An efficient and secure VoIP communication system with chaotic mapping and message digest

An Efficient and Secure VoIP Communication System with Chaotic Mapping and Message Digest

Yijing Jiang[1], Shanyu Tang [2*]

[1] School of Computer Science, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China
[2] School of Computing and Engineering, The University of West London, St Mary's Road, Ealing, London W5 5RF, United Kingdom *shanyu.tang@uwl.ac.uk

Abstract As one of the most popular real-time services on the Internet, Voice over Internet Protocol (VoIP) has attracted researchers' attention in the information security field for its characters of real-time and high flow. To protect data security, a new covert VoIP communications system was proposed in this study to realize secure communications by hiding secret information in VoIP streams. In the proposed algorithm, secret data were divided into blocks after being encrypted with a block cipher, and then each block of secret data was embedded into VoIP streaming packets randomly by using a chaotic mapping. The symmetric key was distributed through an efficient and secure channel, and the message digest was implemented to protect the integrity of secret data. The experimental data were collected by comparing audio data between the sender and the receiver. The experimental results indicated that data embedding had little impact on the quality of speech. Besides, statistical analysis could not detect the secret data embedded in VoIP streams using the block cipher and random numbers generated from chaotic mapping. The proposed covert VoIP communications system not only achieved high quality of VoIP and prevented detection of statistical analysis, but also provided integrity for secret data.

1 Introduction

Voice over Internet Protocol (VoIP) is one of the most popular real-time services on the Internet. VoIP is also called IP phone, which is used to make phone calls through the Internet. VoIP has more advantages than traditional telephones on Public Switched Telephone Network (PSTN), because the Internet allows VoIP to provide high-reliability, global and low-cost or even free services.

Due to the outstanding features of VoIP, more and more users communicate with each other daily by using VoIP products, such as Skype. And VoIP streams play an increasingly important part in the Internet traffic. VoIP has attracted more and more attention from the researchers in the field of information security, and it is also considered to be a better cover object for information hiding compared with traditional static cover objects, such as text files and image files. Steganography is a method of embedding secret information into a cover object, which should not cause unacceptable distortion and arouse observers' attention. Message digest is the output of hash algorithm, which makes the message with different lengths into a fixed-length string, provides error-detecting capacity, and is also widely used in secure communications.

One of the most important services that VoIP provides is the real-time service, which means that attackers do not have sufficient time to detect whether VoIP streams have embedded secret data or not. The real-time characteristic is useful to improve the security of secret data hidden in VoIP streams. However, the real-time requirement makes it more difficult for ones to introduce more security operations to embed and protect the secret data.

Early studies about information hiding were widely applied in various digital media, such as text, image [1-5], audio [6-8], and video [9] files. Because of the insensitivity of Human Visual System (HVS), human eyes cannot make a clear distinction between the original image and the stego-image with data embedded. It is generally recognized that audio steganography is more challenging than image steganography for the wider dynamic range of Human Auditory System (HAS) in comparison with HVS.

The rest of the paper is organized as follows. Section 2 introduces the related work. The proposed VoIP covert communication system is presented in Section 3. Experimental results and analysis are discussed in Section 4. Finally, a summary is given in Section 5.

2 Related work

Generally speaking, modifying the least significant bits (LSB) of the sampled voice data causes the smallest distortion. LSB algorithms have been applied to many covert communication systems. The first VoIP steganographic method to utilize digital voice signal as cover object was proposed by Aoki [10] in 2003, which used steganography to improve its resistance to packet losses and voice quality. Miao et al. [11] proposed an adaptive steganography system which implements different embedding methods in active frames and inactive frames by analysing the character of audio.

Wu et al. [12] proposed a scheme about the least significant bit based on G.711 speech codec, which estimates the hiding capacity of the cover speech in each packet by calculating the statistical characteristics of speech energy. Their experimental results show that it achieved a better embedding capacity than traditional LSB steganography method, and had little effect on the quality of cover speech. However, it is not enough to only consider the variance in speech quality caused by the steganography algorithm, and

the transmission error is inevitable between the caller and callee.

Ito et al. [13] suggested that less distortion would not necessarily improve the subjective speech quality. They introduced an improved LSB algorithm based on tolerable distortion, which can improve the speech quality. Tian et al. [14] proposed a VoIP covert communication model based on LSB steganography. It is an improved model based on the method in [15], which encrypts secret data before embedding them into speech and provides a short-term protection of secret data with short delay.

Xu et al. [16] proposed a steganography algorithm based on G.723.1 codec with 5.3kbit/s coding rate speech transmission model. Its implementation is hiding the secret data by replacing the least significant bit of LSP (Linear Spectrum Pair) quantization parameters with the bits stream of secret data. The experimental results show the method achieved 133.3bit/s steganographic bandwidth.

Mazurczyk et al. developed a steganography method in [17], which forms a covert channel based on VoIP streams. They also suggested two ideas about VoIP steganography. The first one is named network steganography scheme, which utilizes the unused field in Internet protocols, such as UDP, RTP, RTCP protocols. The second one is called Lost Audio Packets Steganography (LACK), which makes use of the delayed audio packets to achieve a covert channel of mixed time storage. Abdullaziz et al. [18] also proposed an active IP identification based network steganography.

Considering VoIP speech distortion caused by the inevitable packet loss and delay in the real networks, Aoki [10] introduced an error concealment method, which improves the side information based on the reconstruction technique at the sending end, and the packets could be fully compatible with traditional VoIP format. Mazurczyk proposed a method in [19], which makes use of the packet loss and retransmission to hide secret data, whereas the voice packet with retransmission could not be utilized by application to play out.

Aoki [20] proposed a lossless steganographic approach for μ-law of G.711 codec, which utilizes the redundancy of G.711 codec to hide secret data without speech distortion. Zhou et al. [21] proposed a state-based steganographic algorithm, which implements on G.723.1 low bit rate speech codec, and modifies the G.723.1 transmission parameters to hide secret data. Huang et al. [22] suggested an algorithm for embedding data in some parameters of inactive speech frames encoded by G.723.1 codec, which achieves a high steganographic capacity. They [23] 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. [24] designed an M-sequence-based LSB steganographic algorithm for embedding information in VoIP streams encoded by G.729a codec. In [25] they also developed an adaptive partial-matching steganography method with triple M sequences, which uses a partial similarity value to evaluate the partial matching between the cover object and secret data. Besides, Chang et al. [26] proposed a steganographic method that hid speech data in the MELP and G.729 encoded speech. It is a method for hiding secret speech in the static speech, which uses multistage vector quantization and dithering techniques to ensure the speech quality. Wu et al. [27] proposed an approach of covert communication by embedding a 2.4Kbps low bit rate MELP speech into G.729 coding speech. The method adapts the techniques of covering code and the interleaving to achieve covert communications.

Rahman et al. [28] proposed a method to prevent the steganographic method, which uses transcoding for covert communication in [29]. It introduces a tamper-proof method based on signature. The signature of the original speech is appended to the RTP packet for providing integrity. But the digital signature in RTP packet can also be modified. When the attacker modified the speech data to hide secret data, she/he could also calculate the signature of the modified speech data to forge the signature of the sender during the whole communication period.

Chaotic map has also been used in steganography technique. Thenmozhi et al. [30] proposed a novel technique for image steganography based on DWT (Discrete Wavelet Transform); the chaotic system was utilized in achieving the encryption. Anees et al. [31] proposed a steganography technique in spatial domain for digital images based upon chaotic maps, which uses three chaotic maps to select the row number, column number and frame number of a specific pixel in the carrier image respectively.

However, all the studies above consider where to hide the secret message only and how to improve the embedding capacity. This study is different from the related works, and proposes a new and completed scheme of covert VoIP communication, including not only the embedding and retrieving process, but also the key distribution, the security and integrity of the secret message. Moreover, both of the communicating parties can embed secret data without disrupting each other.

3 The proposed VoIP covert communication system

3.1 Architecture of covert VoIP communications system

The proposed covert VoIP communication system takes full account of the security and integrity of secret data, which implements the efficient symmetrical encryption algorithm to encrypt secret data and calculate the message digest value of the plaintext of secret data by using MD5 hash algorithm. The bit streams of secret data are not uniformly embedded into the original audio signal, but distributed randomly by using a chaotic sequence generated from the logistic chaotic mapping.

As Figure 1 shows, the embedding end of the proposed VoIP covert communication system is composed of four modules in the embedding end, including pre-processing module, embedding module, calculating message digest module, and network transmission. The calculating message digest module is aiming at generating a unique and fixed-length message authentication code (MAC) of secret data for verification. The main task of the pre-processing module is to pre-process secret data before hiding them, the pre-

processing operation includes compressing and encrypting secret data, in order to send as much information as possible and guaranty the security of secret data. The embedding module is the core of the VoIP steganography system, which in fact determines whether the location of embedding secret data is undetectable. The outputs at the embedding end consist of the message digest value, the key used to encrypt secret data, the initial value of logistic mapping, and the stego-speech samples which contain the embedded secret data.

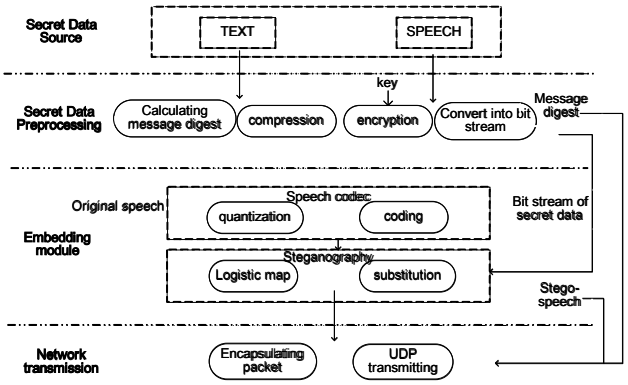

Fig. 1 The embedding end of VoIP steganography system

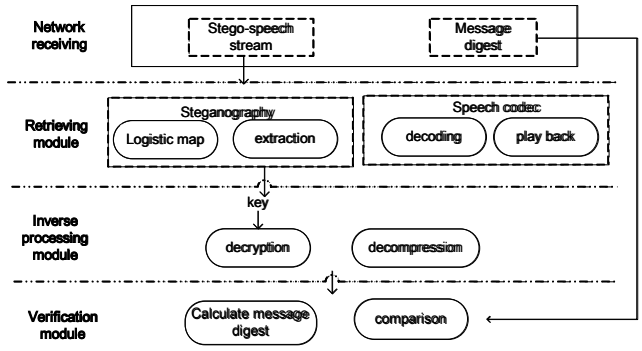

Fig. 2 The retrieving end of VoIP steganography system.

Figure 2 shows the framework of the VoIP steganography system at the receiving end, which consists of four modules, mainly consisting of the network receiving, the retrieving module, the inverse processing module, and the verification module. The process of each module is the inverse of the corresponding module at the embedding end. Before receiving the stego-speech stream, the receiving end gets the message digest value of secret data, the session key, and the initial value of logistic mapping. After the processing of all the modules, the plaintext of secret data and verification result are obtained at the receiving end. Finally the stego-speech samples play back in real-time.

3.2 Calculating message digest module

Message authentication is a security scheme for verifying the integrity of message, making sure that the data received is the same as the data sent. In other words, there is no modification, insertion or removing in communications. Hash algorithm makes the message with different lengths into a fixed-length string, which is called message digest. Message digest provides error-detecting capacity, i.e., message digest varies whereas any bit of the message changes. In addition, attackers cannot get the original message from message digest, as a result of that hash algorithm is one-way function.

Figure 3 shows the application of hash algorithm in the proposed VoIP covert communication system. In Figure 3, M stands for the original secret data, H represents the hash algorithm, and E denotes the encrypting operation. The original secret data is encrypted by AES encryption algorithm and calculated to get the corresponding message digest, respectively. And the message digest and the cipher text of secret data are transmitted separately. The bits stream of cipher text is embedded into speech stream by the steganographic algorithm. The transmitting of message digest is implemented at the signalling phase before communication. At the end of the communication, the received message digest is used to verify the integrity of secret data at the receiving end after the completion of the extraction.

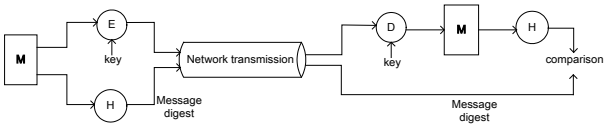

Fig. 3 Message authentication with message digest

3.3 Compression and encryption

Generally, there are two processing operations in the pre-processing module, compressing and encrypting the secret data. The compressing operation is not essential in the VoIP steganography system, it is used to decrease the size of secret data, and the compressing method depends on the type of secret data, such as image, voice, string and so forth. The encryption of secret message is the vital part in the pre-processing module, which makes secret data unreadable and ensures its confidentiality.

Table 1 Comparison of some common symmetric encryption algorithms.

algorithm	Key size	speed	security	cost
DES	56bits	fast	low	middle
3DES	112,168bits	slow	middle	high
AES	128,192,256bits	fast	high	low

The encryption algorithm implemented in the proposed steganography system is one of the symmetric encryption algorithms. As Table 1 shows, compared with some common symmetric encryption algorithms, Advanced Encryption Standard (AES) is more suitable for encrypting messages, because of the fast speed, high security, and low cost of AES encryption algorithm. And AES with 128 bits key length is enough to achieve good performance.

There are four block-cipher modes of operation, electronic codebook (ECB), cipher block chaining (CBC), cipher feedback (CFB), and output feedback (OFB). The plaintext needs to be padded in ECB and CBC mode encryption, and the length of cipher text is different from the length of the original plaintext after encryption. In CFB and OFB mode encryption, there is no padding in plaintext, and the length of cipher text is in accordance with the length of plaintext. In order to reduce the amount of information for embedding, CFB mode encryption is implemented in the proposed VoIP steganography system.

Suppose the original secret data are denoted as M, $M = [m_0, m_1, ..., m_{N-1}]$, and N is the length of original secret data. Encrypting the secret data can be expressed as follows:

$$S = E_{key}(M) \tag{1}$$

Where $S = [s_0, s_1, ..., s_{N-1}]$ is the cipher text, and E is the CFB mode encryption of AES. *Key* is the symmetric key used to encrypt secret data, and it is the session key distributed between the communicating parties.

3.4 Embedding module

The embedding algorithm in our covert VoIP communication system is based on LSB algorithm, but using the chaotic sequence to determine the embedding location.

Logistic mapping is one of the most popular models for discrete nonlinear dynamical systems. A logistic map can be described in equation of:

$$x_{n+1} = rx_n - rx_n^2 = rx_n(1 - x_n) \tag{2}$$

where x_n denotes the value of x_0 after n iterations, a real number between 0 and 1. In discrete-time demographic model [20] x_n represents the ratio of existing population to the maximum possible population after n years. And r is a positive number, standing for a rate for reproduction and starvation.

Figure 4 depicts a bifurcation diagram of the logistic map when x_0 = 0.502. As we can see from Figure 4, when $r \in (0,1]$, the value of x is eventually going to be 0. When $r \in (1,3]$, the value of x quickly approaches the value of r-1/r. It shows chaos characteristics when r changes among the range of (3.57, 4]. When r = 4, the value of x becomes increasingly chaotic. Figure 5 describes the initial condition sensitivity of chaotic system, when x_0 changes slightly near 0.502 and r = 4 in the equation of logistic map. The sensitivity of initial conditions in chaotic system means that a small difference in initial conditions leads to great changes on chaotic orbit. And we can see that the values of x_n randomly fall in the range (0, 1] as the value of n increases, the orbit changes a lot when x_0 changes slightly.

The utilization of chaotic map makes the embedding of secret information randomly, and it is very difficult to predict the initial conditions based upon chaotic sequences. These properties of chaotic map could increase the undetectability of covert communication.

Since the value of x_n is between 0 and 1, some adjustments should be made to get a chaotic sequence of integers from logistic map. Supposing x_0 and r are given as a certain value in equation (2), then we can get a sequence of x, denoted by set X,

$$X = \{x_i \mid i = 0,1,2,...,n\} \tag{3}$$

And

$$R = \{k_i = x_i * 1000 \pmod{p} + 1 \mid i = 0,1,2,...,n\} \tag{4}$$

The sequence of R is utilized in our covert VoIP communication. The value of p is determined to be 16 as the largest interval. And k_i determines the location to embed the bits stream of secret data.

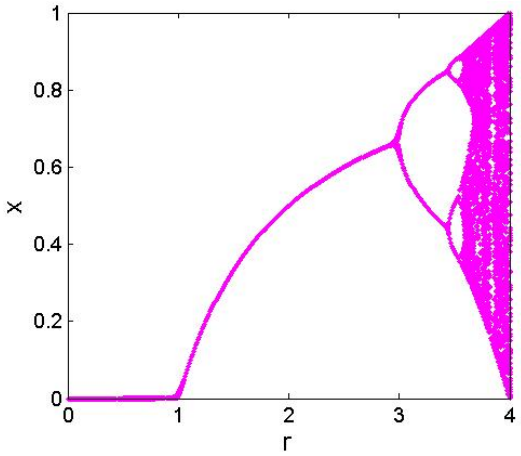

Fig. 4 Bifurcation diagram of logistic map when x_0 = 0.502

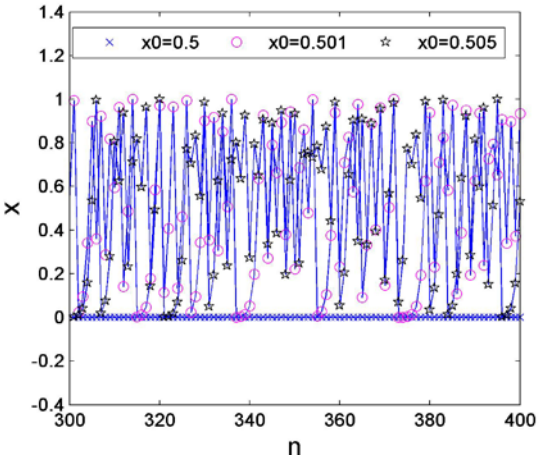

Fig. 5 The sensitivity of initial condition in chaotic system.

In the proposed covert VoIP communication system, assume N is the length of original secret data M, the cipher text of secret data is described as $S = \{s_i \mid i = 0, 1, 2, ..., N\text{-}1\}$. If the number of sample in the speech of each packet is Q, and the least significant bit set of samples is denoted by B =

$\{b_i \mid i = 0, 1, 2, ..., Q\text{-}1\}$. The pseudo-code of embedding process is designed as follows:

Embedding algorithm

```
begin
i ← 0
j ← 0
while   i < N and j < Q
do
      x_i ← chaoticmap()
      k_i ← x_i*1000 (mod p) + 1
      if   j+k_i < Q
      then
            b_{j+ki} ← s_i
            j ← j + k_i,
            i ← i + 1
      end if
      N=N-i
end
```

3.5 Retrieving module

At the signalling phase, the receiver receives the initial values of logistic map. The extraction process is a reverse process of the embedding process. After receiving an audio packet, the least significant bit set of samples is denoted by $B' = \{b'_i \mid i = 0, 1, 2, ..., Q\text{-}1\}$. The logistic chaotic map generates a corresponding random x_i that decides the extraction location. The pseudo-code of retrieving algorithm to extract the original secret message at the receiving end is as follows.

Retrieving algorithm

```
begin
i ← 0
j ← 0
while   i < N and j < Q
do
      x_i ← chaoticmap()
      k_i ← x_i*1000 (mod p) + 1
      if   j+k_i < Q
      then
            s_i ← b_{j+ki}
            j ← j + k_i,
            i ← i + 1
      end if
      N=N-i
end
```

3.6 Inverse processing module and verification module

The inverse processing module is to complete the decryption of the retrieved cipher text. The cipher text is obtained after the retrieving module, expressed as $S' = [s'_0, s'_1, ..., s'_{N-1}]$ which is the cipher text.

$$M' = D_{key}(S') \tag{5}$$

where D is the CFB mode decryption of AES, $M' = [m'_0, m'_1, ..., m'_{N-1}]$ is the plaintext of secret data, which is acquired at the receiving end eventually. *Key* is the symmetric key to decrypt secret data.

To ensure the integrity of secret data through communication, the verification module is the final stage. In the verification module, the same hash algorithm is implemented on the plaintext *S'* to calculate the corresponding message digest. Then the calculated message digest is compared with the message digest previously received; if they are the same, it means the secret data have not been modified during transmission. Otherwise, the secret data may have been tampered and cannot be trusted.

4 Experiment and Results

4.1 Experiment setting

In the experiments, the proposed steganographic algorithm was implemented on our VoIP communication platform called StegPhone. StegPhone is a real VoIP application which was developed based on visual studio 2012 MFC application. The implementations of speech signal acquisition and playback were based on winmm.lib library which is Windows multimedia API, and the real-time transmission of speech was based on jrtplib 3.9.1 library. The cryptography algorithm and hash algorithm used in the proposed system are from crypto++ library 5.6.2 version. The communicating parties' IP address needs to initialize VoIP communications. The speech codec in StegPhone was pulse code modulation (PCM), and the parameters used for sampling and quantizing the cover-speech could be selected. Besides, the VoIP application could be used as normal network phone without steganography, and it is convenient to be set.

There were about 6MB secret messages hidden in the VoIP streams in each VoIP call in the experiment, and the embedding capacity is not deterministic because of the random chaotic numbers used. In the experiments, the largest embedding interval was set to 16, and the steganographic bitrate was between 0.5 - 8 Kbits/s. More than 2 minutes VoIP communications were performed to conceal all the secret information. To clearly show the impact on speech, the first 60000 samples with the hidden secret data were used for analysis.

The VoIP audio samples were obtained by using single-channel and sampling at 8 kHz. Each sample was represented with 16 bits. There was 128 ms speech in each packet with 1024 samples. To evaluate the performance of the proposed steganography system, 20 male speech and 20 female speech

tests were performed in the experiments. In the experiments, the speech quality of the cover-speech and stego-speech streams was tested with DSLA. And the speeches for testing were played back to microphone as the cover-speech, which was to simulate the real-time communication. The audio samples were standard English records. As Figure 6 shows, the VoIP communication was achieved 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, Perceptual Evaluation of Speech Quality (PESQ) [32] scores and Signal-to-Noise Ratio (SNR) values of the speech samples were measured using DSLA, which is high-accuracy equipment made by Malden Electronics Ltd. in the United Kingdom.

The value of x_i in equation (4) was used for determining the embedding location in VoIP streams for embedding secret data. The value of p was determined to be 16 as the largest interval. And the value of r_i was used for determining the embedding location in VoIP streams for embedding secret data.

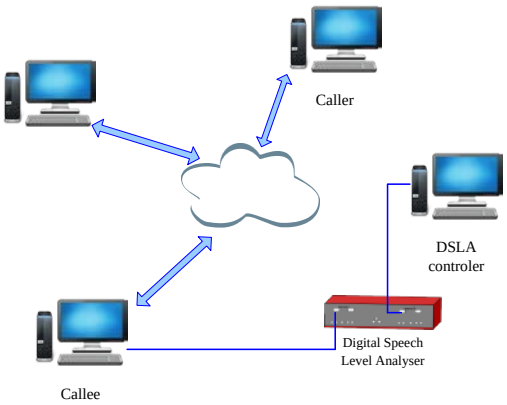

Fig. 6 Speech quality testing by using DSLA (Digital Speech Level Analyser).

4.2 Experimental results

The time-domain and the frequency-domain of speech samples were analyzed. The PESQ P.862.1 scores and SNR values of speech samples measured by DSLA were the experimental results. Female and male speech samples were taken for each test. To compare and analyze the performance of the proposed steganographic algorithm, the corresponding cover-speech samples were used as the references signal, the stego-speech samples were the input of the DSLA to be tested.

Figure 7 shows the waveforms in the time-domain and the spectrums in the frequency-domain of two female and two male speeches in the original cover-speech samples and stego-speech samples with embedding secret data, respectively. As Figure 7 shows, there is almost no difference in the waveforms and spectrums between the cover-speech and stego-speech samples with steganography. This also means that the proposed steganographic algorithm has no or little impact on the time domain and frequency domain of the original cover-speech samples.

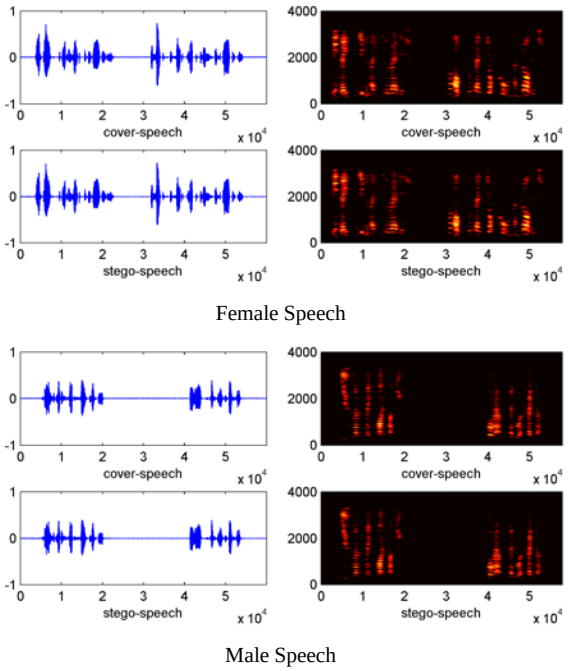

Fig. 7 Comparisons in time-domain and frequency-domain of cover-speech and stego-speech samples with data embedding.

Table 2 Mean PESQ scores and SNR values of the stego-speech samples.

Speech sample	Mean PESQ score	Mean SNR value
Female speech	4.01	40.01
Male speech	4.08	37.40

Table 2 lists the average PESQ scores and SNR values of female and male stego-speech samples with embedding secret data, respectively. As can be seen from Table 2, the average PESQ scores were near 4, which means that the speech quality of the stego-speech samples is acceptable.

4.3 Statistical analysis

The Mann-Whitney-Wilcoxon (M-W-W) test was adopted to evaluate the security of the proposed steganographic algorithm in this study. 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 [33]. Comparisons in probability distributions between the cover-speech and the stego-speech show whether the differences are almost indistinguishable.

The M-W-W test in our experiment analysis could be 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 different. 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. When the sample sizes are sufficiently large, the M-W-W test is based on the standardized test statistic:

$$z^* = \frac{S_2 - E\{S_2\}}{\sigma\{S_2\}} \tag{6}$$

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. To have 95 percent confidence, *i.e.* with a confidence coefficient (1-α) of 0.95, where α is called the level of significance, it therefore requires z(1 - α/2) = z(0.975) = 1.960, where z is the percentile of the standard normal distribution. Thus, the decision rule for the test is as follows:

If | z*| ≤ 1.960, conclude H = 0 (the null hypothesis is true).

If | z*| > 1.960, conclude H = 1 (the null hypothesis is false).

In statistical significance testing, the *P*-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. Researchers often "reject the null hypothesis" when the *P*-value turned 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.

As can be seen from Table 3, female and male speeches in the experiment had more than 60000 samples in the M-W-W test. The P-values were considerably larger than the significance level 0.05 for embedding secret data into female and male speeches, respectively. Moreover, the | z^*| values were smaller than 1.960, and the values of H were 0 for all the tests, 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.

Table 3 M-W-W analysis results

Test	Sample number	Rank sum	z^*	*P*-value	*H*
Female speech	65536	4.2948e+9	-0.0354	0.9717	0
Male speech	65024	4.2276e+9	-0.0755	0.9398	0

5 Conclusion

In this paper, an efficient and secure covert VoIP communications system with chaotic mapping and message digest was proposed to realize secure VoIP communications. In the proposed algorithm, secret data were divided into blocks after being encrypted with a block cipher, and then each block of secret data was embedded into VoIP stream packets randomly by using chaotic mapping. At the signalling phase, the symmetric key of encryption was distributed through an efficient and secure channel, and the message digest was calculated and sent to the receiver for protecting the integrity of secret data. The experimental results indicated that the data embedding had little impact on the quality of speech, and the statistical analysis with Mann-Whitney-Wilcoxon test could not detect the existing of the secret data embedded in VoIP streams. The proposed steganographic system with chaotic mapping and message digest achieved high quality of VoIP, prevented detection of statistical analysis, and provided integrity for secret data.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China under Grant 61272469.

References

1. Yang C. H., Weng C. Y., Wang S. J., Sun H. M.: Adaptive data hiding in edge areas of images with spatial LSB domain systems. IEEE Trans. Information Forensics and Security (3), 488-497 (2008)
2. Lee Y. K., Chen L. H.: High capacity image steganographic model. IEE Proc. Vision Image & Signal Processing, 147 (3), 288-294 (2000)
3. Sedighi, V., Cogranne, R., Fridrich, J.: Content-Adaptive Steganography by Minimizing Statistical Detectability. *Information Forensics and Security, IEEE Transactions on*, *11*(2), 221-234. (2016)
4. Provos N., Honeyman P.: Hide and seek: An introduction to steganography. Security & Privacy, IEEE (1), 32-44 (2003)
5. Al-Dmour H., Al-Ani A.: A steganography embedding method based on edge identification and XOR coding. Expert Systems with Applications (46), 293-306 (2016)
6. Darsana R., Vijayan A.: Audio steganography using modified LSB and PVD. Trends in Networks and Communications (197), 11-20 (2011)
7. Cvejic N., Seppanen T.: Increasing the capacity of LSB-based audio steganography. In Proc. 5th IEEE Workshop on Multimedia Signal Processing (MMSP), St Thomas, pp. 336-338 (2016)
8. Bao P., Ma X. H.: MP3-resistant music steganography based on dynamic range transform. In Proc. IEEE Int. Symp. Intelligent Signal Processing and Communication Systems (ISPACS), pp. 266-271 (2004)
9. Cetin, O., Akar, F., Ozcerit, A. T., Cakiroglu, M., Bayilmis, C.: A blind steganography method based on histograms on video files. The Imaging Science Journal, 60(2), 75-82 (2012)
10. Aoki N.: A packet loss concealment technique for VoIP using steganography. In Proc. Int. Symp. Intelligent Signal Processing and Communication Systems (ISPACS), pp. 470-473 (2003)
11. Miao R., Huang Y. F.: An approach of covert communication based on the adaptive steganography scheme on Voice over IP. In Proc. IEEE Int Conf. Communications (ICC), pp. 1-5 (2011)
12. Wu Z., Yang W.: G.711-based adaptive speech information hiding approach. In Proc. Int. Conf. Intelligent Computing (ICIC), pp. 1139-1144 (2006)
13. Ito, Abe S., Suzuki Y.: Information hiding for G.711 speech based on substitution of Least Significant Bits and estimation of tolerable distortion. IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences, E93-A, 1279-1286 (2010)
14. Tian H., Zhou K., Huang Y., Feng D., Liu J.: A covert communication model based on least significant bits steganography in voice over IP. In Young Computer Scientists (ICYCS 2008) The 9th International Conference for IEEE, pp. 647-652 (2008)
15. Kratzer C., Dittmann J., Vogel T., Hillert R.: Design and evaluation of steganography for Voice-over-IP. In Proc. IEEE Int. Symp. Circuits and Systems, Kos, Greece, pp. 2397-2340 (2006)
16. Xu T., Yang Z.: Simple and effective speech steganography in G.723.1 low-rate codes. In Proc. Int Conf. Wireless communications & Signal Processing (WCSP), Nanjing, China, pp. 1-4 (2009)
17. Mazurczyk W., Szczypiorski K.: Steganography of VoIP streams. Lecture Notes in Computer Science (5332), 1001-1018 (2008)
18. Abdullaziz O. I., Goh V. T., Ling H. C., Wong K.: AIPISteg: An active IP identification based steganographic method. Journal of Network and Computer Applications (63), 150-158 (2016)

19. Mazurczyk W.: Lost Audio Packets Steganography: the first practical evaluation. Security and Communication Networks, 5 (12), 1394-1403 (2012)
20. Aoki N.: A semi-lossless steganography technique for G.711 telephony speech. In Proc. 6th Int. Conf. Intelligent Information Hiding and Multimedia Signal (IIH-MSP), pp. 15-17 (2010)
21. Zhou K., Liu J., Tian H., Li C. H.: State-based steganography in low bit rate speech. In Proc. 20th ACM Int. Conf. Multimedia, pp. 1109-1112 (2012)
22. Huang Y. F., Tang S., Yuan J.: Steganography in inactive frames of VoIP streams encoded by source codec. IEEE Transactions on Information Forensics and Security, 6 (2), 296-306 (2011)
23. Huang Y., Liu C., Tang S., Bai S.: Steganography integration into a low-bit rate speech codec. IEEE Transactions on Information Forensics and Security, 7 (6), 1865-1875 (2012)
24. Tian, H., Zhou, K., Jiang, H., Liu, J., Huang, Y., & Feng, D.: An M-sequence based steganography model for voice over IP. In Communications (ICC09) IEEE International Conference on, pp. 1-5 (2009)
25. Tian H., Jiang H., Zhou K., Feng D.: Adaptive partial-matching steganography for voice over IP using triple M sequences. Computer Communications (34), 2236-2247 (2011)
26. Chang P. C., Yu H. M.: Dither-like data hiding in multistage vector quantization of MELP and G.729 speech coding. Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers (2), pp. 1199-1203 (2002)
27. Wu Z. J., Cao H. J., Li D. Z.: An approach of steganography in G.729 bitstream based on matrix coding and interleaving. Chinese Journal of Electronics, 24 (1), 157-165 (2015)
28. Rahman A., Amritha P. P.: Using Signature for Hidden Communication Prevention in IP Telephony. In Artificial Intelligence and Evolutionary Algorithms in Engineering Systems, Springer India, pp. 489-494 (2015)
29. Mazurczyk W., Szaga P., Szczypiorski K.: Using transcoding for hidden communication in IP telephony. Multimedia Tools and Applications, 70 (3), 2139-2156 (2014)
30. Thenmozhi S, Chandrasekaran M: A novel technique for image steganography using nonlinear chaotic map. Intelligent systems and control (ISCO), 2013 7th international conference on. IEEE, pp. 307-311(2013)
31. Anees A., Siddiqui A. M., Ahmed J., Hussain, I.: A technique for digital steganography using chaotic maps. Nonlinear Dynamics, 75(4), 807-816 (2014)
32. 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)
33. Neter J., Wasserman W., Whitmore G. A., Applied Statistics, 4th edition, Simon & Schuster, Inc., pp. 435-450 (1993)

Yijing Jiang received the Ph.D. degree and the BSc degree in information security from China University of Geosciences.

Dr Jiang is with China University of Geosciences (CUG), 388 Lumo Road, Wuhan 430074, P. R. China. Her research interests include covert communications, VoIP and steganography. Dr Jiang has published four research papers and participated in three externally funded research projects as key research members.

Shanyu Tang (A'08–M'08–SM'10) received the Ph.D. degree from Imperial College London, United Kingdom in 1995.

Dr Shanyu Tang is currently Professor in Cybersecurity at the University of West London. He was Distinguished Professor in Information Security in the School of Computer Science at China University of Geosciences from 2012-2017. Professor Tang is dedicated to adventurous research in fractal computing methods for covert communications, multimedia security, and bio-informatics. He is the principal grant holder of eight externally funded research projects including three grants from the UK government. He has contributed to 96 scientific publications — 57 refereed journal papers including IEEE/ACM TRANSACTIONS and IET journal papers, and held one patent.