer of security in addition to encryption by embedding the encrypted message into steganographic carriers, which helps individuals or organisations protect sensitive information. For example, a message can be steganographically embedded into the least significant bits of frames on a CD. Covert steganographic channels can be used to bypass the censorship in a hostile environment. The covert channel can also be used by the adversary as a possible means of information exchange. A message can be concealed before distribution by splicing it to the end of a copy of a normal audio or video. A disgruntled employee may use steganography to ship out the most commercially sensitive information.

VoIP provides real-time audio communication services over the Internet, and VoIP packets are discarded immediately on arrival. That means that attackers do not normally have sufficient time to detect whether VoIP dynamic streams contain the hidden message or not. The real-time character of VoIP is useful in protecting the message hidden in their streams; however, the real-time requirements make it hard to perform necessary operations to embed the message into the streams without causing signal distortion.

VoIP communications consist of two phases: signalling phase and conversation phase. The signalling phase sets up and negotiates VoIP session parameters between the communicating parties. The most popular signalling protocol is called Session Initiation Protocol (SIP). As mutual authentication is a cryptographic scheme used to convince parties of each other's identity and to exchange session keys, it is typically used only when an extra level of security is needed, especially in VoIP communications [3]. Some key agreement protocols and authentication schemes [4-8] were proposed to improve the VoIP security in the signalling phase, but research on the protection of the VoIP conversation phase falls behind.

Security measures like the Triple Data Encryption Algorithm (3DES) provide protection for radio-frequency-identification communications [9], but their complexity, time consuming and increasing computational power make them unable to provide security for real-time VoIP communications. Thus, an alternative measure like steganography is sought for VoIP communications. Traditional Least Significant Bit (LSB)-based algorithms used in image steganography are prone to statistical analysis. Directly applying LSB to VoIP steganography certainly inherits its vulnerability, but the highly redundant representation in VoIP packets allows subtle modifications that preserve the perceptual content of the underlying packets. The use of a chaotic map in the modification process could strengthen

VoIP steganography due to its added complexity, which is a motivation of this study.

Pulse Code Modulation (PCM) codec is the most basic speech codec that exists in a large majority of speech codecs. So VoIP steganography with PCM was investigated in this study. And a new self-adaptive steganographic method based on a logistic chaotic map, taking into account the voice character of VoIP, was devised to realise real-time covert VoIP communications. The technical merits of this study are summarised as follows:

a) Active voice period detection-based secure, self-adaptive and real-time covert VoIP communications over networks using a logistic chaotic map;

b) Performance evaluation with state-of-the-art network equipment Digital Speech Level Analyser, unlike previous works with performance evaluation being conducted using in-house software with low precision;

c) Security tests carried out using the Mann-Whitney-Wilcoxon method, instead of conventional statistical tests.

The remaining of the paper is organized as follows. In Section 2, the related work is briefly introduced. Section 3 describes in detail our proposed real-time covert VoIP communications scheme based on self-adaptive audio steganography. The experimental setup is given in Section 4. Experimental results, security analysis and performance comparisons are discussed in Section 5. The final section concludes the paper.

2 RELATED WORK

Embedding a message into the payload of VoIP streams is one of the most widely researched steganographic methods. Some related works are introduced below.

Attempts have been made to improve the security of VoIP communications. As the LSB method is one of the most popular data embedding methods due to its low complexity and high capacity, LSB-based embedding techniques have been applied to VoIP communications. Kratzer et al. [10] reported a design of VoIP steganography, which substituted the bitstream of a secret message for the least significant bits of cover audio. Wang et al. [11] proposed a method of using the LSBs of voice samples to carry secret communications, and described a design of real-time speech hiding with G.711 codec, which was implemented in Linphone. They compressed the secret speech with Speex, before embedding it into the LSBs of voice samples. In these studies, the bitstream of the secret message to be hidden was uniformly distributed in cover objects using LSB substitution, which is most likely to be detected by statistical analysis. Besides, their works mainly focused on the designs of steganographic algorithms, neglecting the effects of the characteristic of VoIP conversation on VoIP steganography.

Huang et al. [12] suggested an algorithm for embedding data in some parameters of inactive speech frames encoded by G.723.1 codec, which was a high-capacity steganographic method. In addition, Huang et al. [13] proposed an algorithm for steganography in low bit-rate VoIP audio streams by integrating data hiding into the process of speech encoding.

Tian et al. [14] designed an M-sequence-based LSB steganographic algorithm for embedding information in VoIP

streams encoded by G.729a codec. Tian et al. [15] also proposed an adaptive partial-matching steganographic method with triple M sequences, which used a partial similarity value to evaluate the partial matching between the cover object and the secret data. They introduced three sequences: the first was used to eliminate the correlation between the secret data and the cover object; the second was utilised to guide an adaptive embedding process; the last was used for encrypting synchronization signalling patterns.

Aoki [16] proposed a lossless steganographic approach for u-law of G.711 codec, which embedded a secret message into '0' speech samples by exploiting the characteristic that a '0' speech sample could be represented by two codes '+0' and '-0'. If '0' was required to be embedded into a '0' speech sample, the sign of the speech sample was modified to '-'. Its steganographic capacity depended on the number of '0' speech samples, so its applicability is limited.

In 2017, Tian et al. put forward a bitrate modulating based steganographic algorithm with Hamming matrix encoding [17], but its practicality needs further study.

Balasubramaniyan et al. developed a PinDr0p mechanism to detect and measure single-ended VoIP audio features to identify all the applied voice codecs, and calculate packet loss and noise profiles with over 90% accuracy [18].

Peeters et al. [19] presented a Sonar system that detected the presence of SS7 redirection attacks by securely measuring audio round-trip times between telephony devices, capable of detecting 70.9% of redirected calls between call endpoints of varying attacker proximity (300 - 7100 miles) with low false positive rates (0.3%).

More recently, Jiang et al. designed a reversible data hiding in encrypted domain scheme with low computational complexity for three-dimensional meshes [20]. Zhang et al. suggested a coverless steganographic algorithm based on discrete cosine transform and latent dirichlet allocation topic classification, having robustness against common image processing and better ability to resist steganalysis [21].

Some researchers have engaged in VoIP steganalysis. Steganalysis is the science of detecting the message hidden using steganography, which is to distinguish the stego data from the cover object. Huang et al. [22] proposed a steganalysis method to detect covert VoIP communications, which used a sliding window mechanism and an improved regular singular (RS) algorithm. Huang et al. [23] suggested a steganalysis method that employed the second detection and regression analysis, which not only detected the hidden message in the compressed VoIP speech, but also accurately estimated the data embedding length. However, the successfulness of VoIP steganalysis depends heavily on the steganographic algorithm used in covert VoIP communications.

In summary, a great deal of research has been conducted on the basic techniques of VoIP steganography and steganalysis, but few studies have been carried out to discover the most appropriate data embedding locations in VoIP streams affected by the voice character during active and inactive speech periods. To bridge the knowledge gap, this study presents a new self-adaptive steganographic method using a specially designed Active Voice Period Detection algorithm and a logistic chaotic map, which can achieve real-time self-adaptive covert VoIP communications with negligible signal distortion.

3 PROPOSED REAL-TIME COVERT VOIP COMMUNICATIONS SCHEME

The proposed self-adaptive covert VoIP communications scheme is based on the connectionless User Datagram Protocol (UDP), which focuses on low-overhead operation and reduced latency to meet the real-time requirements of VoIP communications. The covert VoIP communications are realised by embedding secret data into audio signals encoded by PCM codec. The audio signals are chosen by an Active Voice Period Detection (AVPD) algorithm, which decides whether a VoIP packet carries active voice data (Active speech period) or inactive voice data (Inactive speech period). The bitstream of secret data is not uniformly embedded into the audio signals, but is distributed randomly using random sequences generated from a logistic chaotic map.

The initial parameters of the chaotic map can be exchanged between the communicating parties in three ways. Firstly, a device (e.g. smartcard) is used to hold the cryptographic parameters. Having a smart card on the top of VoIP is acceptable for a local VoIP network with few users, but is unlikely to implement in practice due to a massive number of VoIP users in real VoIP communication. Secondly, the parameters can be embedded into VoIP protocols such as the SIP signalling protocol. However, this method would affect and disrupt the functioning of the SIP signalling protocol that establishes connections between the users in the VoIP signalling phase. Thirdly, a key-distribution scheme is utilised to exchange the parameters, which is adopted in this scheme to distribute the initial parameters in the VoIP conversation phase after a connection is established via SIP.

In this study, the use of a key-distribution scheme to exchange the initial parameters of the chaotic map between the VoIP users is to avoid affecting and disrupting the functioning of the SIP signalling protocol that is used to establish connections between the communicating parties in the signalling phase in real-time VoIP communication. So the proposed steganographic channel is integrated with VoIP communication without causing perceptible signal distortion.

Before data embedding, the initial parameters of the chaotic map are exchanged between the communicating parties using the Diffie–Hellman key exchange scheme [24], which is indeed widely used, in the VoIP conversation phase. As Diffie–Hellman provides no authentication of the two communicating partners, this vulnerability needs to be overcome with the use of digital signatures and public-key certificates. Thus, the communicating parties in the proposed covert VoIP system are authenticated by means of elliptic curve digital signatures. The elliptic curve digital signatures are executed in a short space of time (~190 ms) [7], which is shorter than the acceptable latency of 400 ms for one-way VoIP communication, so they do not affect the real-time performance of the VoIP communication system. The extraction of the secret data hidden in the audio signals is carried out on the receiver side, which is a reverse process of data embedding.

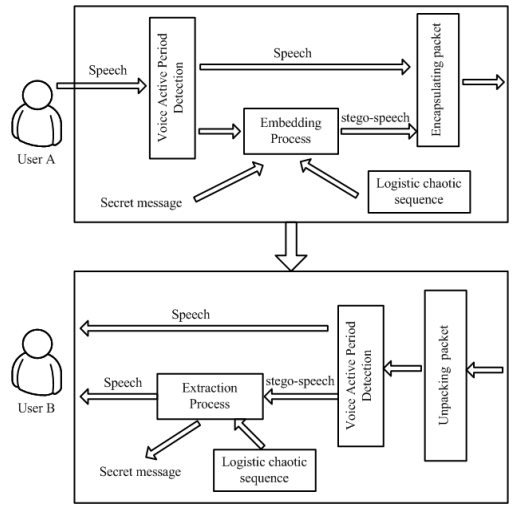


Fig. 1. Sketch of the proposed covert VoIP communications.

Figure 1 shows one way communication of the proposed covert VoIP communications. A speech stream is first detected by the AVPD function to decide whether it is in the active speech period or in the inactive speech period. A logistic chaotic map is used to generate random sequences that choose the embedding locations in the speech stream for the bitstream of secret data to be hidden. The chaotic map-based data embedding algorithm has great complexity, compared to simple LSB replacement with low complexity in literature. The speech stream with the hidden secret data is then encapsulated into packets and sent to the receiver. Otherwise, the speech stream is directly packed into packets and sent to the receiver. On the receiver side, after the arriving packets are unpacked, the speech stream is detected by AVPD until the secret data is extracted. Finally, the speech stream goes through the extraction process or is played back for listening.

The meanings of the symbols, used in the self-adaptive covert VoIP communications scheme, are provided in Table 1.

Table 1 Table of notations

Symbol	Meaning
a	Parameter of Tent map
C	Set of cover objects
c	Cover object
E	Mean of variance
H	Hypothesis value
j	Current sample
M	Secret data set
m	Secret data
N	Number of samples in the speech stream
$P\text{-}value$	Probability
Pc	Probability distribution

p	Largest interval
Q	Space of possible measurements
R	Embedding location set
r	Embedding location
S	Least significant bit set
s	Least significant bit
S_n	Sampling distribution
W	Space of possible measurements
x	Random number
x_0	Initial ratio of the population to the maximum population
x_n	Value of x_0 after n iterations
z^*	Test statistic
μ	Positive number
σ	Square root of variance

3.1 Active Voice Period Detection

There is no fixed voice activity detection module in PCM codec, so an AVPD algorithm is devised to minimise the impact of data embedding on speech quality. The AVPD algorithm uses a threshold to decide whether the speech in an audio packet has fallen into an active or inactive speech period. Although sometimes active and inactive speech periods are not distinguishable accurately, the AVPD can distinguish the majority of active and inactive speech periods in an audio packet.

The threshold for an active voice is determined after analysing audio data in the first few packets, in which only environmental noises exist in normal circumstances at the beginning of the conversation phase. Hence, the AVPD algorithm is effective in a real environment where users speak in a background noise.

Analogue to a delta-sigma modulator as a two-level dynamic quantizer to encode and decode signals [25], PCM codec uses a perception model-based compression method to code audio, which can yield high speech quality. To digitise an audio signal, the first step is sampling, and the nominal value recommended for the sampling rate is 8 kHz. The second step is quantization, i.e. transforming the sampling signal amplitude into a numeric value with binary digits. The sample signal values in inactive speech periods are smaller than those in active speech periods.

To decide whether the speech in an audio packet has fallen into active or inactive speech periods, a threshold value for the energy level of the signal is set to distinguish between active and inactive sample signal values. Normally, at the beginning of the conversation phase, there are few milliseconds of insignificant voice between the communicating parties. The first few packets are then analysed to obtain the threshold signal value which is defined as:

$$Threshold = Max + OffsetValue \quad (1)$$

where *Max* is the maximum value among all the sample signal values of the first few packets, and *OffsetValue* is an offset value which is less than *Max*. A large number of tests showed that the best range of *OffsetValue* is [0, threshold*2/3) with excellent AVPD results.

```
function VAPD: unsigned char *voiceData, int sampleNum
begin
        counter←0
        activeNumber←0
        while (counter < sampleNum)
        do
                if (abs(*voicedata)>Threshold)
                then activeNumber←activeNumber+1
                voicedata←voicedata+1
                counter←counter+1

        if(activeNumber > sampleNum/2)
        then return true
        else return false
end
```

Fig. 2. Pseudo code of the AVPD function for PCM codec.

If more than half of the sample signal values of an audio packet are larger than the *Threshold*, the speech in the audio packet is considered to be in an active speech period. Otherwise, it is in an inactive speech period. Figure 2 shows the pseudo code of the AVPD function in our covert VoIP communications system. If the AVPD function returns a true value, it means that the speech stream is active.

Figure 3 shows an example of a speech waveform where inactive and active speech periods are determined by AVPD. Clearly there are three active speech periods (marked with rectangles) in the figure. A sample signal value of an audio packet might be vulnerable to being modified through the network, but the active or inactive character of the packet is unlikely to be changed. Thus, it is more conducive to distinguish active and inactive speech periods before data embedding.

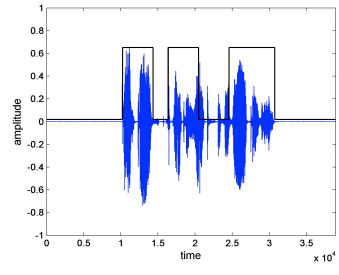


Fig. 3. Example of a speech waveform with active periods determined by AVPD.

3.2 Logistic Chaotic Map to Choose Embedding Locations

Chaos is a stochastic phenomenon of nonlinear deterministic system in the nature. The uncertainty of random sequence is resulted from the internal factor of a chaotic dynamical system. A chaotic map is extremely sensitive to the initial conditions and the parameters of the chaotic map. So a mass of noise-like but determinate random sequences can be obtained from a chaotic map. The random sequence can be reproduced from the same chaotic map with the

constant initial condition. In comparison with other pseudo random sequence generation algorithms such as Mersenne Twister method, using a logistic map to generate a random sequence is straightforward and convenient to implement.

In the proposed scheme, a series of random sequences generated from a logistic chaotic map are used to choose data embedding locations in VoIP streams. The utilization of chaotic map makes data embedding in VoIP streams randomly, and it is unlikely to predict the initial conditions of random sequences. So the properties of the chaotic map can increase the security of covert communications. To meet the real-time requirement, it needs to minimize the time to generate random sequences. Meanwhile, it is necessary to know the initial parameters with infinite precision for sensitivity of the chaotic map, so that the initial parameters can be transmitted securely between the communicating parties in the conversation phase.

A logistic map is one of the most popular models for discrete nonlinear dynamical systems. The map is popularized in a seminal paper by the biologist Robert May, in part as a discrete-time demographic model analogous to the logistic equation first created by Pierre Francois Verhulst [26]. A logistic map is given by

$$x_{n+1} = \mu x_n (1 - x_n) \qquad (2)$$

where x_0 is the initial ratio of the population to the maximum population at year 0, x_n denotes the value of x_0 after n iterations, a number between 0 and 1, and the ratio of existing population to the maximum possible population after n years, and μ is a positive number which stands for a combined rate for reproduction and starvation.

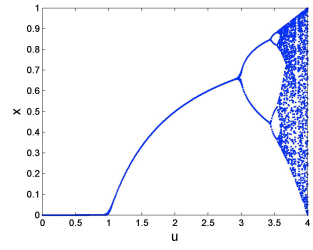


Fig. 4. Bifurcation diagram of a logistic map when $x_0 = 0.52$.

Figure 4 depicts a bifurcation diagram of a logistic map when $x_0 = 0.52$. As Fig. 4 shows, when $\mu \in (0,1]$, the value of x is equal or close to 0. When $\mu \in (1,3]$, x quickly approaches the value of μ-1/μ. It shows chaotic characteristics when μ varies in the range (3.57, 4]. When $\mu = 4$, x becomes increasingly chaotic. Figure 5 shows the ergodic property of chaos, when $x_0 = 0.52$ and $\mu = 4$ in the equation of a logistic map. As can be seen from Fig. 5, the value of x_n randomly falls in the range (0, 1] as n increases between (0,10000].

Tent map is also a discrete-time dynamical system model. The original formula for the Tent map can be written as:

$$\begin{cases} x_{n+1} = \mu x_n & 0 < x_n < \dfrac{1}{2} \\ x_{n+1} = \mu(1 - x_n) & \dfrac{1}{2} < x_n < 1 \end{cases} \qquad (3)$$

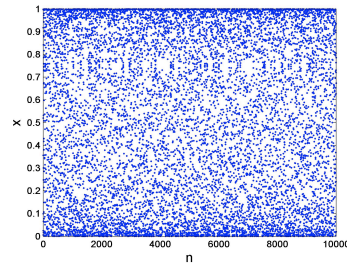


Fig. 5. Values of x_n with n increasing when $x_0 = 0.52$ and $\mu = 4$.

To extend the mapping range, the improved Tent map is obtained:

$$\begin{cases} x_{n+1} = \mu x_n & 0 < x_n < \dfrac{a}{2} \\ x_{n+1} = \mu(a - x_n) & \dfrac{a}{2} < x_n < a \end{cases} \tag{4}$$

In equation (4), $x \in (0,a)$, $\mu \in (0,2)$, and $a \in R$. For the Tent map, the parameters are $a = 16$, $\mu = 1.99$, and $x_0 = 0.552$. The parameter values of the logistic map are $x_0 = 0.52$ and $\mu = 4$. As Figs. 6 and 7 show, after 30000 iterations, the statistical distribution of the numbers generated from the logistic map is 'U' shape distributed. And the statistical distribution of the numbers from the improved Tent map is almost uniformly distributed.

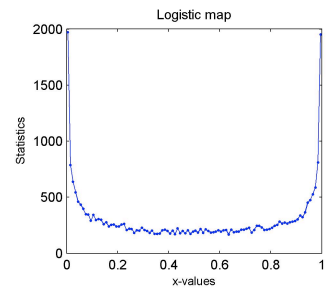


Fig. 6. Statistical distribution of numbers (Y axis) generated from the logistic map (X axis: x_n).

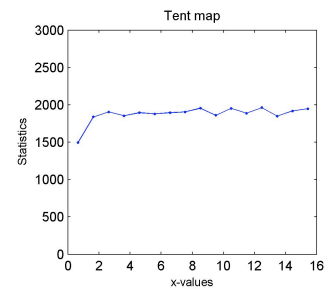


Fig. 7. Statistical distribution of numbers (Y axis) generated from the improved Tent map (X axis: x_n).

Since x is between 0 and 1, some adjustments need to be made to obtain a random sequence of integers from the

logistic map. Suppose x_0 and μ are given as certain values in Equation (2), a sequence of x can then be yielded and denoted by X as

$$X = \{x_i \mid i = 0,1,2,...,n\} \quad (5)$$

and $R = \{r_i = x_i \times 1000 \pmod{p} + 1 \mid i = 0,1,2,...,n\}$ (6)

where R is the embedding location set, r is the embedding location, and p is the largest interval. To get x = 16, the adjustment for the Tent map is $x_n = \text{floor}(x_n) + 1$.

As Fig. 8 shows, after adjustment the statistical distribution of numbers from the improved tent map is still uniformly distributed. As can been seen from Fig. 9, when x is in [1, 16] the numbers generated are around 2000, and the numbers from the logistic map are almost evenly distributed too.

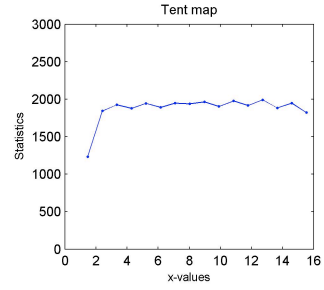


Fig. 8. Statistical distribution of numbers (Y axis) generated from the improved tent map after adjustment (X axis: x_n).

The sequence of R is utilized in the proposed covert VoIP communications. p is an integer, which represents the largest interval, and the value of p is in the range of 2 to 35 [27]. The PESQ scores and SNR values are stable before the embedding interval reaches 35. r_i is used to determine the data embedding locations in VoIP streams to embed the bitstream of secret data.

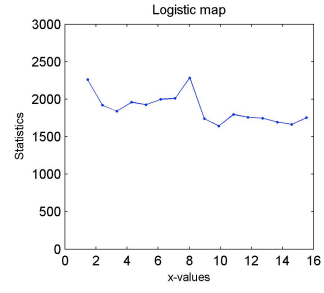


Fig. 9. Statistical distribution of numbers (Y axis) generated from the improved logistic map after adjustment (X axis: x_n).

3.3 Data Embedding Algorithm

A new data embedding algorithm is used to embed secret data in VoIP cover-speech streams. It is based on bit substitution with payloads using a random sequence to determine the data embedding location, i.e. where to embed the bitstream of secret data. At the beginning, there are two choices: either embedding secret data into an active or inactive speech period of VoIP streams. If data embedding in an active speech period is chosen, the embedding algorithm waits until the active speech period starts. An important parameter *ActiveChoice* is set to represent the choice. If the value of *ActiveChoice* is 'false', secret data is embedded in inactive speech periods, and the algorithm continues with inactive speech periods. If the value of *ActiveChoice* is set to 'true', the bitstream of secret data is embedded in active speech periods.

Assuming user A wants to send L bits of secret data M to user B, the secret data is described as $M = \{m_i \mid i = 0,1,2,...,L\text{-}1\}$. If the number of samples in the piece of speech of each packet is N, the least significant bit set of samples can be denoted by $S = \{s_i \mid i = 0,1,2,..., N\text{-}1\}$. The data embedding algorithm is designed as follows:

Step 1: Execute the function of AVPD to detect whether a VoIP packet carries active or inactive voice data, if the returned value of AVPD is equal to *ActivChoice*, go to Step 2; otherwise, go to Step 4. The initial value of j is 0, and s_0 is the least significant bit of the first sample in the chosen piece of speech.

Step 2: Obtain a random x_i from a logistic chaotic map, calculate $r_i = x_i \times 1000 \pmod{p} + 1$ as the data embedding interval to decide the data embedding location in VoIP streams.

Step 3: Suppose the least significant bit of the current sample is s_j, if $j+r_i < N$, replace s_{j+r_i} with m_i, then perform $j = j+r_i$, $i = i+1$, repeat Steps 2 and 3 until the end of the current piece of speech S, i.e., $j+r_i >= N$.

Step 4: Encapsulate the piece of speech in a packet to be sent; go to the next piece of speech, and repeat Step 1 until the secret data is embedded completely.

3.4 Data Extraction Algorithm

In the conversation phase, the receiver obtains the initial parameters of a logistic map, which is used to generate random sequences that randomly choose data embedding locations, and the value of *ActiveChoice* which determines where to extract the secret data hidden in VoIP streams. The same initial parameters and the value of *ActiveChoice* enable the receiver to successfully retrieve the secret data. The extraction process is a reverse phase of the data embedding process. After receiving an audio packet, the least significant bit set of samples can be denoted by $S' = \{s'_i \mid i = 0, 1, 2, ..., N\text{-}1\}$. The logistic chaotic map generates a corresponding random number x_i that decides the extraction location. The following steps are performed to extract the original secret data on the receiver side.

Step 1: Execute the AVPD function to detect whether a VoIP packet carries active or inactive voice data; if the returned value of AVPD is equal to *ActiveChoice*, go to Step 2; otherwise, go to Step 4. The initial value of j is 0, and s'_0 is

the least significant bit of the first sample in the chosen piece of speech.

Step 2: Generate a random number x_i from a logistic chaotic map, and calculate r_i to decide the extraction location.

Step 3: Suppose the least significant bit of the current sample is s'_j, if $j+r_i < N$, get s'_{j+r_i} as m_i, then perform $j = j+r_i$, $i = i+1$, repeat Step 2 and Step 3 until the end of the current piece of speech S', i.e., $j+r_i >= N$.

Step 4: Play audio, receive the next audio packet, and repeat Step 1 until the completion of extracting the secret data M.

4 Experimental Setup

4.1 Covert VoIP Communications System

To evaluate the performance of the proposed steganographic algorithm, speech samples coded by PCM were employed as the cover-speech. The steganographic algorithm was used in our VoIP communications system called StegPhone. In the StegPhone system, end-user terminals (the communicating parties) are connected to a VoIP proxy server through an IP local network. Its end-user interface is shown in Fig. 10.

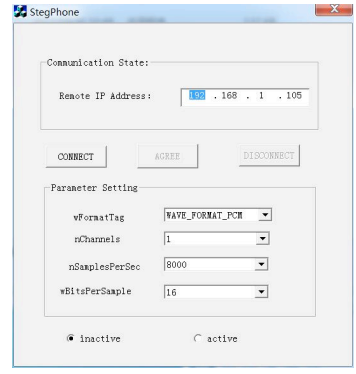


Fig.10. End-user interface.

StegPhone was developed in-house using C++ and MFC. The implement of speech signal acquisition and playback is based on winmm.lib which is a multimedia API, and the real-time transmission of audio packets is carried out by means of jrtplib 3.9.1 library. The transport protocol is UDP.

As Fig. 11 shows, the StegPhone VoIP system used the SIP signalling protocol to establish connections between two communicating parties, the media protocol to handle the real-time audio communication, and the IP protocol for VoIP voice transmission. These protocols were implemented on a VoIP server.

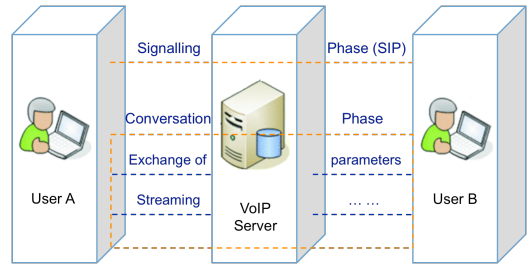


Fig. 11. StegPhone VoIP system.

The communicating party's IP address was needed to initialize VoIP communications, and the parameters used for sampling and quantizing the cover-speech were then selected. It was convenient to choose between data embedding in an active speech period and that in an inactive speech period.

The VoIP audio samples were obtained by using single-channel and sampling at 8 kHz. Each sample was represented with 16 bits. There were 2048 samples in each audio packet. The initial values of the logistic chaotic map were given as follows: $x_0 = 0.52$, $x_1 = 4x_0(1 - x_0)$, ..., $x_{n+1} = 4x_n(1 - x_n)$.

The value of x_i was used for determining the embedding location R in VoIP streams for data embedding. The value of p in equation (6) was determined to be 16 as the largest interval, and r_i was used for determining the embedding location in VoIP streams for data embedding.

In general, a key distribution occurs with short time-consuming and small communications traffic. The exchange of the chaotic initial condition is a kind of key distribution, and 8 bytes are sufficient for the value of x_0 and u to be transmitted. As the length of secret data in covert communications is time-dependent and there are more than 700 bytes secret message in our experiment, it is impracticable to envoy secret data using a key distribution method.

The experiments for testing the speech quality of the cover-speech and stego-speech samples were conducted using Digital Speech Level Analyser (DSLA). In the experiments, a player was used to playback records of English audio as the cover-speech to microphone. The audio samples were standard English records obtained from DSLA. The covert VoIP communications were achieved through a VoIP proxy server (a component of StegPhone) over our laboratory's local area network. Comparisons between cover-speech and stego-speech samples were carried out at the end of the VoIP call. On the receiver side, Perceptual Evaluation of Speech Quality (PESQ) scores and Signal-to-Noise Ratio (SNR) values of the speech samples were measured using DSLA II, which is a high-accuracy equipment made by Malden Electronics Ltd. in the United Kingdom.

4.2 Perceptual Evaluation of Speech Quality

According to ITU-T Recommendation, PESQ is an objective method for end-to-end speech quality assessment of

narrow-band telephone networks and speech codecs [28]. PESQ requires two inputs, the original unprocessed test signal and the degraded version that has been passed through the distorting system. Figure 12 illustrates the processing carried out by PESQ. To compare the signals, the reference speech signal and the degraded signal should be at the same, constant power level. The model begins by aligning the original and degraded signals to a standard listening level. They are then filtered with an input filter to model the standard telephone handset. The system under test may include a delay, which may be variable. To compare the original and degraded signals, they need to be lined up with each other. The signals are aligned in time and then processed through an auditory transform that mimics certain key properties of human hearing and the auditory transform gives a representation in time and frequency of the perceived loudness of the signal, known as the sensation surface. The difference between the sensation surfaces for the signals is known as the error surface. Two error parameters are extracted from the disturbance, are aggregated in frequency and time, and are mapped to a prediction of subjective mean opinion score (MOS).

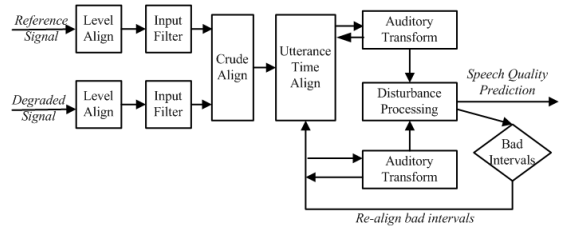


Fig. 12. Signal processing in PESQ.

The aim of the amended recommendation ITU-T P.862.1 is to provide a single mapping from the raw P.862 score to the Listening Quality Objective Mean Opinion Score (MOS-LQO). The mapping from PESQ score to PESQ P.862.1 is given by:

$$PESQ\text{P.862.1} = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.4945 \times PESQScore + 4.6607}} \qquad (7)$$

4.3 Signal-to-Noise Ratio

SNR is a measure that compares the level of a desired signal to the level of a background noise. It is defined as the power ratio between the signal and the background noise, often expressed in decibels. As many signals have a very wide dynamic range, SNRs are often expressed using the logarithmic decibel scale. In decibels, SNR is defined as:

$$SNR_{dB} = 10\log_{10}(\frac{P_{signal}}{P_{noise}}) \qquad (8)$$

where P_{signal} and P_{noise} are the levels of the signal and the noise in decibels, respectively. SNR is usually taken to indicate an average signal-to-noise ratio, as it is possible that instantaneous signal-to-noise ratios are considerably different.

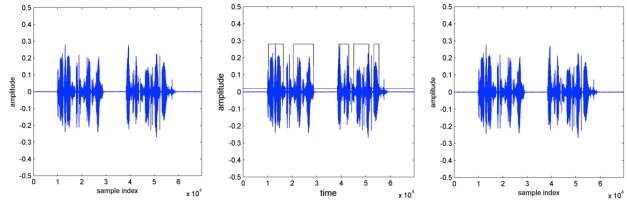
5 Results and Discussion

To compare data embedding in active and inactive speech periods, two sets of tests upon data embedding in active and inactive speech samples were carried out, respectively. Eighty female and male speech samples were used as cover objects in the experiments. Each experiment was repeated 12 times. About 8.7 seconds of audio in a stego-speech sample were taken for each test. To evaluate the performance of the proposed steganographic algorithm, the corresponding 8.7 seconds of the cover-speech were used as the reference. The 8.7 seconds of speech contain 69632 sampling points with 16 bits, i.e. 34 audio packets. The size of the hidden message in the speech stream is 745 bytes, equivalent to about 130 English words in a text file.

The sample signal values of normal VoIP speech are much larger than the threshold according to the definition of AVPD. The sample values after data embedding in VoIP packets vary in a range of -1 to 1, which means that it does not affect the active or inactive character of the VoIP packets. In addition, under the same initial parameters of the chaotic map, the same random sequences are generated and then used to determine the extraction locations for successfully retrieving the secret message on the receiver side. Thus, the rate of successfully extracting the secret message can reach 100% in a network without packet-loss. And all the experiments are on the basis of successfully extracting the hidden message from the received pieces of speech.

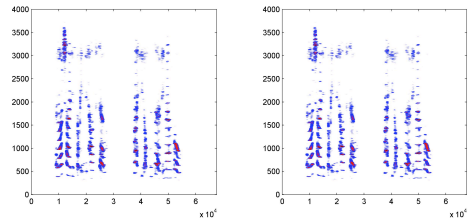
5.1 Active Speech Periods Using Logistic Maps

Figures 13(a), 13(b), 14(a) and 14(b) show the waveforms in the time-domain and the spectrums in the frequency-domain of female and male cover-speech and stego-speech samples with data embedding in active speech periods, respectively, and the AVPD detection results for the cover-speech samples. Close analysis of the figures reveals that there are almost no differences in the waveforms and spectrums between the cover-speech and stego-speech samples. That means that the proposed steganographic algorithm has no or little impact on the original cover-speech in the time and frequency domains.



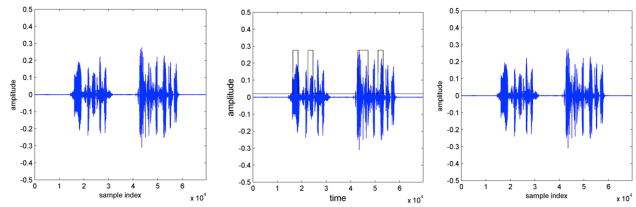
Female cover-speech Female speech AVPD Female stego-speech

Fig. 13(a). Comparisons in time-domain of female cover-speech and stego-speech with data embedding in active speech periods.



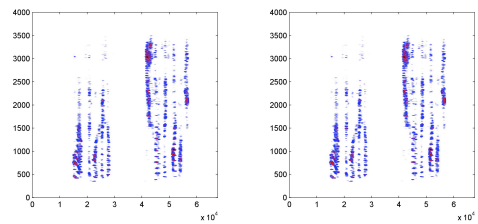
Female cover-speech Female stego-speech

Fig. 13(b). Comparisons in frequency-domain of female cover-speech and stego-speech with data embedding in active speech periods.



Male cover-speech Male speech AVPD Male stego-speech

Fig. 14(a). Comparisons in time-domain of male cover-speech and stego-speech with data embedding in active speech periods.



Male cover-speech Male stego-speech

Fig. 14(b). Comparisons in frequency-domain of male cover-speech and stego-speech with data embedding in active speech periods.

In PESQ measurements, the audio signal captured on the sender side was served as the reference signal for DSLA input, and two speech categories were measured as the degraded signals. One category is the stego-speech with the hidden message received on the receiver side, marked as 'stego'. The other category is the cover-speech without the hidden message received on the receiver side, denoted as 'original'.

Table 2 lists the PESQ P.862.1 scores of female and male cover-speech (original) and stego-speech (stego) samples with data embedding in active speech periods, respectively. The variances shown in the table are the differences in PESQ P.862.1 scores between the 'original' and 'stego' speech samples. As can be seen from Table 2, the impact that the local network alone made on the cover-speech was negligible. The small variances indicated that the proposed steganographic algorithm caused little degradation in speech quality, indicative of effective covert VoIP

communications.

Table 2 PESQ P.862.1 scores for female and male speech samples

Degraded signal		PESQ P.862.1			
		Max	Min	Mean	Variance
Female	Original	4.55	4.54	4.545	---
	Stego	4.21	4.19	4.200	0.345
Male	Original	4.55	4.54	4.545	---
	Stego	4.39	4.39	4.390	0.155

Table 3 SNR values (dB) for female and male speech samples

Degraded signal		Female		Male	
		Original	Stego	Original	Stego
Cover speech	Max	36.60	35.40	34.70	32.40
	Min	36.00	35.00	33.30	32.40
	Mean	36.30	35.20	34.00	32.40
Stego speech	Max	34.50	33.10	32.40	30.70
	Min	33.30	32.00	31.20	30.10
	Mean	33.90	32.55	31.80	30.4
Variance		2.40	2.65	2.20	2.00

Table 3 shows the SNR values and their variances of speech samples for active speech tests. The SNR values of the stego-speech samples were close to those of the cover-speech samples for female and male speech samples, respectively, which signified that the proposed steganographic algorithm had no or little impact on the quality of the cover-speech samples.

5.2 Inactive Speech Periods Using Logistic Maps

Table 4 lists the PESQ P.862.1 scores and their variances of female and male speech samples for inactive speech tests. Analysis of the PESQ variances shows that the proposed steganographic algorithm caused little degradation in speech quality, indicative of effective covert VoIP communications.

Table 4 Results for data embedding in inactive speech periods

Degraded signal		PESQ P.862.1			
		Max	Min	Mean	Variance
Female	Original	4.55	4.54	4.545	---
	Stego	4.02	4.02	4.020	0.525
Male	Original	4.55	4.54	4.545	---

	Stego	3.93	3.93	3.930	0.615

As Table 5 shows, for female or male speech samples, the SNR values of the stego-speech samples are slightly lower than those of the cover-speech samples, which meant that some distortions between the waveforms of 'stego' and 'original' speech samples existed. These results are evidenced by the amplitude increases in the waveforms in the time-domain.

Table 5 SNR results (dB) for data embedding in inactive speech periods

Degraded signal		Female		Male	
		Original	Stego	Original	Stego
Cover speech	Max	36.60	33.50	34.70	33.80
	Min	36.00	33.50	33.30	33.30
	Mean	36.30	33.50	34.00	33.55
Stego speech	Max	34.50	25.20	32.40	25.20
	Min	33.30	24.90	31.20	24.90
	Mean	33.90	25.05	31.80	25.05
Variance		2.40	8.45	2.20	8.50

5.3 Comparisons of Different Tests

Figure 15 shows the variances in the mean PESQ score and the variances in the SNR value between the cover-speech and stego-speech samples. The 'Active female' and 'Inactive female' represent the female speech tests with data embedding in active and inactive speech periods, respectively. The 'Active male' and 'Inactive male' stand for the experimental results for the male speech samples with data embedding in active and inactive speech periods, respectively.

As Fig. 15 shows, for both SNR scores illustrated in histogram and PESQ values plotted as the blue curve, the variances for data embedding in active speech periods are smaller than those for data embedding in inactive speech periods. Since the sample signal values of inactive speech are much smaller and close to 0 resulting in small amplitudes, data embedding in inactive speech periods leads to significant changes to the amplitudes. That means that inactive speech periods are more sensitive to changes in sample values than active speech periods. These results indicated that data embedding in active speech periods has much less impact on the speech quality of VoIP communications for both female and male speech samples.

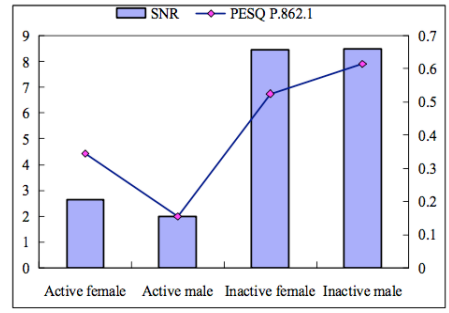


Fig. 15. Comparisons of variances for data embedding in active and inactive speech periods.

To compare the steganographic algorithm with Logistic map and the one with Tent map, the experiments with similar noise backgrounds were carried out. And the results are shown in Fig. 16. It is clear that, for both female and male speech samples, if data embedding in VoIP packets with inactive speech periods occurs, the variance of PESQ P.862.1 scores is estimated to be around 0.5; if data are embedded into VoIP packets with active speech periods, the variance of PESQ P.862.1 is about 0.1.

In addition, the changes in PESQ P.862.1 score between the steganographic algorithms with Logistic map and Tent map in the same tests were about 0.03, which means there is no impact on PESQ, indicating that the steganographic algorithms with Logistic map and Tent map work similarly.

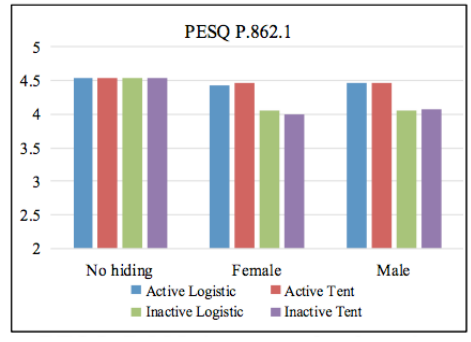


Fig. 16. Average PESQ P.862.1 scores for female and male speech samples.

5.4 Security Analysis

For covert communications, the essential security is that it would not cause any suspicion from attackers. Once secret communications are suspicious, or attackers have noticed that there is an underlying communications channel, the whole covert communications system is not safe, because attackers can intercept and even destroy the communications.

A. Adversarial model

Cachin's definition of steganographic security with a passive adversary is widely accepted in the literature [1, 30]. It assumes that the warden will permit Alice to send any cover object, c, to Bob, provided it is drawn from a probability distribution, P_C, and $P_C(c)$ is the probability of drawing a cover object from this distribution, where C is the set of all cover objects [29].

A measure of security for steganographic systems is the statistical distance (ε) between the cover and stego objects, $\varepsilon = \min_{Q_0 \subseteq C} \left| \sum_{c \in Q_0} P_C(c) - \sum_{c \notin Q_0} P_C(c) \right|$ and $Q_0 \subset Q$, where Q_0 is a plausible space and Q is the space of possible measurements.

If a stochastic process U is used to simulate the sending of the secret message in the ith packet, the total probability distribution of the message over Q can be expressed as

$$P_U(c) = \eta P_M(c) + (1-\eta) P_C(c) \tag{9}$$

where P_M is the probability distribution of the secret message, and η is the probability that '1' appears in a period. As

$$P_M(c) = \begin{cases} \frac{P_C(c)}{1+\varepsilon}, & c \in Q_0 \\ \frac{P_C(c)}{1-\varepsilon}, & c \in Q_1 \end{cases} \tag{10}$$

where Q_1 is the other plausible space of possible measurements, the relative entropy between the cover and stego objects of a steganographic system is given by

$$D(P_C \| P_U) \le \frac{1+\varepsilon}{2} \frac{\varepsilon^2 \eta^2}{1-\varepsilon^2 (1-\eta)^2} \tag{11}$$

For two N-level random sequences in covert VoIP communications, $\eta = 2^{n-1} 2^{n-1} / (2^n - 1)(2^n - 1)$, it has

$$D(P_C \| P_U) \le \varepsilon \tag{12}$$

Equation (12) shows that the proposed covert VoIP communications scheme is ε-secure against passive adversaries.

B. Adversary attacks

The steganographic security follows the same path as security in cryptography [1]. The security of covert steganographic communication lies in the fact that nobody has so far been able to produce an attack substantially faster than brute-force search for the key.

In the proposed steganographic algorithm, the Diffie–Hellman key exchange scheme [24] is used to securely exchange the initial parameters of the chaotic map between the communicating parties that are authenticated using elliptic curve digital signatures [7]. The use of authentication with the digital signatures can prevent man-in-the-middle attacks, which are particularly possible on wireless networks.

However, there are limitations on resisting tampering attacks. Mutual authentication could be used to prevent tampering attacks, and it is a good solution to resist tampering attacks by sending an authentication message of secret data to the receiver. If the verification fails on the receiver side, the secret data would be retransmitted. Although the utilization of Message Authentication Code (MAC) could resist the tampering attack, but the computation of MAC is

costing, which would add latency and lead to speech distortion. Besides, MAC is a kind of redundant message which would reduce the available embedding capacity. Thus, it is difficult to achieve security and efficiency simultaneously in real-time VoIP communications with steganography.

C. Statistical significance test

The Mann-Whitney-Wilcoxon (M-W-W) test was adopted to evaluate the security of the proposed steganographic algorithm. As a non-parametric significance test, the M-W_W can assess whether two independent samples of observations come from the same distribution [30]. Comparisons in probability distributions between the cover-speech and the stego-speech show whether the differences are almost indistinguishable.

When the sample sizes are sufficiently large, the M-W-W test is based on the standardized test statistic z^*:

$$z^* = \frac{S_2 - E\{S_2\}}{\sigma\{S_2\}} \tag{13}$$

where $E\{S_2\}$ and $\sigma\{S_2\}$ are the mean and square root of variance of the sampling distribution S_2 that is the combination of the two samples of observations to be assessed. If $|z^*| \le 1.960$, the null hypothesis is true ($H = 0$); If $|z^*| > 1.960$, the null hypothesis is false ($H = 1$).

In statistical significance testing, the P-value is the probability of obtaining a test statistic at least as extreme as the one that is actually observed, assuming that the null hypothesis is true. The null hypothesis is rejected when the *P*-value turns out to be less than a certain observed significance level, often 0.05 or 0.01. In the test, the significance level was set to be 0.05, and the number of samples in cover-speech and stego-speech were 69632.

Table 6 M-W-W test results for the Logistic map

Test	Rank sum	z^*	*P*-value	*H*
Active female	4.8493e+9	0.0883	0.9297	0
Active male	4.8490e+9	0.0532	0.9575	0
Inactive female	4.8506e+9	0.2626	0.7929	0
Inactive male	4.8525e+9	0.5131	0.6079	0

Table 6 contains the M-W-W test results and the parameters used for comparing the probability distribution drawn from the original VoIP streams (cover-speech) to that drawn from the stego VoIP streams with the hidden message (stego-speech). For the proposed steganographic algorithm, the maximum value for the test statistic (z^*) was 0.5131 for the 'Inactive male' test, given the two sample sizes being 69632 and 69632, respectively. Since $\text{Max}\{|z^*|\} = 0.5131$, i.e. $|z^*| < 1.960$, it concludes $H = 0$, which means that the probability distributions for the original VoIP streams and the stego VoIP streams did not differ, indicating that the proposed steganographic algorithm is undetectable in terms of statistical analysis.

As Table 6 shows, the *P*-values are considerably larger than the significance level 0.05 for data embedding in active

and inactive speech periods. Table 7 shows the values of H are 0 for all the tests with Tent map. There is almost no difference in the results between the steganographic algorithms with Logistic map and Tent map.

Table 7 M-W-W test results for the Tent map

Test	Rank sum	z^*	*P*-value	*H*
Active female	4.849e+9	-0.0226	0.9819	0
Active male	4.849e+9	0.0035	0.9972	0
Inactive female	4.849e+9	-0.2666	0.7898	0
Inactive male	4.849e+9	-0.3155	0.7524	0

Moreover, the values of H are 0 for all the tests, which indicates that the null hypothesis is true, i.e., the cover-speech and the stego-speech does not differ. That means that the proposed steganographic algorithm can withstand steganalysis based on statistical analysis.

5.5 Performance Comparisons between Algorithms

The proposed steganographic algorithm has an adaptive feature, and the embedding location is chosen randomly according to a random sequence. So the exact steganographic bandwidth could not be precisely calculated, but it can be determined in a range which is correlated with the value of p in equation (6). In our experiments, the value of p was 16, and the steganographic bandwidth was between 0.5 - 8 kbits/s in the selected audio signals.

To demonstrate the effectiveness of the proposed algorithm, performance comparison was conducted by comparing steganographic bandwidth, undetectability, and robustness, as shown in Table 8, thereby examining and noting the similarities or differences between the proposed algorithm and other related algorithms.

Table 8 Comparison of VoIP steganographic algorithms

	Steganographic bandwidth (kbits/s)	Undetectability	Robustness
Proposed algorithm	0.5 - 8	Yes	Yes
Wu [31]	20	Yes	n/a
Takahashi [32]	8	n/a	n/a
Liu [33]	0.2	Yes	n/a
Tian [34]	0.8 - 2.6	Yes	n/a
Xu [35]	0.1333	n/a	n/a
Miao [36]	7.5	Yes	n/a

As Table 8 shows, the proposed algorithm achieved a relatively larger steganographic bandwidth with a non-detectable characteristic.

6 Conclusion

A self-adaptive audio steganographic scheme for realising real-time covert VoIP communications over the Internet has been devised in this study. The new scheme was succesfully implemented in our StegPhone communications system. Two sets of tests (data embedding in active and inactive speech periods) were conducted for female and male VoIP speech samples, respectively. Although the PESQ and SNR variances of stego-speech samples for inactive speech periods are slightly larger than those for active speech periods, the M-W-W security analysis shows that the probability distributions drawn from the cover-speech and the stego-speech do not differ for both sets of tests, indicating that the proposed steganographic scheme is statistically undetectable with negligible signal distortion. Further studies are necessary to determine the effectiveness of the proposed scheme for other low bit-rate VoIP codecs.

References

[1] J. Fridrich, Steganography in digital media: principles, algorithms, and applications. Cambridge University Press, Apr. 2014.

[2] Z. Yang, X. Guo, Z. Chen, Y. Huang, RNN-Stega: Linguistic steganography based on recurrent neural networks, IEEE Transactions on Information Forensics and Security. 99 (11) (2018) 1-16.

[3] C. Yang, R. Wang, W. Liu, Secure authentication scheme for session initiation protocol, Computers & Security. 24 (5) (2005) 381-386.

[4] L. Wu, Y. Zhang, F. Wang, A new provably secure authentication and key agreement protocol for SIP using ECC, Computer Standards & Interfaces. 31(2) (2009) 286-291.

[5] J. Tang, Y. Cheng, Y. Hao, W. Song, SIP flooding attack detection with a multi-dimensional sketch design, IEEE Transactions on Dependable and Secure Computing. 11(6) (2014) 582-595.

[6] E.J. Yoon, K.Y. Yoo, C. Kim, Y.S. Hong, M. Jo, H.H. Chen, A secure and efficient SIP authentication scheme for converged VoIP networks, Computer Communications. 33 (14) (2010) 1674-1681.

[7] L. Zhang, S. Tang, J. Chen, S. Zhu, Two-factor remote authentication protocol with user anonymity based on elliptic curve cryptography, Wireless Personal Communications. 81 (1) (2015) 53-75.

[8] C. Wang, Y. Liu, A dependable privacy protection for end-to-end VoIP via Elliptic-Curve Diffie-Hellman and dynamic key changes, Journal of Network and Computer Applications. 34 (5) (2011) 1545-1556.

[9] D. Wang, J. Hu, H. Tan, A highly stable and reliable 13.56-MHz RFID tag IC for contactless payment, IEEE Transactions on Industrial Electronics. 62 (1) (2015) 545-554.

[10] C. Kratzer, J. Dittmann, T. Vogel, R. Hillert, Design and evaluation of steganography for Voice-over-IP, in: Proceedings of IEEE Int. Symp. on Circuits and Systems, Kos, Greece, 21-24 May 2006, pp. 2397-2340.

[11] C.Y. Wang, Q. Wu, Information hiding in real-time VoIP streams, in: Proceedings of IEEE Int. Symp. on Multimedia, Taichung, Taiwan, 10-12 Dec 2007, pp. 255-262.

[12] Y.F. Huang, S. Tang, J. Yuan, Steganography in inactive frames of VoIP streams encoded by source codec, IEEE Trans. Inf. Forensics Security. 6 (2) (2011) 296-306.

[13] Y. Huang, C. Liu, S. Tang, S. Bai, Steganography integration into a low-bit rate speech codec, IEEE Trans. Inf. Forensics Security. 7 (6) (2012) 1865-1875.

[14] H. Tian, K. Zhou, H. Jiang, J. Liu, Y. Huang, D. Feng, An M-sequence based steganography model for voice over IP, in: Proceedings of 44th IEEE International Conference on Communications, Dresden, Germany, 14-18 June 2009, pp. 1-5.

[15] H. Tian, H. Jiang, K. Zhou, D. Feng, Adaptive partial-matching steganography for voice over IP using triple M sequences, Computer Communications. 34 (18) (2011) 2236-2247.

[16] N. Aoki, A technique of lossless steganography for G.711 telephony speech, in: Proceedings of 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Harbin, China, 15-17 August 2008, pp. 608-611.

[17] H. Tian, J. Sun, C. C. Chang, J. Qin, Y. Chen, Hiding information into Voice-Over-IP streams using adaptive bitrate modulation, IEEE Communications Letters. 21 (4) (2017) 749-752.

[18] V. A. Balasubramaniyan, A. Poonawalla, M. Ahamad, M. T. Hunter, P. Traynor, PinDr0p: Using Single-Ended Audio Features To Determine Call Provenance, in: Proceedings of ACM Conference on Computer and Communications Security, Chicago, IL, USA, 2010, pp. 109-120.

[19] C. Peeters et al., Sonar: Detecting SS7 Redirection Attacks with Audio-Based Distance Bounding, in: Proceedings of 2018 IEEE Symposium on Security and Privacy, San Francisco, CA, 2018, pp. 567-582.

[20] R. Jiang, H. Zhou, W. Zhang, N. Yu, Reversible data hiding in encrypted three-dimensional mesh models, IEEE Transactions on Multimedia. 20 (1) (2018) 55-67.

[21] X. Zhang, F. Peng, M. Long, Robust coverless image steganography based on DCT and LDA topic classification, IEEE Transactions on Multimedia. 20 (12) (2018) 3223-3238.

[22] Y. Huang, S. Tang, Y. Zhang, Detection of covert voice-over Internet protocol communications using sliding window-based steganalysis, IET Communications. 5 (7) (2011) 929-936.

[23] Y. Huang, S. Tang, C. Bao, Y. J. Yip, Steganalysis of compressed speech to detect covert voice over Internet protocol channels, IET Information Security. 5 (1) (2011) 26-32.

[24] M. E. Hellman, An overview of public key cryptography, IEEE Communications Magazine. 40 (5) (2002) 42-49.

[25] D.J. Almakhles, A.K. Swain, N.D. Patel, Stability and performance analysis of bit-stream-based feedback control systems, IEEE Transactions on Industrial Electronics. 62 (7) (2015) 4319-4327.

[26] E.W. Weisstein, Logistic equation. MathWorld, 2003.

[27] S. Tang, Y. Jiang, L. Zhang, Z. Zhou, Audio steganography with AES for real-time covert voice over internet protocol communications, Science China Information Sciences. 57 (3) (2014) 1-14.

[28] Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. ITU-T Draft Recommendation P.862 (2000)

[29] I. Cox, M. Miller, J. Bloom, J. Fridrich, T. Kalker, Digital Watermarking and Steganography, 2nd edition. Morgan Kaufmann, 2008.

[30] J. Neter, W. Wasserman, G.A. Whitmore, Applied Statistics, 4th edition, Simon & Schuster, Inc., 1993.

[31] Z. Wu, W. Yang, G.711-based adaptive speech information hiding approach, in: Proceedings of International Conference on Intelligent Computing (ICIC 2006), Kunming, China, 16-19 August 2006, pp. 1139-1144.

[32] T. Takahashi, W. Lee, An assessment of VoIP covert channel threats, in: Proceedings of 3rd Int. Conf. Security and Privacy in Communication Networks 2007, Nice, France, 2007, pp. 371-380.

[33] L. Liu, M. Li, Q. Li, Y. Liang, Perceptually transparent information hiding in G.729 bitstream, in: Proceedings of 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Harbin, China, 2008, pp. 406-409.

[34] H. Tian, K. Zhou, Y. Huang, D. Feng, J. Liu, A covert communication model based on least significant bits steganography in voice over IP, in: Proceedings of 9th International Conference for Young Computer Scientists, China, 2008, pp. 647-652.

[35] T. Xu, Z. Yang, Simple and effective speech steganography in G.723.1 low-rate codec, in: Proceedings of Int. Conf. Wireless Communications & Signal Processing (WCSP 2009), Nanjing, China, 2009, pp. 1-4.

[36] R. Miao, Y. Huang, An approach of covert communications based on the adaptive steganography scheme on Voice over IP, in: Proceedings of IEEE Int. Conf. Communications (ICC 2011), Kyoto, 5-9 June 2011, pp. 1-5.