Covert Voice over Internet Protocol Communications with Packet Loss Based on Fractal Interpolation

YIJING JIANG and SHANYU TANG, China University of Geosciences, University of Salford
LIPING ZHANG and MUZHOU XIONG, China University of Geosciences
YAU JIM YIP, University of Salford

The last few years have witnessed an explosive growth in the research of information hiding in multimedia objects, but few studies have taken into account packet loss in multimedia networks. As one of the most popular real-time services in the Internet, Voice over Internet Protocol (VoIP) contributes to a large part of network traffic for its advantages of real time, high flow, and low cost. So packet loss is inevitable in multimedia networks and affects the performance of VoIP communications. In this study, a fractal-based VoIP steganographic approach was proposed to realize covert VoIP communications in the presence of packet loss. In the proposed scheme, secret data to be hidden were divided into blocks after being encrypted with the block cipher, and each block of the secret data was then embedded into VoIP streaming packets. The VoIP packets went through a packet-loss system based on Gilbert model which simulates a real network situation. And a prediction model based on fractal interpolation was built to decide whether a VoIP packet was suitable for data hiding. The experimental results indicated that the speech quality degradation increased with the escalating packet-loss level. The average variance of speech quality metrics (PESQ score) between the "no-embedding" speech samples and the "with-embedding" stego-speech samples was about 0.717, and the variances narrowed with the increasing packet-loss level. Both the average PESQ scores and the SNR values of stego-speech samples and the data-retrieving rates had almost the same varying trends when the packet-loss level increased, indicating that the success rate of the fractal prediction model played an important role in the performance of covert VoIP communications.

CCS Concepts: • **Information systems → Multimedia streaming**; • **Computer systems organization → Real-time systems**; • **Mathematics of computing → Interpolation**; • **Security and privacy → Distributed systems security**

Additional Key Words and Phrases: Covert communications, VoIP, fractal interpolation, steganography, packet loss

ACM Reference Format:
Yijing Jiang, Shanyu Tang, Liping Zhang, Muzhou Xiong, and Yau Jim Yip. 2016. Covert Voice over Internet Protocol communications with packet loss based on fractal interpolation. ACM Trans. Multimedia Comput. Commun. Appl. 12, 4, Article 54 (August 2016), 20 pages.
DOI: http://dx.doi.org/10.1145/2961053

This work is supported by the National Natural Science Foundation of China, under Grant 61272469 and Grant 61303237.
Authors' addresses: Y. Jiang, S. Tang (corresponding author, Senior Member, IEEE), L. Zhang, and M. Xiong, School of Computer Science, China University of Geosciences, Wuhan 430074, China; emails: yijingjiang2012@gmail.com, shanyu.tang@gmail.com, carolyn321@163.com, mzxiong@foxmail.com; Y. J. Yip, University of Salford, Salford, M5 4WT, United Kingdom; email: y.j.yip@salford.ac.uk.

DOI: http://dx.doi.org/10.1145/2961053

1. INTRODUCTION

Throughout human history, people have been constantly seeking for newer and more effective communication methods. Since the emergence of the World Wide Web in 1990s, the rapid development of information network technology especially the Internet has radically changed the way people communicate with each other. Digital data, including images, audio, video and other information could be spread to all the corners of the world. E-mail, video telephony, and video conference have become new forms of human communications in daily life. Internet provides low-cost and convenient means of access to massive data and communication methods. The Internet has already become an essential part in people's lives.

As a new communication method, Voice over Internet Protocol (VoIP) is flourishing in the field of electronic communications, because it has irreplaceable advantages in providing network multimedia services with low cost, high quality, multi-function and maintainability. With the increasing percentage of VoIP streams in the Internet traffic, VoIP is considered to be a better cover object for information hiding compared with other cover objects such as text files, image files, and audio files. Thus, VoIP steganography has caught attention of a growing number of researchers in the field of information hiding.

Steganography is actually an ancient art. Since human civilization, people had the thought of protecting secret information, and ancient people learned to pass a message with secret signal and code, which was an early form of information hiding. Information technology and the Internet provide modern digital steganography with an ample space for fast development. The modern research of covert communication was derived from the "The prisoners' problem" [Simmons 1984]. In the problem, Bob and Alice were jailed in separate cells and they were trying to prepare an escape plan, but their communication had to pass through and be inspected by the warden Wendy. If Wendy found any suspicious information transferring between Alice and Bob, she would defeat their plan. This problem triggered a boom in the modern research of information hiding.

Steganography is a method of embedding secret data into a cover object, which should not cause unacceptable distortion and arouse observers' attention. Both steganography and encryption technology provide the confidentiality of secret data, but there are significant differences in many aspects. Encryption technology only protects the content of secret data, making them unreadable; thus, unauthorized users can know the existence except the specific details about the secret data. Steganography hides the existence of secret data, such that unauthorized users know neither the existence of the secret data nor the details of them. However, steganography and cryptography are not mutually exclusive, and they could complement each other to improve the security of covert communications.

In comparison with the Public Switched Telephone Network (PSTN), a complex VoIP comes with more security threats, because voice packets are transmitted in such an open network environment. Meanwhile, VoIP communication is a real-time multimedia application, and it has a high quality of service requirements, e.g., the one-way delay, i.e.,, "mouth-to-ear", must be kept less than 400ms to provide acceptable speech quality [ITU-T G.114]. In order to protect the user privacy in VoIP systems, added encryption and other security measures could provide security, but they also inevitably cause some delay, which affects the performance of VoIP communication. To emphasize the problem mentioned above, the main objective of this study was to realize covert VoIP communication with acceptable delay in the presence of packet loss.

The main contribution of this study lies in utilizing a fractal-based packet loss prediction model to reduce the impact of packet loss on covert VoIP communications. And the proposed covert VoIP communication scheme could also be applied to other speech codecs, such as G.729, G.723.1, and so on. The packet loss prediction model could also

be implemented on covert communications over other media like video, to reduce the loss of secret data.

The rest of this article is organized as follows: In Section 2, the related work is briefly introduced. Section 3 describes the proposed covert VoIP communication scheme, consisting of the data embedding steganographic algorithm and the prediction model based on fractal interpolation. The experimental setup, results and discussion, and comparisons with other steganographic methods are detailed in Section 4. Finally, the conclusion and directions for future research are given in Section 5.

2. RELATED WORK

Early steganography is widely applied in a variety of digital media, such as text, image [Yang et al. 2008; Lee and Chen 2000; Marvel et al. 1999], audio [Darsana and Vijayan 2011; Cvejic and Seppänen 2002], and video files [Cetin et al. 2012]. As a result of the insensitivity of Human Visual System (HVS), human eyes cannot make a clear distinction between the original image and the image with a secret message embedded. It is generally recognized that VoIP steganography is more challenging than image steganography for the wider dynamic range of Human Auditory System (HAS) in comparison with HVS.

There have been some efforts to study VoIP steganography. Those steganographic techniques can be roughly divided into two categories. The first one is hiding secret data into protocol headers in TCP/IP protocols, such as the reserved field in protocol headers, and mostly in SIP or SDP protocol [Mazurczyk and Szczypiorski 2008a]. For example, the method was proposed in Huang et al. [2011a]. The second one is hiding secret data in the data block of voice packets, in which different techniques were used to embed secret data [Tang et al. 2014].

2.1. Steganographic Methods that Modify the Network Protocol

For the first category, the packets with data embedded into protocol headers are likely to be dropped by an intelligent router. However, those algorithms have not addressed the synchronization mechanism associated with VoIP communications.

Besides, a steganography method was developed in Mazurczyk and Szczypiorski [2008b], which formed a covert channel based on VoIP streams. The method had the characteristics of most steganographic techniques; it also presented two new ideas. The first one was named the network steganography scheme, which utilized the unused field in Internet protocols, such as UDP (User Datagram Protocol), RTP (Real-Time Transport Protocol), and RTCP (Real-Time Control Protocol) protocols. The second one was called Lost Audio Packets Steganography, which made use of the delayed audio packets to achieve a covert channel of mixed time storage. Their experimental results included only the capacity of the secret data transmitted during a typical VoIP session, excluding the consideration of steganalysis detection.

2.2. Steganographic Methods by Modifying the Voice Data

For the second category, hiding secret data into speech by modifying the voice payload is the most commonly used. For instance, the Least Significant Bit (LSB) substitution is the most basic method. And at the coding stage, secret data were hidden by utilizing the feature of speech codec. Moreover, the secret data were embedded into the speech streams by analyzing the characteristics of voice.

A great deal of research is based on the LSB method. LSB algorithm has been applied to many covert communications systems. The first VoIP steganographic technique was introduced by Aoki [2003]; it was the first use of G.711 codec for steganography to embed secret data into VoIP speech streams. Kratzer et al. [2006] proposed a covert communication model based on LSB VoIP steganography. This article indicated the differences between active voice and quiet voice. But this method simply replaced the

least significant bits of speech with the bitstream of secret data, and so it is easy to be detected by a simple statistical analysis. A design of real-time speech hiding for G.711 codec was suggested [Wang and Wu 2007]. The secret speech was compressed with Speex, before embedding it into the least significant bits of each two samples.

Some improved LSB algorithms for VoIP steganography were also developed. A LSB-based embedding algorithm was suggested [Huang et al. 2008], which guided the embedding process with a pseudo-random sequence. Miao and Huang [2011] analyzed the character of audio and proposed an adaptive steganography system that implemented different embedding approaches in active frames and inactive frames. In addition, they designed an overflow judgment to ensure the synchronization of the transmission of secret data.

A mechanism about the least significant bit based on G.711 speech codec was proposed [Wu and Yang 2006], which calculated the statistical characteristics of speech energy to estimate the hiding capacity of the cover speech used to hide secret data. Experiments showed that this method achieved a greater steganographic capacity than traditional LSB steganography methods and had less effect on the quality of cover speech. In the literature [Ito et al. 2010], researchers proposed an improved LSB algorithm based on tolerable distortion, which could improve quality of speech.

At the coding stage, secret data are hidden by utilizing the feature of speech codec. Chang and Yu [2002] proposed a steganography method that hides speech data in the MELP and G.729 encoded speech. Tian et al. [2008] suggested a VoIP covert communication model based on LSB steganography. It was an improved model based on the method in Kratzer et al. [2006]. Xu and Yang [2009] proposed a steganography algorithm based on G.723.1 codec with 5.3 Kbits/s coding rate, which achieved 133.3 bits/s steganographic bandwidth. A state-based steganography algorithm was proposed in Zhou et al. [2012], which was implemented on G.723.1 low bit rate speech codec. The steganography algorithm modified the G.723.1 transmission parameters to hide secret data. A lossless steganographic approach for u-law of G.711 codec [Aoki 2010] used the redundancy of G.711 codec to hide secret data without speech distortion. In Huang et al. [2011b], an algorithm for embedding data in some parameters of inactive speech frames encoded by G.723.1 codec was suggested, which is a high-capacity steganography method. In addition, in Huang et al. [2012], they also proposed an algorithm for steganography in low-bit-rate VoIP audio streams by integrating information hiding into the process of speech encoding. Tian et al. [2009] designed an M-sequence-based LSB steganographic algorithm for embedding information in VoIP streams encoded by G.729a codec. In Tian et al. [2011], they also proposed an adaptive partial-matching steganography method with triple M sequences, which used a partial similarity value to evaluate the partial matching between the cover object and secret data.

In addition, an effective steganography scheme for hiding secret speech in narrowband speech was proposed in the literature [Guerchi and Djebbar 2009]. The embedding process was designed in the high-frequency and low-amplitude part of speech, and the output was the stego-speech with secret speech, which had quality of speech similar to the original speech. The process occurred in the frequency domain [Rabie 2006; Guerchi et al. 2008], and the digital information was embedded in the amplitude component.

2.3. Summary

The above research work focuses on the steganography algorithm instead of the synchronization mechanism that is essential in real covert VoIP communication applications. Due to inevitable packet loss in the VoIP network, the design of the steganographic algorithm should take into account the impact of packet loss on the efficiency of steganography. When packet loss occurs, not only would the voice data in the audio

packets be lost, but also the secret data hidden in the voice data. However, few studies have taken into account of packet loss in VoIP communications. Taking into account the VoIP speech distortion caused by an inevitable packet loss and delay, Aoki [2003] proposed an error concealment method. The method improved the side information based on the reconstruction technique at the sending end, and the packets were fully compatible with traditional VoIP format. The method in Mazurczyk [2012] made use of packet loss and retransmission to hide secret data, although the voice packet with retransmission was not used by applications to play. Improved LACK steganographic method, utilizing intentionally excessively delayed packets to hide secret data without modifying speech data, was also proposed by Mazurczyk et al. [2014]. The hidden data insertion rate is time dependent between 0 and 10 Kbits/s. However, sometimes the excessively delayed packets may be discarded, so the packets could not arrive at the receiver.

To address the packet-loss problem with VoIP communications, a packet-loss fractal prediction model was developed in this study to reduce the impact of packet loss on covert VoIP communications. In consideration of the retrieving of secret data with packet-loss occurrence, the preprocessing of the secret data was conducted in the divided blocks, i.e., the secret data embedded in each packet was pre-processed independently. Since the retrieving of secret data between packets is not associated, the packet-loss events would have no effect on the retrieving of secret data in the packets that are not missing. There is a wide recognition that packet loss is out of control in the real network environment. So VoIP communications in our experiments were implemented in a local area network linked constantly to the Internet, the Gilbert packet-loss model was modified to simulate packet loss in a real network environment, and then it is convenient to control the packet loss rate to compare the experimental results.

3. PROPOSED STEGANOGRAPHY SCHEME FOR COVERT VOIP COMMUNICATIONS WITH PACKET LOSS

The proposed covert VoIP communication scheme was based on the connectionless UDP which emphasizes low-overhead operation and reduces latency to meet the real-time requirement, instead of TCP which is a reliable, ordered, and connection-oriented but time-consuming communication protocol. A covert VoIP communication was achieved by embedding secret data into an audio signal encoded by Pulse Code Modulation (PCM) codec. The audio signal was chosen by a packet-loss prediction model, which decides whether an audio packet would be discarded. The bitstreams of secret data were embedded into the chosen audio signal using a data embedding steganographic algorithm with variable embedding intervals. The extraction of secret data hidden in the audio packet was carried out at the receiving end, which was a reverse process of the data embedding phase. When the proposed scheme is implemented in a distributed environment, the network traffic statistic data are needed. The network traffic data can be collected from the statistic packets numbers that sent and lost in a typical VoIP communication. Then, the network traffic statistical data can be used in the packet loss prediction model of covert communications. And the Gilbert packet-loss model is executed in a real packet-loss network environment. The packet-loss prediction model is implemented at the sending end.

Figure 1 shows the framework of the proposed covert VoIP communication in which it was predicted whether each voice packet would be discarded or not. If a voice packet was predicted to be not missing, it would be used to embed secret data. Otherwise, the voice packet would keep its original speech and would not be used for hiding the secret data. Then, every packet was decided whether to be transmitted to the receiver or not. At the receiving end, once the arriving packets were unpacked, the retrieving operation was carried out until the completion of extracting the secret data. Finally, the speech went through the extraction process or was played back for listening. The details of

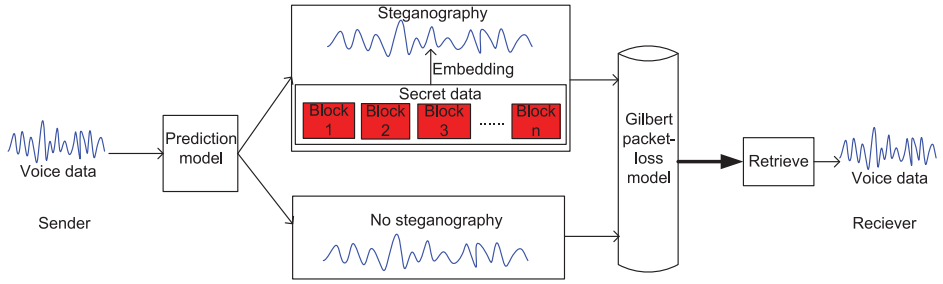


Fig. 1. The framework of the proposed covert VoIP communication scheme.

```
Begin
S=encrypt(M),
B=ConvertToBinary(S),
do
{
if (B[i]==0), V(k)=V(k)&0xfe,
if (B[i]==1), V(k)=V(k)|0x01,
k=k+R,
length=length-k,
}while(length!=0)
end
```

Fig. 2. The pseudo-code of the data embedding function used.

the data-embedding steganographic algorithm and the fractal prediction model are described below.

3.1. Data Embedding Steganographic Algorithm

The steganographic algorithm used for the proposed covert VoIP communication scheme is a data embedding algorithm with variable embedding intervals incorporating Advanced Encryption Standard (AES). The covert VoIP communication system was implemented by embedding a secret message encrypted with symmetric cryptography AES-128 into audio signals encoded by PCM codec. The secret message was encrypted and embedded in the payload of VoIP audio packets (Figure 2) before sending them to meet the real-time requirements of VoIP communications. At the receiver end, the corresponding algorithm was employed to retrieve the secret message, and then to decrypt it to get the original secret message.

In the data embedding function (Figure 2), M is the secret message to be hidden in an audio packet, and the length of M is an integral multiple of 16 bytes and smaller than the size of the audio packet; V is the voice in the payload of VoIP packet, and the length is the size of the audio in a packet.

3.2. Packet-Loss Fractal Prediction Model

In this study, a packet-loss fractal prediction model was built to reduce the impact of packet loss on covert VoIP communications by predicting the number of lost packets in next packet-loss event.

Fractal interpolation is a method to construct fractal curves, which was proposed by American mathematician M.F. Barnsley in 1986. The principle is to construct the corresponding Iterated Function System (IFS) according to a given set of interpolation points. The attractor of the IFS is the function graph through the given set of points.

Table I. Notations of Packet-Loss Events in a VoIP Communication

n	0	1	2	...	i	...	$N-1$	N
x_n	$x_0 = 0$	$x_1 = 1$	$x_2 = 2$	...	$x_i = \mathrm{i}$	...	$x_{N-1} = N-1$	$x_N = N$
y_n	y_0	y_1	y_2	...	y_i	...	y_{N-1}	y_N
x_i is the sequence number of the x_ith packet-loss event.								
y_i is the number of packets lost in the x_ith packet-loss event.								

When the fractal interpolation method was applied to VoIP communications, packet-loss events would be counted in VoIP communications, and the given set of points is the number of packets lost in the corresponding packet loss event. Table I shows N packet-loss events occurred, and the number of packets lost in each packet-loss event. And these data were then analyzed to predict the number of packets lost in the $(N+1)$th packet loss event using the packet-loss prediction model of VoIP communication.

Generally, the following affine transformation can be IFS to construct a fractal interpolation curve:

$$w_i \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_i & 0 \\ c_i & d_i \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e_i \\ f_i \end{bmatrix}, \tag{1}$$

where $i = 1, 2, \ldots, N$. (x, y) are the coordinates of a point, and w_i is the affine transformation relationship right of the equal, the parameters a_i, c_i, and d_i are the matrix elements, and e_i, f_i are the constant components after transformation. If the IFS attractor has to go through the given interpolation point (x_i, y_i), the transformation must meet the requirements of the following conditions,

$$w_i \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_{i-1} \\ y_{i-1} \end{bmatrix} \tag{2}$$

$$w_i \begin{bmatrix} x_N \\ y_N \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{3}$$

In other words, the left endpoint (x_0, y_0) of the whole range is mapped to the left endpoint $(x_{i\text{-}1}, y_{i\text{-}1})$ of the subinterval, and the right endpoint (x_N, y_N) of the whole range is mapped to the right endpoint (x_i, y_i) of the subinterval. In accordance with Equations (1), (2), and (3) and the given parameter d_i, it is easy to acquire the other parameters.

$$a_i = \frac{x_i - x_{i-1}}{x_N - x_0} \tag{4}$$

$$e_i = \frac{x_N x_{i-1} - x_0 x_i}{x_N - x_0} \tag{5}$$

$$c_i = \frac{y_i - y_{i-1}}{x_N - x_0} - d_i \frac{y_N - y_0}{x_N - x_0} \tag{6}$$

$$f_i = \frac{y_{i-1} x_N - y_i x_0}{x_N - x_0} - d_i \frac{x_N y_0 - x_0 y_N}{x_N - x_0}, \tag{7}$$

where a_i, c_i, e_i, f_i are the parameters of the affine transformation, which means the fractal feature parameters of the N packet-loss events. After obtaining the parameters, the IFS attractor can be determined. With the increasing number of times of iteration, the fitting degree of the interpolation curve continues to be improved. And it will be a stable and unchanging interpolation curve after lots of iterations, which is not only through the sampling points but also much closer to the original curve.

From the iterations with different initial points, it is easy to find that the affine transformation IFS obtained based on fractal interpolation has similar data with the historical data, and no matter what the initial point, it will gradually approach the IFS attractor after iterations.

According to the self-similarity and scale invariance of fractals, the extension portion near the interval can keep the fractal characteristic of the interval. When the point $P(x_p, y_p)$ falls on the extension portion of the curve, it also has the same fractal property. And when $x_p = N+1$, y_p is the number of packets lost in $(N+1)$th packet-loss event. It is much faster to iterate from the point $P(x_p, y_p)$ by using the affine transformation IFS for approaching the IFS attractor than to iterate from the point $Q(x_p, y_q)$ which has the same horizontal coordinate with P. The deviation between the attractor and the point set after iteration from point P is smaller than iteration from the point Q. In other words, the point with the same horizontal coordinate and the smallest deviation can be regarded as the predicted extrapolation point that the proposed model is looking for. Then y_{N+1} will be calculated, which is the number of packets lost in the $(N+1)$th packet-loss event predicted from the packet-loss prediction model.

Fractal geometry is actually a natural geometry. Fractal interpolation function fits the curves with strong volatility by using the self-similar structure of phenomena in the nature. And it has been proved to be a very effective tool. In Leland et al. [1994] and Willinger et al. [1997], it was proved that Ethernet LAN traffic is statistically self-similar by rigorous statistical analysis of hundreds of millions of high quality Ethernet traffic measurements collected between 1989 and 1992. Song et al. [2005, 2006] also reported that the self-similarity arose in different fields such as biology, technology, and sociology in a given length-scale, and networks showed a fractal topology. It has demonstrated that network traffic is self-similar or fractal. It is well known that burst packet loss rate exists in heavy network load, and packet loss scarcely occurs in low network load, indicating that packet loss is also of fractal. Since fractal prediction methods were well applied to short-term prediction [Xiu et al. 2014], in this study, it was a first attempt to apply fractal prediction into predicting packet loss in VoIP communications.

Based on the above thought, a fractal extrapolation method for VoIP communications was proposed in this study to predict the number of lost packets in next packet-loss event. In the proposed packet-loss prediction model, the horizontal coordinate is the sequence number of packet-loss event, and the vertical coordinate is the number of packets lost in the corresponding packet-loss event. The interval, which is also called the whole range, denotes all the packet-loss events that have been collected. And the points in the interval are the records of the number of packets lost in the collected packet-loss events. The point P with the horizontal coordinate x_p means the x_pth packet-loss event, and x_p is in the extension portion of the interval. The vertical coordinate y_p is the number of packets lost in the x_pth packet-loss event, which is the predicted value the model seeks.

The corresponding algorithm for the proposed packet-loss prediction model is designed as follows:

Step (1). Choose the sample points in interval which represents all the packet-loss events that have been collected in previous experiments. For each point, the horizontal coordinate is the sequence number of packet-loss event, and the vertical coordinate is the number of packets lost in the corresponding packet-loss event. Calculate the parameters of the affine transformation to obtain the IFS in the interval, and mark the maximum of all the vertical coordinates in the interval as $y_{\max}$ which is the largest number of packets lost in the previous experiment.

Step (2). According to the prediction requirement, determine the horizontal coordinate as x_p, which means the x_pth packet-loss event would be predicted. Initialize its

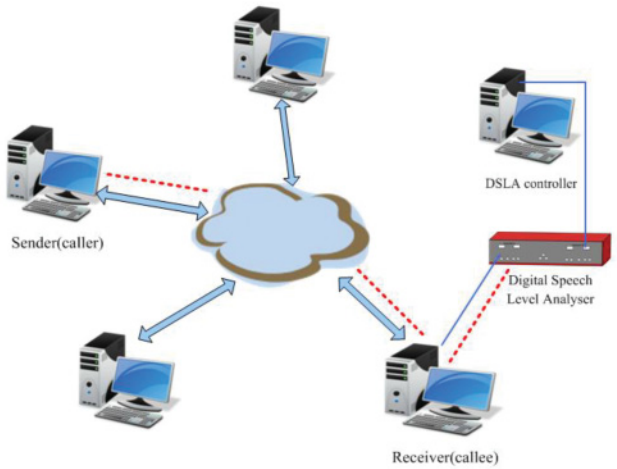


Fig. 3. Speech quality measurements using DSLA.

vertical coordinate as y_p, which is an assumed value of number of packets that would be lost in the x_pth packet-loss event. Iterate from the point $P(x_p, y_p)$ to obtain the corresponding points set. Compare them with the average value of the original interpolated points in the interval, get the smallest deviation as e_0, $e_{\min} = e_0$.

Step (3). Increase the vertical coordinate with a given length y_e, obtain the corresponding point set from the iteration starting from the point $P'(x_p, y_p+y_e)$. Get the deviation e' after compared with the original interpolated points in the interval. Take a smaller value between e' and $e_{\min}$ to replace the $e_{\min}$, and record the corresponding point P with the smaller deviation. Repeat Step (3) until the vertical coordinate exceeds $y_{\max}$.

Step (4). The recorded point $P(x_p, y_p+ky_e)$ with the smallest deviation is the predicted value that the model desires, which indicates that y_p+ky_e packets would be lost in the x_pth packet-loss event.

4. RESULTS AND DISCUSSION

4.1. Experiment Settings

To evaluate the performance of the proposed covert VoIP communication scheme in the presence of packet loss, different VoIP speech samples coded by PCM G.711 A-law were employed as the cover-speech. The proposed covert communication scheme was implemented on our VoIP communication platform called StegPhone. StegPhone is a real VoIP software application that has been developed based on MFC. The implement of speech signal acquisition and playback was based on winmm.lib which is a Windows multimedia API, and the real-time transmission of VoIP streams was developed based on jrtplib 3.9.1 library. The callee's IP address needs to be provided to initialize VoIP communication, and the parameters used for sampling and quantizing the cover-speech were then selected. The VoIP audio samples were obtained using mono channel and sampling at 8 kHz. Each sample was quantized to be represented with 16 bits, and so there were 1024 samples in each audio packet. The Gilbert packet-loss model was used to estimate and set the packet loss rate.

Figure 3 describes the experiments of testing the speech quality of the cover-speech and the stego-speech streams with Digital Speech Level Analyser (DSLA), which is high-accuracy speech testing equipment made by Malden Electronics Ltd. in the United Kingdom. DSLA can be used to achieve a real-time measurement of

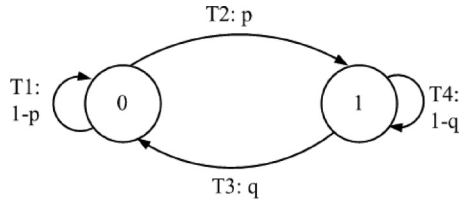


Fig. 4. Gilbert model.

VoIP speech quality by connecting to a microphone or earpiece device. And a set of parameters about speech quality can be obtained after the analysis of DSLA, such as PESQ, SNR, utterance and frame time offset statistics, frame-by-frame score data, voice activity and clipping analysis and so on. In the experiments, we used a player to play back records of English audio as the input to microphone for simulating the caller's speech. The audio samples were standard English and Chinese speech records downloaded from the ITU (International Telecommunication Union) website. The VoIP communication test was completed over our laboratory's local area network. Comparisons between cover-speech samples and stego-speech samples were carried out at the end of the VoIP call. At the receiving end, we measured Perceptual Evaluation of Speech Quality (PESQ) scores and Signal-to-Noise Ratio (SNR) values of cover-speech samples and stego-speech samples using DSLA.

4.2. Gilbert Model to Simulate a Network with Packet Loss

In the information theory, the binary channel is a classic noisy channel model. The average packet loss rate is used to describe the packet-loss feature of a network in such a memoryless channel model. But in actual networks, channels usually have a "memory" that there is a short-term correlation between packet-loss events. A Markov model could be used to capture the correlation of packet-loss temporarily, and it adopts a two-state Markov model, which is known as Gilbert model. The Gilbert model is often used to describe a burst packet loss of networks, and it was proposed by Gilbert in 1960. Following is a brief description of Gilbert model.

As Figure 4 shows, the general description of Gilbert model utilizes the probability of two independent events. Assuming a random variable X denotes whether the packet-loss event has occurred, $X = 0$ means no packet loss, i.e., the packet arrives at the destination. And $X = 1$ represents packet loss. In Figure 4, p stands for the probability of state "0" changing to state "1", and q is the probability that state "1" changes to state "0". The number of times that the packet-loss event has happened is decided by the consecutive packet-loss event, which means the number of times the transition of state "0" changes to state "1" is the same as the number of times the transition of state "1" transfers to state "0".

Assuming the sequence of the packet loss state is $\{i_1, i_2, \ldots, i_n\}$, $i_k = 0$ or 1, $1 \leq k \leq n$, where 0 indicates that the packet is lost, 1 means no packet loss. Thus, there are four possible events, which can be represented by T1, T2, T3, and T4. The probabilities of the four events' occurrences are $1 - p$, p, $1 - q$, and q, respectively. And $p + q < 1$.

$$\begin{aligned} &\mathrm{T1}: 0 \longrightarrow 0 \\ &\mathrm{T2}: 0 \longrightarrow 1 \\ &\mathrm{T3}: 1 \longrightarrow 0 \\ &\mathrm{T4}: 1 \longrightarrow 1 \end{aligned}$$

The specific procedures of the Gilbert packet-loss model at the sending end are described as follows:

Table II. Different Packet-Loss Levels

Packet-Loss Level	Level 1	Level 2	Level 3	Level 4	Level 5
p	0.02	0.05	0.07	0.10	0.20
q	0.90	0.85	0.67	0.70	0.50
p/q (Packet-loss rate)	0.022	0.059	0.104	0.143	0.400

Step (1). Predict whether the current packet is lost or not. If the prediction result is that the current packet is lost, there will be no embedding operation. If the current packet is predicted to survive, secret data are then embedded into the current packet using the steganographic algorithm in Tang et al. [2014].

Step (2). Decide whether to discard the current packet according to the Gilbert packet-loss model. If the previous packet state is "0", when T2 occurs, discard the current packet, and the state changes to "1"; when T1 occurs, include the current packet in the transmission queue, keep the state as "0". If the previous packet state is "1", when T3 occurs, include the current packet in the transmission queue, the state changes to "0"; when T1 occurs, discard the current packet, and maintain the state as "1".

In our experiments, the parameters of p, q were set ahead the communication, different values of p, q denoting different packet-loss rates. The Gilbert model was used to simulate the real network scenarios with different packet-loss rates. A random sequence with "0", "1" was generated using the Gilbert packet-loss model. "0" indicates that the packet is lost, which means the current voice packet would be discarded; "1" means no packet loss, that is to say that the corresponding voice packet would be sent without any change.

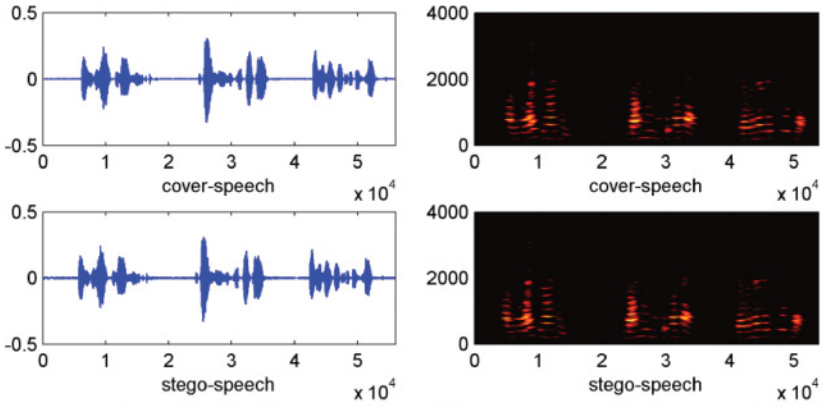
The Gilbert model with two-state Markov modes was used to create a network environment with packet loss in our laboratory, which was able to set a given packet-loss rate. The parameters p and q were set by the sender to obtain an intended packet loss rate.

As Table II shows, there were five different packet-loss levels setting in the experiments. Generally, the maximum packet-loss rate for a speech codec is about 5%, which means the speech quality is acceptable when the packet-loss rate is less than 5%. However, in order to measure the prediction accuracy and the data retrieving rate at different packet-loss levels, some packet-loss rates more than 5% were also implemented.

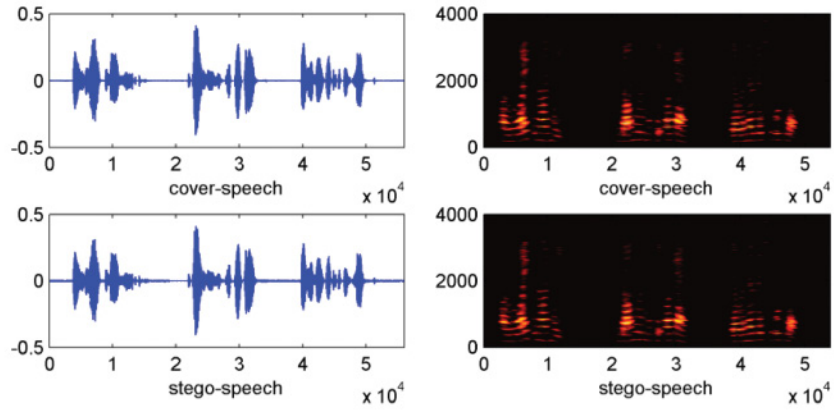
4.3. VoIP Steganography Results and Discussion

4.3.1. Experimental Results. To compare the results between embedding secret data in VoIP streams and no embedding in the presence of packet loss, two sets of tests were carried out, respectively. Female and male, English and Chinese speech tests were performed in each set of experiments. Experimental analysis included the waveforms in the time-domain and the spectrums in the frequency-domain of speech samples. The PESQ P.862.1 scores and SNR values of speech samples were measured using DSLA. To compare and analyze the performance of the proposed covert VoIP scheme in the presence of packet loss, the cover-speech samples were used as the reference, and the corresponding stego-speech samples were the degraded speech in the speech quality test by DSLA.

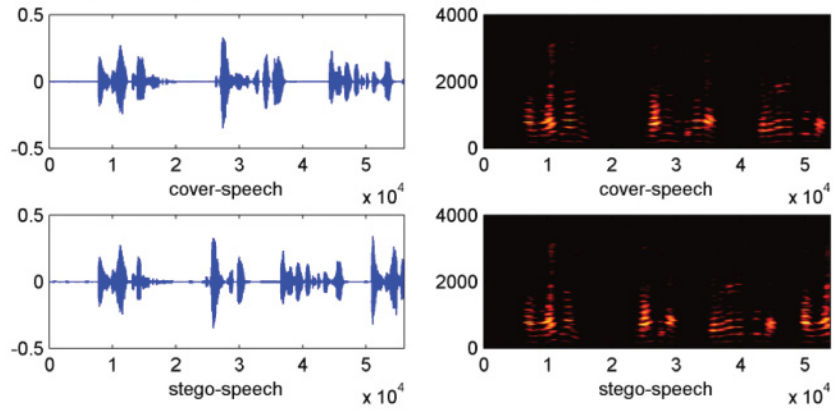
Figure 5 shows the waveforms in the time-domain and the spectrums in the frequency-domain of the original cover-speech samples and the stego-speech samples with embedding secret data in the Gilbert packet-loss network, respectively. As Figures 5(a1)(a2)(b1)(b2) show, there were almost no differences in the waveforms and the spectrums. This also meant that the proposed steganographic algorithm had no or little impact on the time domain and the frequency domain of the original cover-speech



(a1) Waveform comparison at packet-loss Level-1. (a2) Spectrum comparison at packet-loss Level-1.

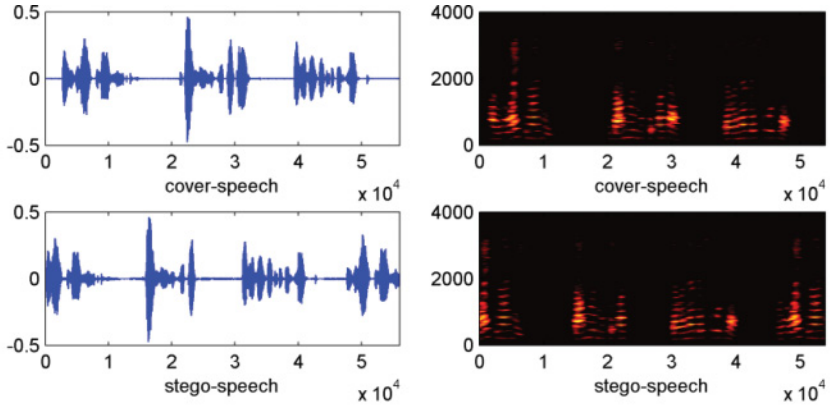


(b1) Waveform comparison at packet-loss Level-2. (b2) Spectrum comparison at packet-loss Level-2.

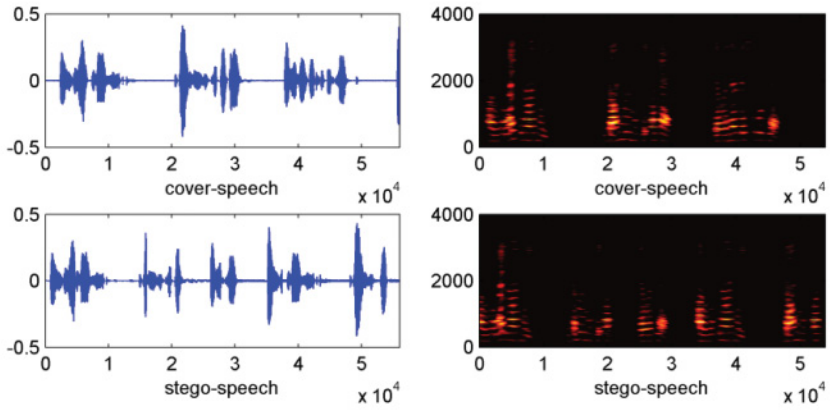


(c1) Waveform comparison at packet-loss Level-3. (c2) Spectrum comparison at packet-loss Level-3.

Fig. 5. Comparisons in the time-domain and the frequency-domain of cover-speech samples and stego-speech samples with the hidden secret data.



(d1) Waveform comparison at packet-loss Level-4. **(d2)** Spectrum comparison at packet-loss Level-4.



(e1) Waveform comparison at packet-loss Level-5. **(e2)** Spectrum comparison at packet-loss Level-5.

Fig. 5. Continued

samples. Compared the cover-speech samples with the stego-speech samples, it is obvious to see in the waveforms that some speeches are missing at packet-loss Level-3, Level-4, and Level-5. And the spectrums in the frequency domain are also different between the cover-speech samples and stego-speech samples at packet-loss Level-3, Level-4, and Level-5.

The stego-speech and cover-speech samples were analyzed by the cross-correlation function xcorr() in Matlab software to compare the differences of the speech samples at different packet-loss levels. The function xcorr(s_1, s_2) returns the cross-correlation of two discrete-time sequences s_1 and s_2, which stand for the cover-speech and stego-speech samples in our test. Cross-correlation measures the similarity between s_2 and shifted (lagged) copies of s_2 as a function of the lag.

Figure 6 shows the cross-correlation results of cover and stego-speech samples at different packet-loss levels. X axis represents the lag, expressing the delay as a number of samples and in seconds. Y axis denotes the cross-correlation degree. If $x = 0$, the greater the y value, the more similar the cover and stego-speech samples are, which means the fewer packets lost. And when $x \neq 0$, the more points of the y value of zero, the

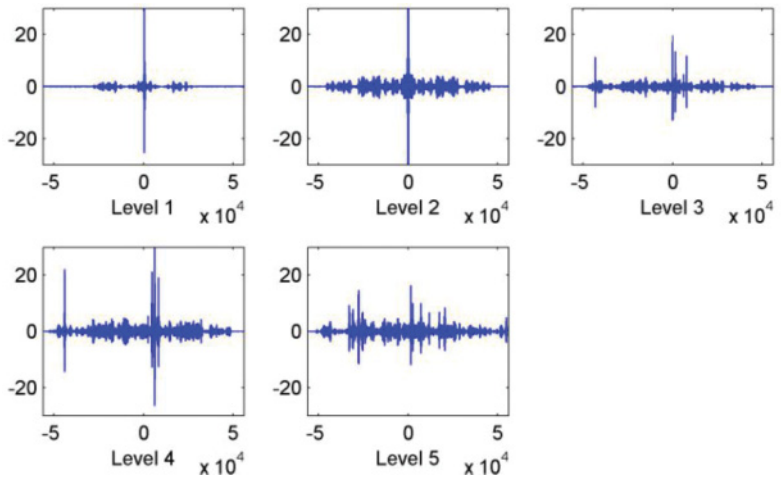


Fig. 6. The cross-correlation results of cover speech and stego speech at different packet-loss levels.

greater similarity of the cover and stego-speech samples. As can be seen in Figure 6, the cover-speech samples and the stego-speech samples at packet-loss Level-1 and Level-2 have more similarity. And the differences are greater with higher packet-loss levels.

In the PESQ measurements, the cover-speech samples played by the caller were used as the reference signal for DSLA input, and two speech categories were measured as the degraded signals. One category was the stego-speech samples with the hidden data based on the Gilbert packet-loss model received at the receiving end, marked as "with-embedding". The other category was the stego-speech samples without data embedding based on the Gilbert packet-loss model received at the receiving end, denoted as "no-embedding". Figures 7 and 8 show the differences and variation tendency of the PESQ and SNR values at different packet-loss levels. The data in Figures 6 and 7 are the average values of the results obtained from 15 repeated experimental measurements.

Figure 7 shows comparisons in the average PESQ score between the "no-embedding" stego-speech samples and the "with-embedding" stego-speech samples. As the packet-loss level increased, the speech quality of stego-speech samples appeared a decreasing trend. And the "with-embedding" stego-speech samples had worse speech quality than the "no-embedding" stego-speech samples. When the packet-loss level reached Level-3, the PESQ scores of the "no-embedding" stego-speech samples were smaller than 3, which is basically unacceptable in view of VoIP communication. The average variance of PESQ score between the "no-embedding" stego-speech samples and the "with-embedding" stego-speech samples was about 0.717, and the variances narrowed with the increasing packet-loss level, indicating that the packet-loss factor has a greater impact on speech samples than the applied steganographic algorithm. However, the PESQ scores were unusually high at the packet-loss Level-4 compared with those at the packet-loss level-3. The reason may be that most packet-loss events occurred in inactive speech periods at Level-4, which means packet loss in inactive speech periods had less impact on the speech quality. And comparing with embedding secret data, packet loss in active speech periods had more impact on quality of speech. So sometimes the "with-embedding" stego-speech samples have higher quality of speech than the corresponding "no-embedding" speech samples. Figure 8 shows comparisons in the mean SNR values between the "no-embedding" stego-speech samples and the "with-embedding" stego-speech samples. The SNR values had almost the same changing trends as the packet-loss level increased.

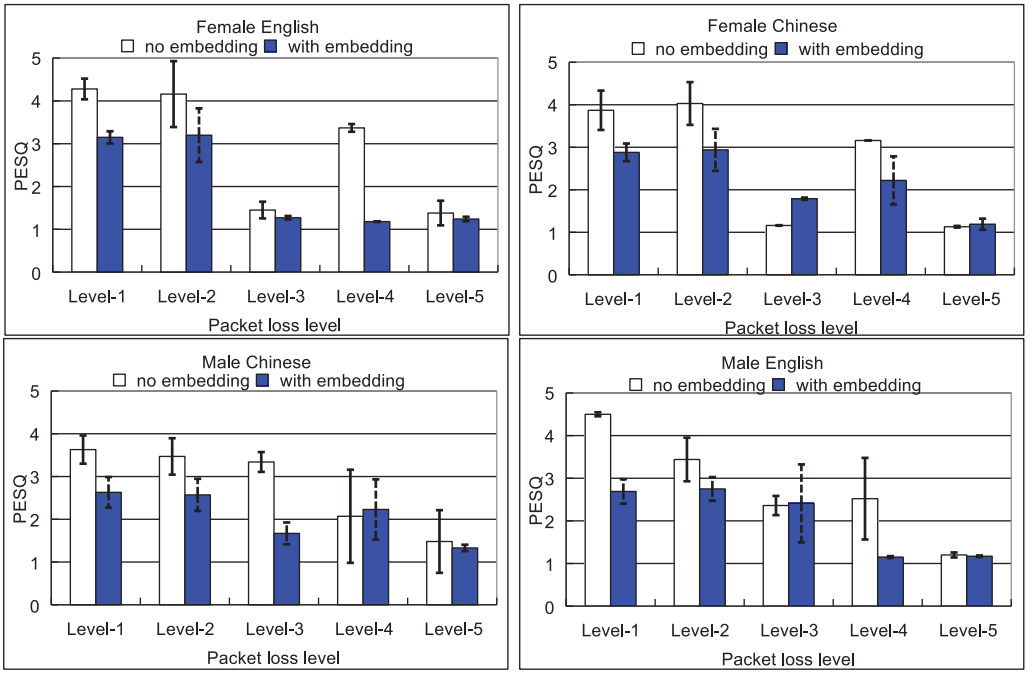


Fig. 7. Comparisons of the average PESQ scores between the "no-embedding" speech samples and the "with-embedding" stego-speech samples.

4.3.2. Statistical Undetectability. The Mann-Whitney-Wilcoxon (M-W-W) test was adopted to evaluate the statistical undetectability of the proposed steganographic algorithm. The M-W-W test is one of the best-known non-parametric significance tests, which can be used for assessing whether two independent samples of observations come from the same distribution. In our experiment, the M-W-W test was used to test the null hypothesis that samples in the cover-speech and the stego-speech samples are from continuous distributions with equal medians, i.e., the cover-speech and the stego-speech do not differ, against the alternative that they are not differ. The result, $H = 1$, indicates a rejection of the null hypothesis, and $H = 0$ indicates a failure to reject the null hypothesis at a significance level. In our test, the significance level was set to be 0.05. In computation of the M-W-W test, the length of speech samples was 6 seconds, and the number of samples (data points) in cover-speech and stego-speech was 56000. And the M-W-W test was conducted by using the ranksum() function in Matlab software. As shown in Table III, the values of H were 0 at different packet-loss levels, which indicated that the null hypothesis was true, i.e., the cover-speech and the stego-speech did not differ. This means that the proposed steganographic algorithm can withstand steganalysis based on statistical analysis.

4.3.3. Robustness Analysis. The VoIP communication system is based on the UDP protocol, which is connectionless and unreliable delivery. And there is no retransmission in a VoIP application. A packet-loss prediction was proposed to decide whether the VoIP packet with hidden data would be lost in a poor communication channel, therefore reducing the loss of secret data for increasing the robustness of covert communications. The higher the accuracy of the packet-loss prediction method, the better the robustness.

Figure 9 presents changes in the data-retrieving rate at different packet-loss levels. Close analysis of Figures 8 and 9 shows that both the average PESQ scores and the SNR

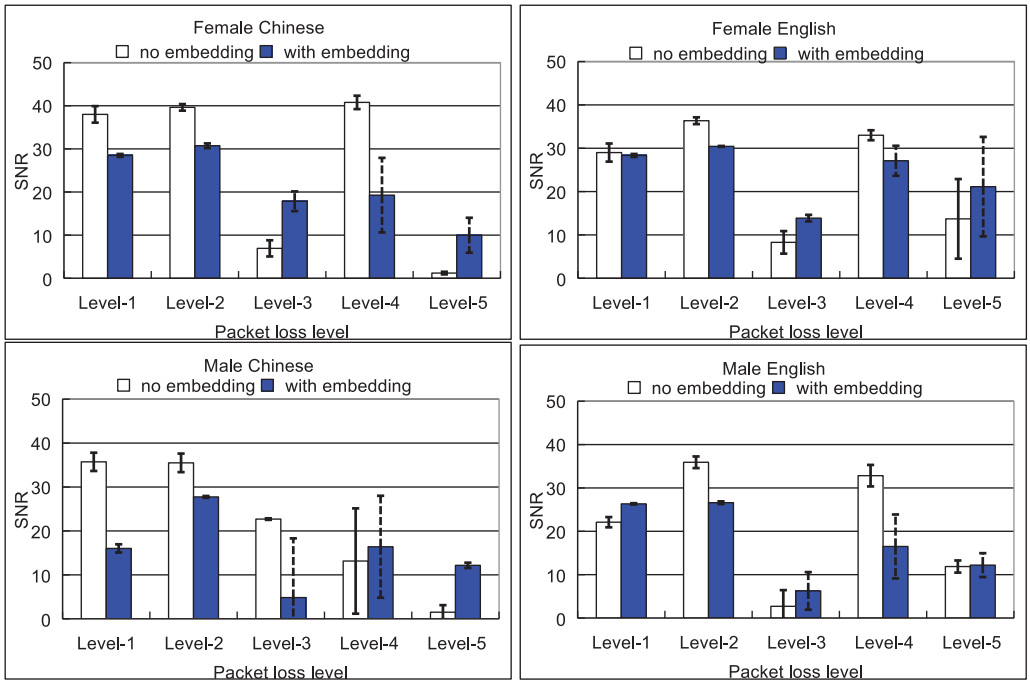


Fig. 8. Comparisons of the mean SNR values between the "no-embedding" speech samples and the "with-embedding" stego-speech samples.

Table III. M-W-W Analysis Results

Test	Number of Samples	Rank sum	z*	P-value	H
Level 1	56000	3.1378e+9	0.3014	0.7631	0
Level 2	56000	3.1365e+9	0.0744	0.9407	0
Level 3	56000	3.1366e+9	0.0851	0.9321	0
Level 4	56000	3.1396e+9	0.6450	0.5189	0
Level 5	56000	3.1368e+9	0.1230	0.9021	0

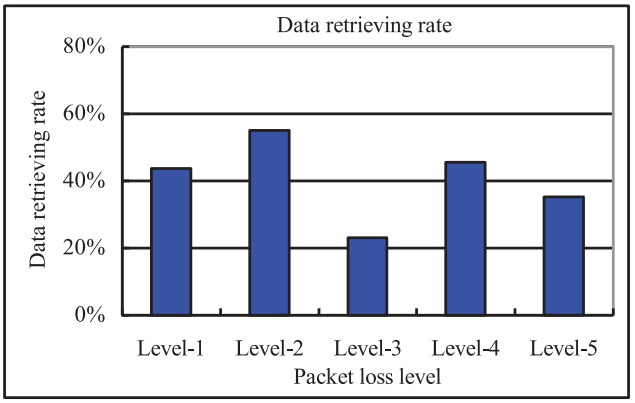


Fig. 9. Changes in the data retrieving rate with the increasing packet-loss.

Table IV. Comparisons of Our Scheme with Other VoIP Steganographic Methods

	Steganographic bandwidth (Kbits/s)	Undetectability	Consideration of packet loss
Proposed scheme	0.5	+	+
Miao and Huang [2011]	7.5	+	N/A
Wu and Yang [2006]	20	+	N/A
Huang et al. [2012]	0.1333	+	N/A
Tian et al. [2008]	0.8-2.6	+	N/A
Xu and Yang [2009]	0.1333	N/A	N/A
Takahashi and Lee [2007]	8	N/A	N/A
Liu et al. [2008]	0.2	+	N/A
Mazurczyk et al. [2014]	0∼10 (Time dependent)	+	+

Some data in Table IV was derived from Mazurczyk [2013].

values of stego-speech samples had the same varying trend with the data-retrieving rate as the packet-loss level increased, especially the speech quality was unusually high at the packet-loss level-4 compared with that at the packet-loss level-3. These results suggest that the success rate of the fractal prediction model affects the performance of the VoIP communication with steganography.

The data-retrieving rate stands for the accuracy of the packet-loss prediction. The higher the data-retrieving rate, the greater the robustness of the covert communication. Some secret data hidden in VoIP packets would be lost in a bad communication channel; the packet-loss prediction could save some but not all the secret data. An underlying secret data retransmission scheme based on TCP was proposed to succeed covert communications; however, it had a big impact on the real-time quality because of the time-consuming TCP delivery. As a result, the retransmission method was not included in the proposed scheme.

4.4. Comparisons with Other Steganographic Methods

To confirm the effectiveness of the proposed covert VoIP communication scheme in the presence of packet loss, performance comparisons were conducted by comparing the steganographic bandwidth, undetectability, and consideration of packet loss.

The steganographic bandwidth is an important performance metrics for VoIP steganography. In our experiments, the VoIP audio samples were obtained using mono channel and sampling at 8 kHz, which means that there were 128 ms speech in each packet with 1024 samples. And the steganographic bandwidth can be calculated by Equation (8):

$$\text{Steganographic bandwidth} = \frac{\text{The length of the hidden secret data in each packet (bit)}}{\text{Speech length in each packet (ms)}} \quad (8)$$

The value of the variable embedding interval used in the data-embedding steganographic algorithm was set as 16 for the proposed covert VoIP communication scheme in our experiments, and the average steganographic bandwidth in the experiments was determined to be 0.5kbit/s in the selected audio signals.

Table IV shows comparisons in the steganographic bandwidth, undetectability, and consideration of packet loss between the proposed covert VoIP communication scheme and some other existing VoIP steganographic methods. Results show that the proposed scheme has an acceptable steganographic bandwidth with undetectability, but

only our scheme had taken into account packet loss for VoIP communications with steganography.

5. CONCLUSIONS

In this study, a fractal-based VoIP steganographic approach was proposed to realize covert VoIP communications in the presence of packet loss. The prediction model based on fractal interpolation was built to decide whether a VoIP packet was suitable for data hiding. The proposed steganography scheme for covert VoIP communication with packet loss was implemented on our StegPhone VoIP communication platform. Female and male, English and Chinese speech tests were carried out in each set of experiments. The experimental results included the waveforms in the time-domain and the spectrums in the frequency domain of speech samples, and the PESQ scores and SNR values of the "no-embedding" and "with-embedding" stego-speech samples. As the packet-loss level increased, the speech quality of stego-speeches had a decreasing trend. Both the average PESQ scores and the SNR values of stego-speech samples and the data retrieving rates had almost the same varying trend when the packet-loss level increased, revealing that the success rate of the fractal prediction has a significant effect on the performance of VoIP communications with steganography.

The proposed covert VoIP communication scheme with fractal packet-loss prediction has a feature of generalizability, which means it can be applied to other speech codecs, such as G.729, G.723.1, and Speex. Covert VoIP communications based on different speech codecs need to employ different steganographic methods, resulting in different speech qualities and embedding capacities. When the proposed scheme is implemented with other speech codecs, the corresponding embedding algorithms should be used to obtain good quality of speech. Moreover, the packet-loss prediction model can also reduce the impact of packet loss on the secret data.

Further studies are necessary to determine the effectiveness of the proposed covert VoIP communication scheme when other low bit-rate VoIP codecs are used. A great effort should be made to achieve a high success rate with the prediction model.

REFERENCES

Naofumi Aoki. 2003. A packet loss concealment technique for VoIP using steganography. In *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* (Awaji Island, Japan). 470–473.

Naofumi Aoki. 2010. A semi-lossless steganography technique for G.711 telephony speech. In *Proceedings of the 6th International Conference on Intelligent Information Hiding and Multimedia Signal* (*IIH-MSP*) (Darmstadt, Germany). 15–17. DOI:10.1109/IIHMSP.2010.136

M. F. Barnsley. 1986. Fractal functions and interpolation. *Constructive Approximation* 2, 1, 303–329.

O. Cetin, F. Akar, A. T. Ozcerit, M. Cakiroglu, and C. Bayilmis. 2012. A blind steganography method based on histograms on video files. *Imaging Science Journal* 60, 75–82. DOI:10.1179/1743131X11Y.0000000004

Pao-Chi Chang and Hsin-Min Yu. 2002. Dither-like data hiding in multistage vector quantization of MELP and G.729 speech coding. In *Proceedings of the 36th Asilomar Conference on Signals, Systems and Computers*. (Pacific Grove, CA). 1199–1203. DOI:10.1109/ACSSC.2002.1196972

Nedeljko Cvejic and Tapio Seppänen. 2002. Increasing the capacity of LSB-based audio steganography. In *Proceedings of the 5th IEEE Workshop on Multimedia Signal Processing* (*MMSP*). IEEE, 336–338. DOI:10.1109/MMSP.2002.1203314

R. Darsana and Asha Vijayan. 2011. Audio steganography using modified LSB and PVD. In *Trends in Network and Communications*. Springer, Berlin, 1–20. DOI:10.1007/978-3-642-22543-7_2

E. N. Gilbert. 1960. Capacity of a burst-noise channel. *Bell System Technical Journal* 39, 1253–1265. DOI:10.1002/j.1538-7305.1960.tb03959.x

Driss Guerchi and Fatiha Djebbar. 2009. Narrowband speech hiding using vector quantization. *International Journal of Information and Communication Engineering* 5, 8.

D. Guerchi, H. M. Harmain, T. Rabie, and E. E. Mohamed. 2008. Speech secrecy: An FFT-based approach. *International Journal of Mathematics and Computer Science*. 1–19.

Yongfeng Huang, Bo Xiao, and Honghua Xiao. 2008. Implementation of covert communication based on steganography. In *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (Harbin, China). 1512–1515. DOI:10.1109/IIH-MSP.2008.174

Yongfeng Huang, Chenghao Liu, Shanyu Tang, and Sen Bai. 2012. Steganography integration into a low-bit rate speech codec. *IEEE Transactions on Information Forensics and Security* 1865–1875. DOI:10.1109/TIFS.2012.2218599

Yongfeng Huang, Shanyu Tang, and Jian Yuan. 2011b. Steganography in inactive frames of VoIP streams encoded by source codec. *IEEE Transactions on Information Forensics and Security* 296–306. DOI:10.1109/TIFS.2011.2108649

Yongfeng Huang, Shanyu Tang, C. Bao, and Yau Jim Yip. 2011a. Steganalysis of compressed speech to detect covert voice over Internet protocol channels. *IET Information Security* 26–32. DOI:10.1049/iet-ifs.2010.0032

Akinori Ito, Shun'ichiro Abe, and Yoiti Suzuki. 2010. Information hiding for G.711 speech based on substitution of least significant bits and estimation of tolerable distortion. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*. 1279–1286.

ITU-T Recommendation G.114. 2003. One way transmission time, 2003.

Christian Kratzer, Jana Dittmann, Thomas Vogel, and Reyk Hillert. 2006. Design and evaluation of steganography for Voice-over-IP. In *Proceedings of IEEE International Symposium on Circuits and Systems* (Kos, Greece). 2397–2340. DOI:10.1109/ISCAS.2006.1693105

Yeuan-Kuen Lee and Ling-Hwei Chen. 2000. High capacity image steganographic model. *IEE Proceedings of Vision, Image and Signal Processing* 147, 3, 288–294. DOI:10.1049/ip-vis:20000341

Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. 1994. On the self-similar nature of Ethernet traffic. *IEEE-ACM Transactions on Networking*. 1–15. DOI:10.1109/90.282603

Lihua Liu, Mingyu Li, Qiong Li, and Yan Liang. 2008. Perceptually transparent information hiding in G.729 bitstream. In *Proceedings of the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (Harbin, China). 406–409. DOI:10.1109/IIH-MSP.2008.297

Lisa M. Marvel, Charles G. Boncelet Jr, and Charles T. Retter. 1999. Spread spectrum image steganography. *IEEE Transactions on Image Processing* 8, 8, 1075–1083. DOI:10.1109/83.777088

Wojciech Mazurczyk. 2012. Lost Audio Packets Steganography: The first practical evaluation. *Security and Communication Networks* 1394–1403. DOI:10.1002/sec.502

Wojciech Mazurczyk. 2013. VoIP steganography and its detection-a survey. *ACM Computing Surveys* 46, 2, 20. DOI:10.1145/2543581.2543587

Wojciech Mazurczyk and Krzysztof Szczypiorski. 2008a. Covert channels in SIP for VoIP signalling. *Global E-Security*. Springer, Berlin, 65–72. DOI:10.1007/978-3-540-69403-8_9

Wojciech Mazurczyk and Krzysztof Szczypiorski. 2008b. Steganography of VoIP streams. *On the Move to Meaningful Internet Systems: OTM*. Springer Berlin, 1001–1018. DOI:10.1007/978-3-540-88873-4_6

Wojciech Mazurczyk, Józef Lubacz, and Krzysztof Szczypiorski. 2014. *On Steganography in Lost Audio Packets, Security and Communication Networks* 7, 12, 2602–2615. DOI:10.1002/sec.388

Rui Miao and Yongfeng Huang. 2011. An approach of covert communication based on the adaptive steganography scheme on Voice over IP. In *Proceedings of the IEEE International Conference on Communications (ICC)* (Kyoto, Japan). 1–5. DOI:10.1109/icc.2011.5962657

Tamer Rabie. 2006. A novel compression technique for super resolution color photography. In *Proceedings of IEEE International Conference on Innovations in Information Technology (IIT)* (Dubai). 1–5. DOI:10.1109/INNOVATIONS.2006.301979

Gustavus J. Simmons. 1984. The prisoners' problem and the subliminal channel. In *Advances in Cryptology*. Springer US, 51–67.

Chaoming Song, Shlomo Havlin, and Hernan A. Makse. 2005. Self-similarity of complex networks. *Nature* 392–395. DOI:10.1038/nature03248

Chaoming Song, Shlomo Havlin, and Hernan A. Makse. 2006. Origins of fractality in the growth of complex networks. *Nature Physics*, 275–281. DOI:10.1038/nphys266

Takehiro Takahashi and Wenke Lee. 2007. An assessment of VoIP covert channel threats. In *Proceedings of the 3rd International Conference on Security and Privacy in Communication Networks* (Nice, France). 371–380. DOI:10.1109/SECCOM.2007.4550357

Shanyu Tang, Yijing Jiang, Liping Zhang, and Zhangbin Zhou. 2014. Audio steganography with AES for real-time covert voice over internet protocol communications. *Science China Information Sciences* 1–14. DOI:10.1007/s11432-014-5063-2

Hui Tian, Hong Jiang, Ke Zhou, and Dan Feng. 2011. Adaptive partial-matching steganography for voice over IP using triple M sequences. *Computer Communications*. Elsevier, 2236–2247. DOI:10.1016/j.comcom.2011.07.003

Hui Tian, Ke Zhou, Hong Jiang, and Jin Liu. 2009. An m-sequence based steganography model for voice over IP. In *Proceedings of the 44th IEEE International Conference on Communications* (Dresden, Germany). 1–5. DOI:10.1109/ICC.2009.5198737

Hui Tian, Ke Zhou, Yongfeng Huang, and Dan Feng. 2008. A covert communication model based on least significant bits steganography in voice over IP. In *Proceedings of the 9th International Conference for Young Computer Scientists*. 647–652. DOI:10.1109/ICYCS.2008.394

Chungyi Wang and Quincy Wu. 2007. Information hiding in real-time VoIP streams. In *Proceedings of the 9th IEEE International Symposium on Multimedia* (Taichung, Taiwan). 255–262. DOI:10.1109/ISM.2007.4412381

Walter Willinger, Murad S. Taqqu, R. Sherman, and D. V. Wilson. 1997. Self-similarity through high-variability: statistical analysis of ethernet LAN traffic at the source level. *IEEE-ACM Transactions on Networking* 71–86. DOI:10.1109/90.554723

Zhijun Wu and Wei Yang. 2006. G.711-Based adaptive speech information hiding approach. In *Proceedings of the International Conference on Intelligent Computing* (*ICIC*). 1139–1144. DOI:10.1007/11816157_141

Chunbo Xiu, Tiantian Wang, Meng Tian, Yanqing Li, and Yi Cheng. 2014. Short-term prediction method of wind speed series based on fractal interpolation. *Chaos Solitons & Fractals* 89–97. DOI:10.1016/j.chaos.2014.07.013

Tingting Xu and Zhen Yang. 2009. Simple and effective speech steganography in G.723.1 low-rate codes. In *Proceedings of International Conference on Wireless communications & Signal Processing (WCSP)* (Nanjing, China). 1–4. DOI:10.1109/WCSP.2009.5371745

Cheng-Hsing Yang, Chi-Yao Weng, Shiuh-Jeng Wang, and Hung-Min Sun. 2008. Adaptive data hiding in edge areas of images with spatial LSB domain systems. *IEEE Transactions on Information Forensics and Security* 3, 3, 488–497. DOI:10.1109/TIFS.2008.926097

Ke Zhou, Jin Liu, Hui Tian, and Chunhua Li. 2012. State-based steganography in low bit rate speech. In *Proceedings of the 20th ACM International Conference on Multimedia*. 1109–1112. DOI:10.1145/2393347.2396395

Received January 2016; revised April 2016; accepted June 2016