Covert communication over VoIP streaming media with dynamic key distribution and authentication

Covert Communication over VoIP Streaming Media with Dynamic Key Distribution and Authentication

Jinghui Peng, Shanyu Tang

University of West London, London W5 5RF, UK

***Abstract*—Voice over Internet Protocol (VoIP) is widely embedded into commercial and industrial applications. VoIP streams can be used as innocuous cover objects to hide secret data in steganographic systems. The security offered by VoIP signalling protocols is likely to be compromised due to a sharp increase in computing power. This paper describes a theoretical and experimental investigation of covert steganographic communications over VoIP streaming media. A new information theoretical model of secure covert VoIP communications was constructed to depict the security scenarios in steganographic systems against passive attacks. A one way accumulation-based steganographic algorithm was devised to integrate dynamic key updating and exchange with data embedding and extraction, so as to protect steganographic systems from adversary attacks. Theoretical analysis of steganographic security using information theory proves that the proposed model for covert VoIP communications is secure against a passive adversary. The effectiveness of the steganographic algorithm for covert VoIP communications was examined by means of performance and robustness measurements. The results reveal that the algorithm has no or little impact on real-time VoIP communications in terms of imperceptibility, speech quality and signal distortion, and is more secure and effective at improving the security of covert VoIP communications than other related algorithms with comparable data embedding rates.**

***Index Terms*—Authentication, covert communication, key distribution, steganography, VoIP.**

I. INTRODUCTION

THE past decade has witnessed the rapid development of embedded Voice over Internet Protocol (VoIP) for commercial and

industrial applications. Widening access to the Internet greatly facilitates the use of multimedia applications in people's daily lives. Evolving network technology such as streaming has enjoyed a rise in popularity. However, security measures are struggling to keep up with the pace of change in attack tactics.

Encryption and decryption technologies are normally used to address data security and privacy issues. There are symmetric encryption and public-key encryption that enable the translation of a plaintext message into ciphertext. However, an increase in computing power has led to decryption of several encryption algorithms, such as MD5 [1], DES [2] and SHA-1 [3], indicating possible vulnerabilities in the encryption primitives. It is generally recognised that encrypted messages are obvious, and when intercepted, it is clear that the communicating parties are communicating secretly.

As a sub-branch, digital steganography is defined as 'the art of concealed communication by hiding messages in seemingly innocuous objects' and 'the very existence of a steganographic message is secret' [4]. Steganography in static cover objects, such as text, BMP or JPEG images, and WAV or MP3 audio files, has been explored extensively [5]-[7]. Network protocols and streaming media [8], such as VoIP, are also used to realise covert steganographic communications.

There has been a large body of research into steganographic algorithms for covert communications over streaming media, but the key distribution problem in covert steganographic communications has been sidestepped. In fact, the successfulness of steganographic algorithms for covert communications relies largely on the transmission of secret keys between the communicating parties. Security in transmission of secret keys is more crucial for covert VoIP communications because of the timing and loss of packets, i.e. covert VoIP communications require continuous embedding and necessary synchronization between the communicating parties. So far there are no reliable and secure key transmission schemes that could be put into use for covert VoIP communications. Thus, secure key transmission for covert steganographic communications is worth studying apart from designing effective steganographic algorithms for them.

The main purpose of this study is to explore the potential of one way accumulation-based dynamic key updating and transmission for innovative applications in the field of covert steganographic communications over streaming media. The findings from the study make several contributions to the current literature as follows:

1) A novel information theoretical model of steganographic VoIP communication is constructed to realise secure covert VoIP communications, achieving high data embedding capacities comparable to other related algorithms.
2) A new dynamic steganographic algorithm is devised for covert VoIP communications. It includes one way accumulation integrating into dynamic key updating and exchange, which can protect steganographic systems from man-in-the-middle attacks, which threaten covert steganographic communications.

The rest of this paper is organised as follows: Section II describes the related work. A novel theoretical model of steganographic VoIP communication is presented in Section III. Section IV details a new steganographic algorithm for covert VoIP communications. Security analysis of the new algorithm is discussed in Section V. In Section VI, experiments including performance and security measurements are depicted, and the results are discussed in detail. The paper is concluded in Section VII.

II. Related Work

A great deal of research has been conducted on algorithm design and cover object selection for covert steganographic communications over streaming media, but little effort has been made to explore the potential of using dynamic key distribution to improve the security of steganographic systems.

Dittmann et al. first studied VoIP steganography and decryption techniques and suggested their algorithm [9]. Aoki developed a lossless steganographic technique for G.711 telephony speech [10], with the embedding capacity depending on the number of

‘0’ in audio signals, so the practical application was limited. Liu et al. analysed the parameters of G.729 coded speech frames to identify the parameters and effective bits of G.729 speech coding, which were used for steganography [11]. Yu et al. designed a VoIP steganographic scheme [12], but its validity needed to be confirmed. Aoki proposed a semi-lossless steganographic technique for G.711 telephony speech [13], with bandwidth improved from 24 bit/s to 400 bit/s, depending on the background noise signal level. Huang et al. devised a high capacity steganographic algorithm for embedding data in various speech parameters of the inactive frames of low bit rate audio streams encoded with G.723.1 source codec [14]. Tian et al. [15] put forward a method to improve the performance of steganography by adding some similarity between the hidden message and the cover object to strike a balance between steganography transparency and bandwidth, but the similarity limited the choice of hidden messages. Gope et al. [16] presented an authentication protocol for wireless sensors networks over which streaming media are transmitted; the protocol provided various imperative security properties such as user anonymity, untraceability, forward/backward secrecy, and perfect forward secrecy. Tian et al. improved the security of quantization-index-modulation steganography in low bit-rate speech streams [17].

In 2016, Qi et al. used Discrete Spring Transform to eliminate redundancy in multimedia signals and improve speech quality [18]. Liu et al. reported the use of a matrix embedding method to achieve steganography in linear predictive coding for low bit-rate speech codec [19]. Janicki investigated pitch-based steganography using Speex voice codec [20] to complement Aoki’s work [13]. More recently, Tian et al. [21] suggested a bitrate modulation steganographic algorithm with Hamming matrix encoding, but its practicality needed further study. Xin et al. proposed an adaptive audio steganographic algorithm for covert wireless communication, which was based on variable low bit coding [22]. Zhang et al. [23] suggested a F5 and simplified wet paper code (SWPC) based algorithm for VoIP steganography, but the tradeoff between the embedding rate and the embedding efficiency was not discussed. Overall, previous steganography studies mainly focused on steganographic algorithm design.

Public key cryptography provides solutions to the key distribution problem. A public key scheme such as RSA involves a public key and a private key; once the communicating parties compute the shared secret key they can use it as an encryption/decryption key [24]. The two keys are related due to inverse operations, so there must be no easily computational method of deriving the private key from the public key. Public key schemes by themselves do not provide authentication of the communicating parties and are thus particularly vulnerable to man-in-the-middle attacks.

Key regression, time-evolving and multicast key distribution schemes were suggested for key management in cryptographic storage systems [25]. They provided a means of deriving a sequence of temporally related keys from the most recent key. The effectiveness of these schemes for covert steganographic communications over streaming media is unknown since they are unable to address some key issues such as packet loss and synchronisation.

A number of authors have investigated key distribution schemes for image steganography. Dagar proposed an image steganographic algorithm that used two secret keys to randomise the bit hiding process and enhance the security of hidden messages [26]. Gutiérrez-Cárdenas suggested a PRNG key distribution scheme for image steganography, which used a picture to conceal a message with unaltered pixel information, so it could be secure against steganalysis detection [27]. Patel et al. reported LSB-based image steganography using dynamic key cryptography in which the dynamic feature of the key was enabled by rotating the key and each key rotation produced a new key [28]. Anwar et al. [29] used the Lifting Wavelet Transform (LWT) and dynamic key techniques to perform steganography in audio files, but the practicality needed further study. However, these key distribution schemes designed for image steganography cannot be used directly in covert steganographic communications over VoIP streams due to the timing and loss of packets.

Although the above investigations examined key distribution schemes for image steganography, only few references in the literature described key distribution schemes for VoIP steganography. This was the motivation behind the present study.

In comparison with existing steganographic algorithms for VoIP steganography, such as FIPlP [30], CNV [31], MELP [32] [33], parameter-LSB [21] and HiF [34], the proposed algorithm is more secure for taking into account key distribution and authentication, and is more effective in terms of data embedding capacity and imperceptibility.

III. Theoretical Model of Steganographic VoIP Communication

A new theoretical model is devised for covert steganographic communications over VoIP streaming media in this study. The model is based on steganography and cryptography and is depicted in Fig. 1. A secret message to be hidden (M) is encrypted with a secret key generated from a random number generator to form an encrypted message; the message is segmented into distinct parts which are embedded in a series of packets of media streams, namely cover objects (C). S in the figure denotes the packet containing a hidden message.

This model integrates dynamic key distribution and authentication with data embedding and extraction. Three sequences in the model simulate the continuity of data embedding, time-variant size of hidden message in each packet and key pairs respectively, taking into account continuous data embedding and necessary synchronisation of sender and receiver due to packet loss.

The random sequence in Fig. 1, $A = \{a_1, a_2, \ldots a_n\}$, is a group of zeros, ones, twos and threes, e.g. {1,2,2,3,0,…}, describing the continuity of data embedding. The ones, twos and threes denote a packet containing the beginning, the continuation or the end of the hidden message, respectively, and zeros mean a packet does not contain the hidden message.

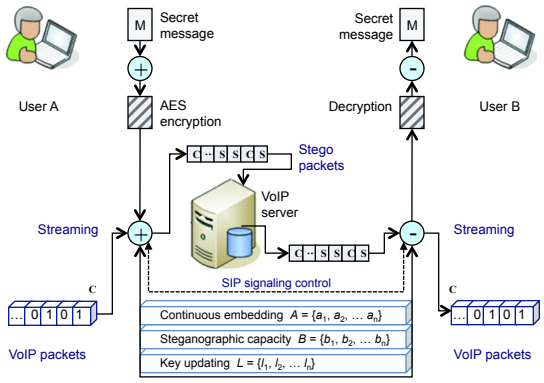

Fig. 1. Model of covert VoIP communications.

The sequence, $B = \{b_1, b_2, \ldots b_n\}$, is a set of steganographic capacity, corresponding to varying numbers of bits of the hidden message embedded in a series of streaming media packets. This sequence enables the receiver to determine the size of the hidden message embedded in each packet.

The sequence, $L = \{l_1, l_2, \ldots l_n\}$, represents a set of private / public key pairs of 1024 bits each. The public and private keys are correlated in the public-key scheme where the sender and the receiver compute the shared key using knowledge of the public key based on discrete exponential and logarithm (hash) functions.

According to information theory, Kullback–Leibler (KL) divergence is used as a measure of security for steganographic systems [5].

The statistical distance (ε) between the cover object and the stego object can be expressed as $\varepsilon = \sum_{w \in W_0} P_C(w) - \sum_{w \notin W_0} P_C(w)$ and $W_0 \subset W$, where P_C is the probability distribution of the cover object, w is the measurement, W_0 is the plausible space, and W is the total space of possible measurements.

The total probability distribution of the secret message sent over the space W is given by

$$P_U(w) = P(w_i \in W_2)P_S(w) + P(w_i \in W_3)P_C(w) \qquad (1)$$

where P_S is the probability distribution of the stego object, and W_2 and W_3 are the observation spaces relating to the stego object and the cover object, respectively. Equation (1) becomes $P_U(w) = \eta P_S(w) + (1-\eta) P_C(w)$, where η is the probability that '1' appears in a period. As

$$P_S(w) = \begin{cases} P_C(w)/1+\varepsilon, & w \in W_0 \\ P_C(w)/1-\varepsilon, & w \in W_1 \end{cases} \qquad (2)$$

where W_0 and W_1 are the plausible spaces of possible measurements, then the relative entropy between the cover object and the stego object for covert steganographic communications is given by

$$D(P_C \,\|\, P_U) \le \frac{1+\varepsilon}{2} \frac{\varepsilon^2 \eta^2}{1-\varepsilon^2(1-\eta)^2} \qquad (3)$$

As Fig. 1 shows, there are two N-level true random sequences (A and B) used to model the dynamic and synchronisation characteristics of covert communications, $\eta = 2^{n-1}\,2^{n-1}/(2^n - 1)(2^n - 1)$, then equation (3) becomes

$$D(P_C \,\|\, P_U) \le \varepsilon^2 \big/ (16 - 9\varepsilon^2) \le \varepsilon \qquad (4)$$

Hence, it proves the proposed model for covert communications over VoIP streaming media is secure against a passive adversary.

IV. Steganographic Algorithm for Covert VoIP Communication

In this study, a novel steganographic algorithm is devised to integrate one way accumulation-based dynamic key distribution with data embedding and extraction for covert VoIP communications.

A. Accumulation-Based Key Distribution

The key updating and transmission algorithm is schematically shown in Fig. 2 for an illustrational purpose. The algorithm includes recurrences of Receiver validation (Step 1), Key transmission (Step 2), and Key updating (Step 3).

Fig. 2. Schematic description of key updating and transmission.

A one way cryptographic accumulation function is used to validate the communicating party, Bob, as shown in Fig. 2. Assuming $T = \{x_1, \ldots, x_n\}$ be the set of items $x_1, \ldots, x_n$ stored by Alice (the communicating party), she selects secure primes p and q that are suitably large, and a suitably large base g that is relatively prime to a big composite number N:

$$N = pq \tag{5}$$

The values of g and N are then made available to the public, but the values p and q are kept secret. Moreover, Alice computes the following value

$$Z = g^{x_1 x_2 \cdots x_n} \bmod N \tag{6}$$

and a partial accumulated hash value Z_i for Bob (x_i), *i.e.* the accumulation of all the values in the set T besides x_i.

$$Z_i = g^{x_1 x_2 \cdots x_{i-1} x_{i+1} \cdots x_n} \bmod N \tag{7}$$

Having computed Z_i, Alice sends Z_i and N to Bob, as well as the signed pair (Z, t) where t is the current timestamp. Bob determines whether t is current and (Z, t) is indeed signed by Alice. Bob then computes

$$Z^b = Z_i^{x_i} \bmod N \tag{8}$$

and returns the value of Z^b to Alice. Alice compares Z^b with Z that is calculated using equation (6). If the two agree, then Alice knows that the sender is Bob in possession of x_i. Indeed, it is generally accepted to be computational infeasible for someone who does not know the values of p and q to compute a value Q such that

$$Z = Q^{x_i} \bmod N \tag{9}$$

when $x_i \notin T$, indicating that the one way accumulation function is secure in terms of cryptography.

After the receiver validation step, the private key used for encryption and decryption of a secret message is transmitted securely under Integer Decisional Diffie-Hellman assumption. Alice and Bob agree to use a prime number p and base g (g is assumed to be known by adversaries). Alice selects a secret integer α, calculates

$$U = g^{\alpha} \bmod p \tag{10}$$

and sends the value of U to Bob. Bob chooses a secret integer β, computes

$$V = g^{\beta} \bmod p \tag{11}$$

and sends the value of V to Alice. Alice calculates

$$(g^{\beta} \bmod p)^{\alpha} \bmod p$$

and Bob computes

$$(g^{\alpha} \bmod p)^{\beta} \bmod p$$

The values of $(g^{\beta} \bmod p)^{\alpha} \bmod p$ and $(g^{\alpha} \bmod p)^{\beta} \bmod p$ are the same because groups are power associative according to mathematical principles [24]. So both Alice and Bob are now in possession of the group element $g^{\alpha\beta} \bmod p$, which can serve as the shared secret key (K) to encrypt and decrypt the secret message:

$$K = (g^{\beta} \bmod p)^{\alpha} = (g^{\alpha} \bmod p)^{\beta} \qquad (12)$$

Key updating is the process of determining whether the stego packets received by Bob contain the complete hidden message. As Fig. 1 shows, the random sequence A is used to identify whether a packet contains the beginning, the continuation or the end of the hidden message. Using the sequence, Bob knows whether the secret message he decrypts with the shared key is complete or not.

To enable key exchange and transmission successfully in case of heavy packet loss, which is common in VoIP communications, the proposed algorithm contains a special re-distribution function as follows:

If a packet containing part of the secret message is lost, Bob sends Alice the value of Z^b again to initiate a repetition of Steps 1, 2 and 3, as shown in Fig. 2, *i.e.* Alice validates Bob as the 'legal' receiver (Step 1), and embeds the same secret message again (Step 2) until Bob receives all the packets used to embed the entire secret message (Step 3). Thus, this function can help achieve synchronization between the sender and the receiver, thereby eliminating the effect of packet loss on key exchange in covert steganographic communications over streaming media.

B. Data Embedding

Random number keys are used to encrypt the secret message to be hidden. The encrypted message is subsequently segmented into distinct parts, which are then embedded in a series of packets of VoIP streams at different data embedding capacities and various data embedding locations. For simplicity, 16 bytes of VoIP streams have a data embedding capacity of two bytes. If the data embedding interval (R) is set to one, the secret message is randomly embedded in VoIP streams at an interval of two bytes, *i.e.* one byte of the secret message is randomly embedded in 16 bytes of VoIP streams. When the interval is set to two, the secret message is randomly embedded in VoIP streams at an interval of three bytes, so in this case 22 bytes of VoIP streams contain one byte of the secret message, and so on.

The process of embedding the secret message in VoIP streams is designed as follows:

Step A: first embed the secret message length (LoM) in VoIP streams, and set length to LoM.

Step B: embed the secret message into the first packet. Compute the length of the secret message embedded in the first packet (m_1) and the length of the secret message embedded in other packets (m_k).

Procedure DE First_packet

---

if (LoM < m_1)
  then (
    encrypt $M(0, m_0 - 1)$ to form $E(0, m_0 - 1)$
    embed $E(0, m_0 - 1)$ in the bit stream $BIT = \{bit(0), bit(1), \ldots, bit((m_0 - 1) * 8)\}$
    if (bit(i) == 0)
      then (V(k) ← V(k)&0xfe)

else (V(k) ← V(k)|0x01)

k ← k + R

length ← 0

end

)

else (

encrypt $M(0, m_1 - 1)$ to form $E(0, m_1 - 1)$

embed $E(0, m_1 - 1)$ in the bit stream $BIT = \{bit(0), bit(1), \ldots, bit((m_1 - 1) * 8)\}$

if (bit(i) == 0)

then (V(k) ← V(k)&0xfe)

else (V(k) ← V(k)|0x01)

k ← k + R

length ← length - m_1

)

---

where m_0 is the total size of the secret message to be hidden, and k is the sequence number.

Step C: embed the secret message into the other packets.

Procedure DE Other_packets

---

while (length > m_k) **do**

(

encrypt $M(m_1, m_1 + m_k - 1)$ to form $E(m_1, m_1 + m_k - 1)$,

embed $E(m_1, m_1 + m_k - 1)$ in the bit stream $BIT = \{bit(0), bit(1), \ldots, bit((m_k - 1) * 8)\}$

if (bit(i) == 0)

then (V(k) ← V(k)&0xfe)

else (V(k) ← V(k)|0x01)

k ← k + R

length ← length - m_k

)

Step D: compute the length of the secret message embedded in the last packet (m_n)

Procedure DE Last_packet

---

encrypt M(LoM− length, LoM− 1) to form E(LoM− length, LoM− length + m_n − 1)

embed E(LoM−length, LoM−length+ m_n −1) into $BIT = \{bit(0), bit(1), \ldots, bit((m_n - 1) * 8)\}$

if (bit(i) == 0)

then (V(k) ← V(k)&0xfe)

else (V(k) ← V(k)|0x01)

k ← k + R

length ← 0

end

For example, the first 16 bytes of the first VoIP packet is used to embed the length of the secret message to be hidden, and the remaining of the first packet and the other VoIP packets are used to embed the secret message itself. The size of the remaining of the first packet is about 4080 bytes, with a data embedding capacity up to 510 bytes (12.5%). To reduce the encryption time, the first 496 bytes of the secret message are embedded in the first VoIP packet. AES is used to encrypt the first 496 bytes of the secret message, and the resulting ciphertext is then embedded in VoIP packets using the data embedding algorithm above. As for the other VoIP packets, 512 bytes of the secret message (encrypted with AES) are embedded in each VoIP packet. The remaining of the secret message is embedded in the last VoIP packet with the size of LLoM. The LLoM value may not be a multiple of 16 bytes, so it is possibly necessary to adjust it to a multiple of 16 bytes in some cases.

C. Data Extraction

The extraction of the hidden secret message, steganographically embedded in VoIP streams using the data embedding algorithm above, from the stego VoIP streams is the inverse process of the data embedding algorithm described in the previous section. The corresponding extraction algorithm is used to retrieve the secret message encrypted with AES, and decrypt it with the same secret keys to obtain the original secret message from stego VoIP packets.

V. Security Analysis

This section theoretically examines the security of the new steganographic algorithm for covert VoIP communications, and shows how the algorithm can resist possible adversary attacks, which threaten existing VoIP steganographic algorithms.

A. Authentication for Communicating Parties

Covert channels based on one way accumulation are used to conduct key updating and transmission for covert steganographic communications over VoIP streams.

The general form of a cryptographic accumulator can be defined as follows: first start with a 'seed' value y_0, denoting the empty set, then define the accumulation value incrementally from y_0 for a set of elements $T = \{x_1, \ldots, x_n\}$, so that $y_i = f(y_{i-1}, x_i)$, where f is a one way function whose final value does not depend on the order of the x_i's [35]. So a source can digitally sign the value of y_n in order to enable a third party to produce a short proof for any element x_i belonging to T – namely, swap x_i with x_n and recompute y_{n-1} from scratch – the pair (x_i, y_{n-1}) is a cryptographically-secure assertion for the membership of x_i in the set T.

A new key distribution algorithm is devised in this study to offer secure key updating and transmission for covert steganographic communications over streaming media, i.e. a one way cryptographic accumulator along with Diffie-Hellman key exchange are integrated with data embedding during steganography. The use of the accumulator provides cryptographic authentication for the communicating parties, thereby preventing covert steganographic communications from adversary attacks.

B. Man-in-the-Middle Attacks

The integration of a new one way cryptographic accumulator and Diffie-Hellman based key exchange is used to provide secure key exchange for covert steganographic communications over streaming media. The algorithm can ensure secure key updating and transmission, and then protect steganographic systems from adversary attacks.

Dynamic key distribution in the algorithm means the keys are almost unlikely to be compromised, because it enables steganographic systems to continually and randomly generate new private keys that the communicating parties share automatically. As the private key is changed continuously, a compromised key in the system could only decrypt a small amount of encoded information using today's supercomputers.

The man-in-the-middle attack is defined as a form of active eavesdropping in which an attacker makes independent connections with the communicating parties and relays messages between them, making the communicating parties believe that they are talking directly to each other over a connection; in fact the entire conversation is controlled by the attacker [8].

In the new steganographic algorithm, a one way cryptography accumulator is used to conduct authentication between the communicating parties to prevent possible man-in-middle attacks. As described in Section IV, Alice authenticates Bob (*i.e.* Bob in possession of a valid x_i) by determining whether Z^b (computed using equation (8) and sent to Alice by Bob) is equal to Z (computed by Alice using equation (6)). If the third party John wants to launch a man-in-the-middle attack, he has to pass the verification process first. Obviously without the knowledge of the two primes of p and q, John cannot pass the verification process of the receiver. In addition, as John cannot guess the high entropy element x_i correctly, he cannot work out the value of Z^b equal to Z. Thus, John cannot launch the man-in-the-middle attack successfully to cheat Alice, indicating that the proposed steganographic algorithm can prevent man-in-the-middle attacks.

VI. Results and Discussion

This section summarises the findings, interprets what the steganographic communications results mean, and explains the significance of the results.

VoIP experiments were used to evaluate the performance and security of the proposed dynamic key distribution-based steganographic algorithm for covert communications over streaming media. Performance measurements were carried out using Digital Speech Level Analyser (DSLA) [36], as shown in Fig. 3, which determines ITU-T P.862 objective speech quality scoring plus improved Mean Opinion Score (MOS) prediction according to ITU-T P.862.1 [37]. The experiments were repeated 12 times to assess the effectiveness and security of the new steganographic algorithm.

In the experiments, VoIP streams with PCM format encoded with G.711 codec were used as cover objects for covert steganographic communications over streaming media. The secret message to be hidden was encrypted with random number keys, and the encrypted message was divided into segments that were then embedded in a series of packets of VoIP streams. The imperceptibility of the resulting stego VoIP streams was evaluated, and the data embedding capacity was computed accordingly for each set of experiments.

PESQ score is calculated according to P.862 and PESQ P.862.1 gives a quality score on a MOS-like scale for narrowband listening. The aim of the amended recommendation ITU-T P.862.1 is to provide a single mapping from the raw P.862 score to the Listening Quality Objective Mean Opinion Score (MOS-LQO). The mapping from PESQ score to PESQ P.862.1 is performed as follows:

$$PESQ\text{P.862.1} = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.4945 \times PESQScore + 4.6607}} \qquad (13)$$

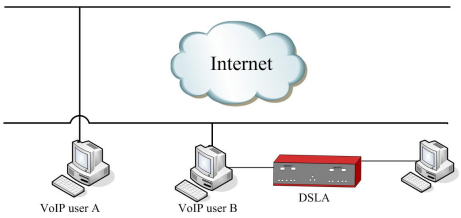

Fig. 3. Diagram of measurements for covert VoIP communications.

A. Imperceptibility

The essential security of covert VoIP communication is that it would not cause any suspicion from attackers. Experiments were carried out to determine the degree of imperceptibility of steganographic systems using the proposed algorithm.

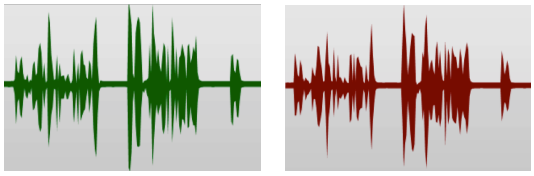

Fig. 4. Waveform screenshots of the original and stego VoIP streams (x-axis represents time, y-axis represents amplitude).

Figure 4 shows the waveform screenshots of the original VoIP streams (left image) and the stego VoIP streams that contain a secret message (right image), respectively. As Fig. 4 shows, there was a remarkable resemblance between the two waveforms. Listening tests indicated that a human perceptual system could not distinguish the differences between the original VoIP streams and the stego VoIP streams with the hidden message. The results show very little distortion occurred in the time domain as a result of steganography in VoIP streams.

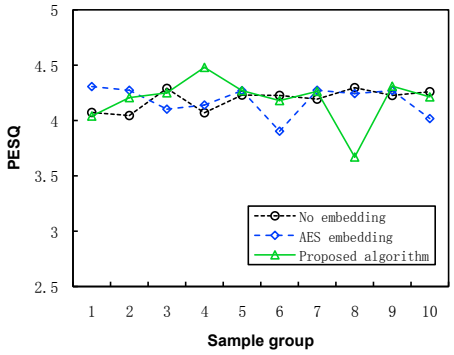

Fig. 5. PESQ values for the original VoIP streams and the stego VoIP streams with AES steganography or the proposed algorithm.

Figure 5 shows a comparison of the mean PESQ values for the original VoIP streams, the stego VoIP streams with AES embedding-based steganography, and the stego VoIP streams using the proposed steganographic algorithm. The number of sample group is shown on the horizontal axis. For each sample group, 12 repeated experiments were carried out to yield the mean PESQ value. The multiple-line graph demonstrates that the PESQ values were reasonably stable except for Testing 8 in which the PESQ was close to 3.5 (the lower boundary of good speech quality). This is in line with the expectation that the new steganographic algorithm would cause no or little degradation in speech quality whilst improving the security of steganographic systems by means of dynamic key distribution, indicative of secure and robust covert VoIP communications.

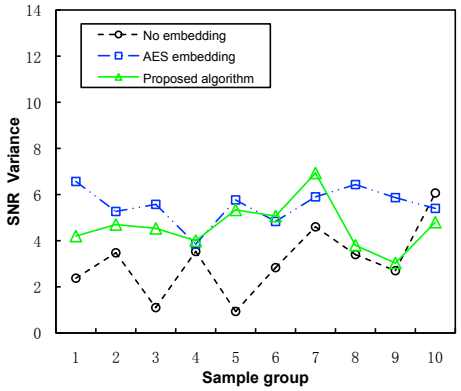

Fig. 6. SNR values for the original VoIP streams and the stego VoIP streams with AES steganography or the proposed algorithm.

Figure 6 shows a comparison of the mean SNR values between the original VoIP streams, the stego VoIP streams with AES embedding-based steganography, and the stego VoIP streams using the proposed steganographic algorithm. The SNR variance of the original streams was estimated to be 3.1, that of the stego streams with AES steganography was around 5.5, and that of the stego streams using the new algorithm was around 4.6. These results indicate that covert VoIP communications using the new steganographic algorithm have much better imperceptibility with a greater level of security than AES steganography.

Overall, the results above indicate that the proposed steganographic algorithm has no or little impact on real-time VoIP communications in terms of speech quality, signal distortion and imperceptibility. The differences in PESQ and SNR between the original and stego VoIP streams were so minor that distortion resulted from covert communications using the new algorithm was imperceptible, indicating that the proposed steganographic algorithm is effective at breaking through the key exchange bottleneck occurs in covert steganographic communications over streaming media, and protecting steganographic systems from man-in-the-middle attacks.

B. Effects of Data Embedding Intervals

The secret message to be hidden is embedded in a series of packets of VoIP streams at various data embedding intervals, so as to study the effects of the increased complexity of the proposed steganographic algorithm on covert steganographic communications over streaming media.

Figure 7 shows changes in the mean PESQ values of the stego VoIP streams that contain the hidden message encrypted with AES before embedding at different data embedding interval distances in streaming media. As Fig. 7 shows, the average PESQ values of the stego streams decreased slightly before the data embedding interval distance reached 3, and then increased as the interval distance increased, showing an upward trend gradually close to the mean PESQ value of the VoIP streams without data embedding (4.31). The results indicate the advantage of the increased complexity of the proposed steganographic algorithm.

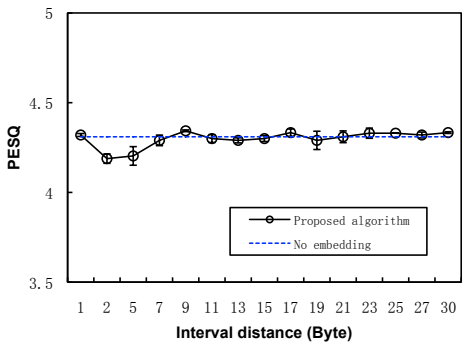

Fig. 7. PESQ values of the stego VoIP streams at various interval distances.

C. Effects of Hidden Message Size

To study the effects of the size of the hidden message on the data embedding capacity of the proposed steganographic algorithm, a series of covert VoIP communications experiments were carried out.

Figure 8 shows changes in the average PESQ values of the stego VoIP streams taken from covert VoIP communications using the new steganographic algorithm at different sizes of the hidden message. Each data point is the mean value based on 12 repeated experiments. As Fig. 8 shows, the average PESQ values of the stego streams decreased with the hidden message size increasing. The average PESQ values of the stego VoIP streams were still greater than 3.5, the lower threshold of covert VoIP communications for the codec used in the experiments, before the hidden message size reached 1186 bytes, which can be regarded as the maximum data embedding capacity. When the size of hidden message exceeds the embedding capacity of the covert object, speech quality would decrease seriously, leading to unavailability of real-time covert communication.

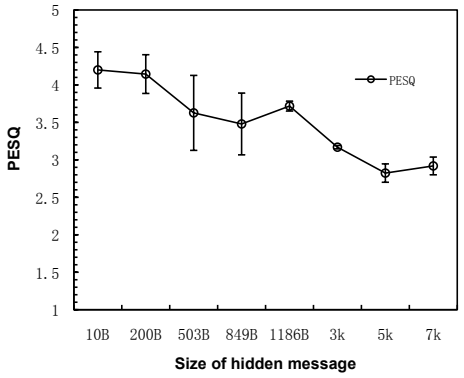

Fig. 8. PESQ values of the stego VoIP streams varying with the hidden message size.

D. Statistical Undetectability Analysis

Steganographic communications aim to conceal the existence of hidden messages from both human perceptual systems and computational detection. Statistical undetectability is normally used to evaluate the security of a steganographic system [5]. A secure steganographic system should be statistically undetectable.

In order to evaluate the security of the proposed steganographic algorithm, t-test was used in this study to perform statistical undetectability analysis. The t-test is a statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis [38]. A two-sample t-test is used when it can be assumed that two distributions have the same variance. The *t* statistic used to test whether the difference between the two samples is significant can be calculated as follows:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{14}$$

where S_1^2 and S_2^2 are the variances of the samples; n_1 and n_2 are the sample sizes.

The default alpha of 0.05 is normally used as the threshold. When the calculated *P*-value is less than the threshold, there is a significant difference between the two samples.

TABLE I
UNDETECTABILITY ANALYSIS RESULTS USING T-TEST

	Original stream samples	Stego stream samples
Mean	-3.25151E-07	5.30214E-06

The original and stego VoIP stream samples were tested in our experiments. Among the variables that appear in Table I, the *P*-value is 0.970405697, which is greater than 0.05, indicating no significant difference between the tested samples. The results show that the covert communication system using the proposed steganographic algorithm is secure in terms of statistical undetectability analysis.

E. Comparisons with Other Related Algorithms

To confirm the effectiveness of the proposed steganographic algorithm, comparisons of the data embedding capacity, number of communication passes, message size for authentication, collision resistance, computational overhead and bandwidth between the proposed algorithm and other related algorithms were conducted for covert steganographic communications over VoIP streams.

Some steganographic algorithms, such as FIPIP [30], CNV [31], MELP [32] [33], parameter-LSB [21], and HiF [34], have been suggested for VoIP steganography. These algorithms achieved different levels of data embedding in streaming media. For comparison purposes, these existing algorithms and the proposed steganographic algorithm were used in the experiments to steganographically embed the secret message in VoIP streams, respectively.

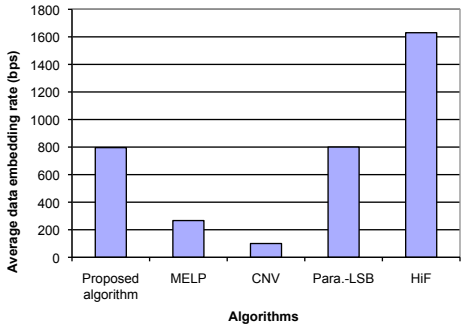

Fig. 9. Comparison of data embedding rates between the proposed algorithm and other related algorithms.

Figure 9 shows a comparison of the average data embedding capacity between the proposed steganographic algorithm and other four steganographic algorithms. For each algorithm, 12 repeated experiments on covert VoIP communications were carried out to determine the average data embedding rate. As Fig. 9 shows, the average data embedding rate of the proposed algorithm was much higher than those of the MELP and CNV algorithms, approximately equal to that of the parameter-LSB algorithm, and lagged behind the HiF algorithm. The average data embedding rate of the proposed algorithm was lower than that of the HiF algorithm, which is possible due to the fact that the HiF algorithm made good use of redundancy in the inactive frames of VoIP streams; however, the proposed algorithm uses various data embedding intervals in VoIP streams to steganographically embed the secret message, thereby achieving much higher level of security than the HiF algorithm. The data embedding results suggest that the proposed steganographic algorithm has great effectiveness in terms of the average data embedding rate, which is sometimes one of the important factors in designing a steganographic algorithm for covert steganographic communications over streaming media.

TABLE II

COMPARISON OF THE NUMBER OF REQUIRED COMMUNICATION PASSES

Algorithm	Number of required passes
Proposed	10
FIPlP	5
HiF	5
CNV	N/A
MELP	4
parameter-LSB	N/A

Table II shows a comparison of the number of the required communication passes between the proposed steganographic algorithm and other related algorithms. The 'N/A' means the value is not available to the public. As the table shows, the number of required communication passes of the proposed algorithm is 10 times, which is higher than FIPlP, HiF and MELP, providing cryptographic authentication for covert communication systems.

TABLE III

COMPARISON OF REQUIRED MESSAGE SIZE FOR AUTHENTICATION AND COLLISION RESISTANCE

Algorithm	Required message size	Collision resistance
Proposed	4096 bytes	$2^{1152/2}$
FIPlP	4096 bytes	$2^{192/2}$
HiF	N/A	$2^{128/2}$
CNV	N/A	N/A
MELP	N/A	$2^{128/2}$
parameter-LSB	N/A	N/A

Table III shows a comparison of the required message sizes for authentication and collision resistance between the proposed algorithm and other related algorithms. As for the proposed covert communication system and FIPlP, the size of audio data in each packet for authentication is 4096 bytes. The strength of steganographic systems against brute-force attacks depends on the block length for key construction and the key size. With a birthday attack (a typical cryptographic attack), it is possible to find a collision of an *n*-bit key in $2^{n/2}$. MELP used MD5 (128-bit) to produce a shorter hash value to calculate a checksum for the hidden message, with a collision of $2^{128/2}$. As Table III shows, a birthday attack on the proposed system produces a collision with a work factor of approximately $2^{1152/2}$, which is viewed as adequate to provide sufficient collision resistance as to today's computing power.

TABLE IV

COMPARISON OF COMPUTATIONAL OVERHEAD AND BANDWIDTH

Algorithm	Computational overhead (ms)	Bandwidth (kbits/s)
Proposed	88.45	0.80
FIPlP	N/A	0.50
HiF	N/A	0.44
CNV	240	0.10
MELP	2760	0.43
parameter-LSB	N/A	1.70

Table IV shows a comparison of computational overhead and bandwidth between the proposed algorithm and other related algorithms. The computational overhead of the proposed algorithm is 88.45 ms, which is acceptable for real-time covert VoIP communication. As with MELP, the computational overhead value is the average of those estimated at four sampling rates used. The parameter-LSB algorithm had the highest steganographic bandwidth due to the use of less secure, simple LSB, and the bandwidth values for other five algorithms were comparable. As can be seen from Table IV and Figs. 5 and 6, the proposed algorithm achieved a relatively larger steganographic bandwidth (0.80 kbits/s) with negligible signal distortion, which is one of the most important performance metrics that assess covert communication over streaming media.

As for covert VoIP communication, it is very unlikely that an adversary will be able to obtain many different copies of a given stego VoIP packet due to real-time, dynamic and streaming features; therefore, collusion attacks are of less or no concern. That means the covert VoIP communication system has incredible strength and great resilience considering collusion attacks. Security analysis of the new algorithm for covert VoIP communication is detailed in Section V.

VII. Conclusion

The purpose of the current study was to explore the potential of one way accumulation-based dynamic key distribution for innovative applications in the field of covert steganographic communications. The new steganographic algorithm devised for covert steganographic communications over VoIP streams is found to be secure against a passive adversary. The evidence from the study suggests that the algorithm can protect steganographic systems from adversary attacks such as man-in-the-middle attacks. Security analysis and experimental results show the effectiveness of the proposed algorithm with imperceptible distortion of the original signal (5% change in PESQ), a greater data embedding rate (~ 800 bps) and a larger steganographic bandwidth (0.80 kbits/s). This new algorithm provides better cryptographic authentication and security in terms of key distribution and collision resistance with a work factor of $2^{1152/2}$. The findings from this study add to a growing body of literature on steganography in streaming media.

Further research should therefore concentrate on investigation into the effectiveness of the steganographic algorithm for VoIP communications over a wide area network where heavy packet loss occurs.

Acknowledgment

The author would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

[1] X. Wang and H. Yu, "How to break MD5 and other hash functions," in Proc. Annual International Conference on the Theory and Applications of Cryptographic Techniques, Berlin, Heidelberg, 2005, pp. 19-35.

[2] EFF. Frequently Asked Questions (FAQ) About the "EFF DES Cracker" Machine, 2016.

[3] M. Stevens, et al. "The first collision for full SHA-1," in Proc. Annual International Cryptology Conference, Springer, Cham, 2017, pp. 570-596.

[4] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, 2nd ed., Burlington, 2008, pp. 4-13.

[5] J. Fridrich, *Steganography in digital media: principles, algorithms, and applications*, 2nd ed., Cambridge University Press, 2014, pp. 107-129.

[6] A. D. Ker and T. Pevný, "The steganographer is the outlier: Realistic large-scale steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 9, iss. 9, pp. 1424-1435, Sep. 2014.

[7] Z. Yang, X. Guo, Z. Chen, and Y. Huang, "RNN-Stega: Linguistic steganography based on recurrent neural networks," *IEEE Transactions on Information Forensics and Security*, vol. 99, no. 11, pp. 1-16, 2019.

[8] W. Stallings, *Cryptography and Network Security: Principles and Practice*, 7th ed., Pearson, London, 2017, pp. 252-279.

[9] J. Dittmann, D. Hesse, and R. Hillert, “Steganography and steganalysis in Voice-over IP scenarios: operational aspects and first experiences with a new steganalysis tool set,” in *Proc. SPIE 5681: Security, Steganography, and Watermarking of Multimedia Contents VII*, San Jose, USA, Mar. 2005, pp. 607-618.

[10] N. Aoki, “A technique of lossless steganography for G.711 telephony speech,” in *Proc. 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Harbin, China, 2008, pp. 608-611.

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

[12] Z. Yu, C. Thomborson, C. Wang, J. Fu, and J. Wang, “A security model for VoIP steganography,” in *Proc. 2009 International Conference on Multimedia Information Networking and Security*, Washington, USA, 2009, pp. 35-40.

[13] N. Aoki, “A semi-lossless steganography technique for G.711 telephony speech,” in *Proc. Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Darmstadt, Germany, 2010, pp. 534-537.

[14] Y. F. Huang, S. Tang, and J. Yuan, “Steganography in inactive frames of VoIP streams encoded by source codec,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 296-306, 2011.

[15] H. Tian, H. Jiang, K. Zhou, and D. Feng, “Transparency-orientated encoding strategies for Voice-over-IP steganography,” *The Computer Journal*, vol. 55, no. 6, pp. 702-716, 2012.

[16] P. Gope and T. Hwang, “A Realistic Lightweight Anonymous Authentication Protocol for Securing Real-Time Application Data Access in Wireless Sensor Networks,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7124-7132, Nov. 2016.

[17] H. Tian, J. Liu, and S. Li, “Improving security of quantization-index-modulation steganography in low bit-rate speech streams,” *Multimedia Systems*, vol. 20, no. 2, pp. 143-154, 2014.

[18] Q. Qi, D. Peng, and H. Sharif, “DST approach to enhance audio quality on lost audio packet steganography,” *Eurasip Journal on Information Security*, vol. 1, pp. 20, 2016.

[19] P. Liu, S. Li, and H. Wang, “Steganography integrated into linear predictive coding for low bit-rate speech codec,” *Multimedia Tools & Applications*, vol. 9, pp. 1-23, 2016.

[20] A. Janicki, “Pitch-based steganography for Speex voice codec,” *Security and Communication Networks*, vol. 9, iss. 15, pp. 2923-2933, 2016.

[21] H. Tian, J. Sun, C. C. Chang, J. Qin, and Y. Chen, “Hiding information into Voice-Over-IP streams using adaptive bitrate modulation,” *IEEE Communications Letters*, vol. 21, no. 4, pp. 749-752, Apr. 2017.

[22] G. Xin, Y. Liu, T. Yang, and Y. Cao, “An adaptive audio steganography for covert wireless communication,” *Security and Communication Networks*, vol. 2018, Article ID 7096271, pp. 1-10, 2018.

[23] L. Zhang, X. Hu, W. Rasheed, T. Huang, and C. Zhao, “An Enhanced Steganographic Code and its Application in Voice-Over-IP Steganography,” *IEEE Access*, vol. 7, pp. 97187-97195, 2019.

[24] M. E. Hellman, “An overview of public key cryptography,” *IEEE Communications Magazine*, vol. 50, pp. 42-49, 2002.

[25] K. Fu, S. Kamara, and T. Kohno, “Key regression: Enabling efficient key distribution for secure distributed storage,” presented at the 13th Annual Network and Distributed System Security Symposium (NDSS ’06), Microsoft, 2006 (research.microsoft.com/apps/pubs/default.aspx?id=102086).

[26] S. Dagar, “Highly randomized image steganography using secret keys,” in *International Conference on Recent Advances and Innovations in Engineering* (ICRAIE-2014), Jaipur, 2014, pp. 1-5. doi: 10.1109/ICRAIE.2014.6909116

[27] J. M. Gutiérrez-Cárdenas, “Secret Key Steganography with Message Obfuscation by Pseudo-random Number Generators,” in *2014 IEEE 38th International Computer Software and Applications Conference Workshops*, Vasteras, 2014, pp. 164-168. doi: 10.1109/COMPSACW.2014.31

[28] N. Patel, and S. Meena, “LSB based image steganography using dynamic key cryptography,” in *2016 International Conference on Emerging Trends in Communication Technologies* (ETCT), Dehradun, India, 2016, pp. 1-5. doi: 10.1109/ETCT.2016.7882955

[29] M. Anwar, M. Sarosa, and E. Rohadi, “Audio Steganography Using Lifting Wavelet Transform and Dynamic Key,” in *Proc. 2019 International Conference of Artificial Intelligence and Information Technology* (ICAIIT), Yogyakarta, Indonesia, 13-15 March 2019, pp. 133-137.

[30] Y. Jiang, et al. “Covert voice over Internet protocol communications with packet loss based on fractal interpolation,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 4, pp. 54:1-54:20, 2016.

[31] B. Xiao, Y. F. Huang, and S. Tang, “An approach to information hiding in low bit rate speech stream,” in *Proc. IEEE GLOBECOM 2008*, Dec. 2008, USA, IEEE Press, pp. 371–375.

[32] C. Krätzer, J. Dittmann, T. Vogel, and R. Hillert, “Design and evaluation of steganography for voice-over-ip,” in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, pp. 2397–3234.

[33] J. Dittmann, D. Hesse, and R. Hillert, “Steganography and steganalysis in voice over IP scenarios: Operational aspects and first experiences with a new steganalysis tool set,” in *Security, Steganography, and Watermarking of Multimedia Contents* VII. San Jose, CA: Electronic Imaging Science and Technology, 2005, pp. 607–618.

[34] Y. F. Huang and S. Tang, "Covert Voice over Internet Protocol communications based on spatial model," *Science China Technological Sciences*, Springer, vol. 59, no. 1, pp. 117-127, 2016.

[35] M. Goodrich, R. Tamassia, and J. Hasic, "An efficient dynamic and distributed cryptographic accumulator," *Tech. Rep., Johns Hopkins Information Security Institute*, pp. 1-12, 2002.

[36] DSLA II Getting Started Guide Revision 3.0, Malden Electronics Ltd, UK, 2013.

[37] ITU-T Recommendation, *P. 862. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, Feb. 2001.

[38] J. F. Box, "Guinness, Gosset, Fisher, and Small Samples," *Statistical Science*, vol. 2, no. 1, pp. 45-52, 1987.