Wireless Networks (2020) 26:5817–5823
https://doi.org/10.1007/s11276-020-02312-4

Reliability enhancement and packet loss recovery of any steganographic method in voice over IP

Parvaneh Amirzade Dana[1] · Zahra Esmaeilbeig[1] · Mohammad-Reza Sadeghi[1]

Published online: 26 March 2020

Abstract
VoIP data is transmitted through a transport protocol called user datagram protocol (UDP) which is intrinsically unreliable. The quality of the voice or multimedia trasmitted during a VoIP session is not much affected after a few packet loss. However, if a secret message is embedded inside VoIP packets using any steganographic method, the integrity of the secret message can be undermined due to the packets being lost during transmission. In this paper, we propose a scheme which is capable of enhancing the reliability of any VoIP steganographic method. We first distribute k message bits into k successive RTP packets. Then, parity bits are used for reconstruction of lost bits caused by packet loss. The implementation of our scheme on matrix embedding using binary Hamming codes steganography results in a reasonable reliability, a good speech quality and a very high steganographic bandwidth of 3050 bps.

Keywords Steganography · Voice over IP · UDP

Mathematics Subject Classification 05C15 · 05C20 · 68Q25

1 Introduction

Steganography is the technique that conceals secret information inside an innocent carrier. This carrier, also known as steganographic cover, has evolved from images (JPGs or bitmaps) and sound files (mp3s or WAV files) to video files which can host large secret messages. Voice over IP (VoIP) is an important service of the IP-based network and is rapidly replaced by classical telephony services. Due to the proliferation of VoIP, together with the large volume of voice data transmitted during a call, in recent years this service has shown a stupendously great potential for information hiding [1–3]. The data is kept secret among the bits of a digital voice over the Internet protocol conversation transmitted over the network by means of protocols such as SIP, RTP and UDP. Moreover, steganography detection or steganalysis of VoIP data is hard to accomplish due to the ephemerality of the carrier.

VoIP steganographic methods presented in the literature can be classified into two main categories. The first category of studies uses the digital representation of the transmitted voice as the steganographic camouflage. The second category targets the VoIP protocol fields such as SIP (signalling protocol), RTP (transport protocol) and RTCP (control protocol). In the following, we will elaborate on some techniques in both cases:

1.1 Previous VoIP steganography

Previous approaches for VoIP steganography can be categorized into two distinct groups: first, the conventional embedding methods, which are commonly used in the literature for image, audio and video steganography, can be applied to voice payload. The second category uses different network protocols as a covert channel.

✉ Mohammad-Reza Sadeghi
msadeghi@aut.ac.ir

Parvaneh Amirzade Dana
pamirzade@gmail.com

Zahra Esmaeilbeig
zahraesmailbeig@yahoo.com

[1] Department of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran

1.1.1 Voice payload steganography

The first VoIP steganographic method suggests using least significant bits (LSB) of voice samples to provide the resistance against packet loss [4, 5]. Later, Zhijun et al. in [6] use energy characteristics of G.711 encoded speech signal to increase the embedding capacity in each packet that has a little degradation in the quality of speech. Also, in [7] the feasibility of methods other than LSB embedding like DSSS (direct sequence spread spectrum), FHSS (frequency-hopping spread spectrum) and echo hiding is cosidered for VoIP steganography. Experimental results show that a steganographic bandwidth of 20 *bps* can be achieved using these techniques.

Higher steganographic bandwidth of 133.3 *bps* is achieved in [8] by embedding the bit stream of a secret message in least significant bit of LSP (linear spectrum pairs) quantization parameter.

In order to minimize the total number of changes needed to carry out a message and thus to increase the embedding efficiency, matrix embedding is introduced by Crandall [9], and popularized by Westfeld who incorporated a typical implementation using binary Hamming codes in his *F*5 algorithm [10]. In [11], the authors use "divide and rule" strategy to achieve an optimal performance in matrix embedding based on Hamming codes.

1.1.2 Packets time relations and protocol steganography

These methods attempt to embed the secret data either in unused or optional fields of protocol headers or modulated inter-packet times. These solutions can employ specifc fields of VoIP protocols such as signalling protocol (SIP), transport protocol (RTP) and control protocol (RTCP) [12–14]. Some other methods of this category attempt to embed the secret data by altering inter-packet time relation [15], the sequence order of packets [16] or making intentional losses [17].

1.1.3 Hybrid steganographic methods

Lost audio packets steganography (LACK) [18] modifies both payload of the packets and their time dependencies. At the transmitter, some selected audio packets are intentionally delayed before transmitting. If the delay of such packets at the receiver is considered excessive, they will not be passed to the voice decoder for reconstruction by a receiver which is not aware of the steganographic procedure. The payload of the intentionally delayed packets is used to transmit secret information to receivers aware of the procedure, so no extra packets are generated. Although the whole payload of an RTP packet carries the secret message, the number of packets that can be intentionally delayed is limited by the QoS requirements of a VoIP call. In order to maintain the quality of the decoded voice at an acceptable level, the packet lost ratio needs to be limited. For example, G.711 codec can tolerate a maximum of 3% packet loss. Whereas, loss tolerance is 1% for G.723.1 and 2% for G.729A. Therefore, LACK can transmit 320 *bps* of the secret message by imposing 0.5% packet loss.

1.2 Challanges of VOIP steganography

The followings are two most essential challanges of designing steganography schemes for VoIP.

1.2.1 Unreliabilty of transport

VoIP data is transmitted through the UDP protocol which is intrinsically unreliable. This protocol is suitable for transmitting audio or video, however, it is not reliable when used for steganography. In other words, the quality of voice or video will not be much affected after a few packet loss but the integrity of the embedded secret message can be undermined.

1.2.2 Latency

Naturally, a VoIP call is extremely prone to the media degradation due to the packet latency. So, any processing overhead from steganographic system into the cover-medium or delay due to inspection of potential cover-medium packets will cause a considerable degradation on the Quality of Service (QoS).

It is necessary to design a steganographic system whose performance is equivalent to the fact that there is no incorporated in the VoIP system.

1.3 Our motivation

Many of the previous proposed VoIP steganographic methods such as those in [1, 23] do not consider the problem of reliability, while sending the stego-messages, or just assume that the receiver requests to resend the secret message, in case it is not fully received [18]. However, resending reduces the steganographic bandwidth while the resent message could be lost again. In this paper, we propose a scheme which is capable of enhancing the reliability of any voice payload steganographic system in VoIP.

1.4 Related work

One way to address the reliability problem is changing signal codec [19]. In order to handle the packet loss, the authors use G.711 speech codec for steganography and propose to transmit multiple pieces of arbitrary

information. Similarly, in [20], Neal et al. reduce the probability of packet loss by applying the G.711 codec.

Zhang et al. [21], propose a method that applies one packet loss prediction model and Gilbert packet loss model to decide whether an audio packet would be discarded for embedding or not.

ReLACK is another method to achieve the reliability which is proposed in [24].In that work, a modified (n, k) secret sharing scheme based on Lagrange interpolation is used to make the renowned LACK scheme tolerant of the packet loss. There is a major problem with this scheme which is summarized as follows. At the receiver, the parameter k is calculated according to the difference between the sequence number of the first two delayed packets. However, because of the unreliability of the link, these two packets may not be received and consequently a false value for k is used to detect the secret message. The first experimental results of LACK are presented in [25] which consider a controlled LAN network, so no RTP packets are lost or excessively delayed, unless intended. This assumption is usually far from real conditions. Moreover, although the whole payload of the packets are used to embed the secret message, the steganographic bandwidth is reduced by a factor of $\frac{n}{k}$ in comparison with LACK.

Our contributions in this paper are listed as follows.

(1) We introduce a scheme for distributing the bits of the secret message among the packets such that if a packet containing the secret message is lost, it is feasible to recover the lost secret bits. For this purpose, for each message segment a parity bit is defined and every bit of each segment is embedded in different packets. In this way, losing each packet leads to lose only one bit of the message segment that is recoverable by the parity bit.
(2) To compensate the bandwidth reduction, we use the matrix embedding based on Hamming codes to embed the message in payload.

The rest of the paper is organized as follows. Section 2 presents a detailed description of our proposed method. The experiments conducted to evaluate the performance of the proposed method are reported in Sect. 3 and a conclusion is drawn in Sect. 4.

2 Proposed method

In this work, we use an LSB embedding method with embedding relative payload of $\alpha = \frac{m}{l}$ in which l is the number of LSBs in each RTP packet and m is the number of data bits that can be embedded in one RTP packet. In order to recover one packet loss in each k RTP packets, we propose a scheme to enhance the reliability of any VoIP steganographic method.

We divide the packets generated in a sender into groups of k packets and the given secret message into vectors of k bits. As will be explained next, k is chosen such that $m \geq k^2 + 2k$. This allows us to embed at least $k^2 + 2k$ data bits in each packet. In our proposed scheme, three types of data are embedded in each packet: packet number, secret message bits and parity-check bits, (see Fig. 1). In each embedding iteration, a message is embedded in k packets. To recover message bits embedded in the lost packet, it is necessary to mark the k packets of a group by allocating $n = \lceil \log_2 k \rceil$ bits to each packet number. We also assign $2k$ bits to parity-check bits. Thus, $k^2 - n$ bits of the secret message can be embedded in each RTP packet.

As illustrated in Fig. 1, in each embedding iteration, $k^2 - n$ message vectors and k RTP packets are buffered as input. For the sake of packet marking, n vectors of the form shown in Fig. 2 are added to input message vectors and so k^2 data vectors are obtained. As illustrated in Fig. 2, for $k = 6$, a binary representation of 0 to $k - 1$ is computed and reordered as follows: the first bits form the first vector, the second bits form the second vector and, by continuing this process, the n-th bits form the n-th vector.

In order to construct the packet parity vectors $L_1, \ldots, L_k$, k^2 data vectors are divided into k groups of k vectors. For every $1 \leq i, j \leq k$, $L_i = (L_{i1}, \ldots, L_{ik})$ where L_{ij} is obtained by performing XOR operation on bits of the j-th vector of the i-th group. Finally, a master parity vector L is computed as $L = L_1 \oplus L_2 \oplus \ldots \oplus L_k$.

In each iteration, the first bits of all k^2 input vectors are embedded in the first RTP packet, the second bits are embedded in the second RTP packet and, by continuing the same process, the k-th bits in the k-th packet. Then, for every $1 \leq i \leq k$, L_i is embedded in the i-th packet and L is embedded in all packets (see Fig. 1). The above-mentioned procedure is repeated until all vectors of the secret message are embedded and sent to the receiver.

Assuming that the sender and the receiver have shared the steganographic scheme and the parameter k as a key, each RTP packet is inputted as a carrier of a secret message to the extraction iteration. The receiver, at each iteration, allocates a buffer of $k(k^2 - \log_2 k)$ for the secret message bits, extracted from a group of k RTP packets. It should be noted that extracted bits from packets with the same master parity vector L are saved in this buffer. After receiving each RTP packet, data bits are extracted from LSBs according to the embedding method. The first $n = \lceil \log_2 k \rceil$ extracted bits represent the packet number. As indicated in Fig. 1, the next $k^2 - n$ extracted bits are placed in the buffer according to the packet number. The next

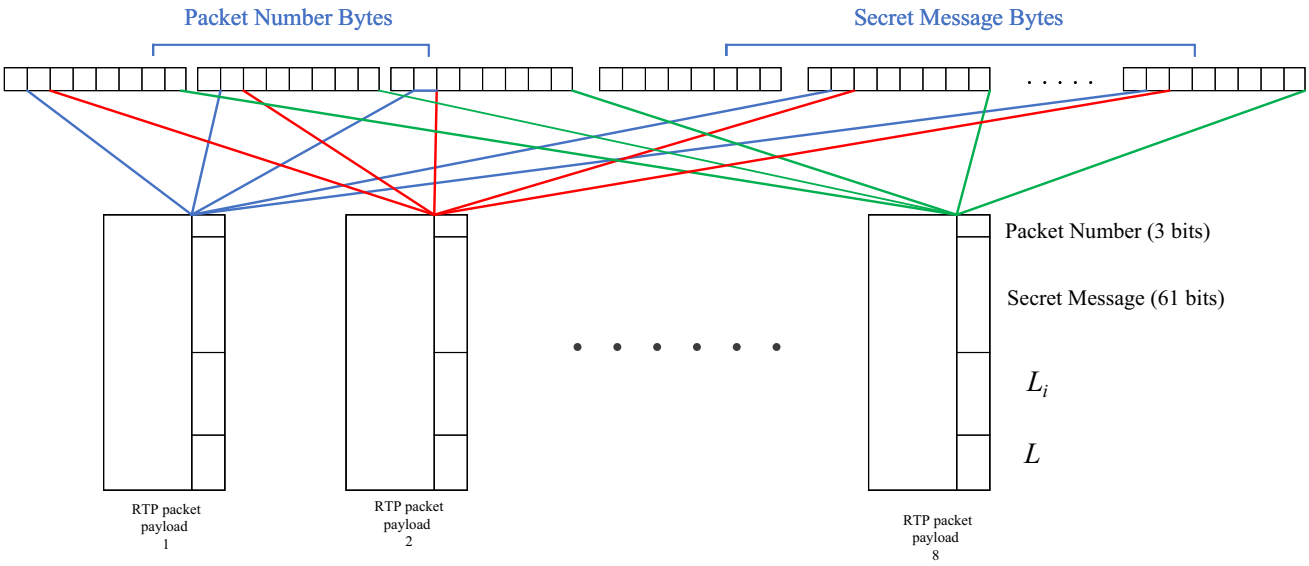

Fig. 1 The proposed scheme for distribution of packet number, secret message and parity bits in packets for $k = 8$

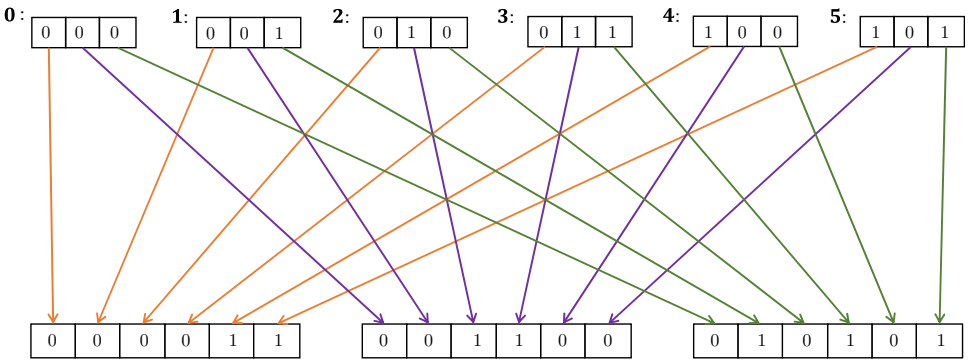

Fig. 2 proposed scheme for construction of packet number vectors for $k = 6$

k extracted bits represent the packet parity vector and the last k extracted bits form the master parity vector L.

This procedure is repeated until a packet with a different master parity vector L is received. If the number of the received packets with the same master parity vector is k, then the message is completely received in this iteration. Otherwise, if only one packet in a group is lost, our scheme is able to recover the lost bits. The receiver, having the packet numbers, can find the lost packet number. Suppose that j is the lost packet number. Receiver is aware of the loss of the j-th bit of each secret message vector and that it should be recovered. First, the parity vector of the j-th packet is computed as follows:

$$L_j = L \oplus L_1 \oplus \cdots \oplus L_{j-1} \oplus L_{j+1} \oplus \cdots \oplus L_k. \tag{2.1}$$

Suppose that $\mathbf{m} = (m_1, \ldots, m_k)$ is a vector in the i-th group of the data vectors. The lost bit m_j is recovered by bit-wise XOR as below:

$$m_j = L_{ij} + (m_1 + \cdots + m_{j-1} + m_{j+1} + \cdots + m_k). \tag{2.2}$$

Provided that the QoS requirement of a voice call over the network is less than 1% packet loss, the probability of losing more than one packet in each iteration would be less than 10^{-3}. Consequently, our scheme enhances the reliability in terms of the packet loss at the cost of sacrificing the following percentage of the available steganographic bandwidth:

$$\frac{2k + log_2 k}{k^2 + 2k} \times 100. \tag{2.3}$$

3 Performance evaluation

3.1 Embedding method selection

Our scheme, unlike other previous works in [24], is applicable to enhance the reliability of any VoIP steganographic method. However, in order to verify the

performance of our proposed scheme, for enhancing the reliability in VoIP steganography, we choose the matrix embedding (ME) using binary Hamming codes. By minimizing the number of changes due to embedding, ME secures a lower distortion in cover signal in comparison with the conventional LSB matching [26].

3.1.1 Matrix embedding using binary hamming codes

Matrix embedding is a type of syndrome coding. In a linear binary code $\mathbf{C}(n,k)$ with the parity-check matrix $\mathbf{H}$, we have $\mathbf{Hx}=\mathbf{0}$ if $\mathbf{x}\in\mathbf{C}$, otherwise, $\mathbf{Hx}=\mathbf{s}$ is a syndrome. Suppose that the sender intends to embed a message $\mathbf{m}$ with $p=n-k$ bits into a vector $\mathbf{x}$ of 2^p-1 LSBs. The sender and recipient share a $p\times(2^p-1)$ binary matrix $\mathbf{H}$ that contains all non-zero binary vectors of length p as its columns. An example of such a matrix for $p=3$ is

$$\mathbf{H}=\begin{bmatrix}0&0&0&1&1&1&1\\0&1&1&0&0&1&1\\1&0&1&0&1&0&1\end{bmatrix}. \tag{3.1}$$

The sender computes $\mathbf{m}-\mathbf{Hx}$ onto the syndrome which is the binary representation of the index of the bit that should be flipped in the cover $\mathbf{x}$ and keeps the remaining bits unchanged. This altered cover denoted by $\mathbf{y}$ is sent to the receiver. At the reciever point, message bits can be extracted as the syndrome of $\mathbf{y}$:

$$\mathbf{Hy}=\mathbf{m}. \tag{3.2}$$

This algorithm is capable of embedding p bits of the secret message in 2^p-1 bits of the cover, which leads to an embedding efficiency $e=\frac{\text{Number of embedded bits}}{\text{Average number of changes}}$ of

$$e_p=\frac{p}{1-2^{-p}}. \tag{3.3}$$

In Conventional LSB matching, on average, two bits are embedded using one change or, equivalently, the embedding efficiency is 2. According to Eq. (3.3), it is possible to substantially increase the embedding efficiency and thus to embed the same payload with fewer embedding changes (see Table 1).

The parameter p in the matrix embedding is chosen according to the secret message length and cover which is the payload length of an RTP packet in this case. According to Table 1, the larger p leads to a higher embedding efficiency and a lower embedding relative payload α. Therefore, a compromise between the embedding efficiency and the payload must be sought. To achieve this goal, Tian et al. [11] have proposed a divide-and-rule strategy to obtain an optimal embedding efficiency.

Table 1 Relative payload and embedding efficiency in bits per change for matrix embedding using binary Hamming codes

p	e_p	α_p
1	2.000	1.000
2	2.667	0.667
3	3.429	0.429
4	4.267	0.267
5	5.161	0.161
6	6.093	0.093
7	7.055	0.055
8	8.031	0.031
9	9.018	0.018

3.1.2 Optimal matrix embedding

In this scheme, the cover and the message are segmented to achieve the maximum embedding efficiency. In the proposed algorithm, for a given message and a cover length, a search over all possible segmentations is done and the optimal segmentation on the message and cover is returned. For instance, according to the lookup table provided in [11], in order to embed 20 bits of the message in 40 LSBs of the cover, we can obtain an embedding efficiency of 3.0769 by dividing the message as $\{5,5,5,5\}$ and the cover as $\{10,10,10,10\}$. Subsequently, to embed 5 bits of the secret message in 10 bits of the cover, we can employ the matrix embedding using Hamming codes with $p=2$ and $p=3$.

In order to implement our scheme, we perform the steganography operations on G.711 coded speech. Under this setting, RTP packets with a 160-byte payload is generated to carry Voice-over-IP. Since we apply LSB embedding, in the sender side upon generating each packet, we have 160 bits of the cover signal to be embedded. In this work, our goal is to obtain a trade-off between the embedding relative payload and the embedding efficiency. Therefore, by using the optimal matrix embedding in [11], we achieve the steganographic bandwidth of 80 bit-per-packet and the embedding efficiency of 3.0769, by dividing the message into segments of 5 bits and the cover into segments of 10 bits.

3.2 Implementation results

To evaluate the feasibility and effectiveness of our proposed method for real time steganography in Voice-over-IP, we utilize the open source VoIP project Linphone to construct a covert communication system [27]. This project supports typical codecs, such as ITU-T G.711,G.729, G.722, speex, gsm, etc. In the experiments, G.711 is set as the codec which encodes the cover speech at 64 kbps. We also choose $k=8$ to compensate for one lost packet in each 8 RTP packets. According to our proposed scheme in

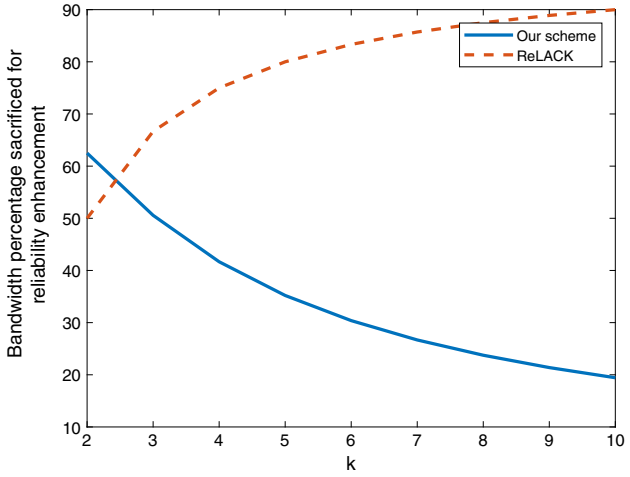

Fig. 3 Comparison of Bandwidth sacrifice between our scheme and ReLACK

Sect. 2, we can embed $k^2 - \log_2 k = 61$ bits of the secret message in each RTP packet. Consequently, an steganographic capacity of 3050 bps is obtained which is significantly higher in comparison with the ReLACK method mentioned in [24]. Assume that we want to compensate one packet loss in a group of k using the ReLACK method. According to [24], steganographic bandwidth reduces by a factor of k. Therefore, $1 - \frac{1}{k}$ percent of the bandwidth is sacrificed whereas in our scheme this percentage is lower according to Eq. (2.3). Fig. 3 illustrates a comparison between our scheme and ReLACK in terms of their bandwidth reduction.

In order to examine the impact of our proposed steganographic system on quality of speech, we utilize the audio recordings from the TIMIT [28] continuous speech corpus. To assess the quality of speech after embedding, we employed the perceptual evaluation of speech quality (PESQ) method presented in the ITU-T P.862 Recommendation [29], which compares an original speech signal with a degraded signal and outputs a PESQ score in the range between -0.5 and 4.5 as a prediction of the perceived quality. In Fig. 4 an average PESQ measurement of ten different recordings after embedding is illustrated.

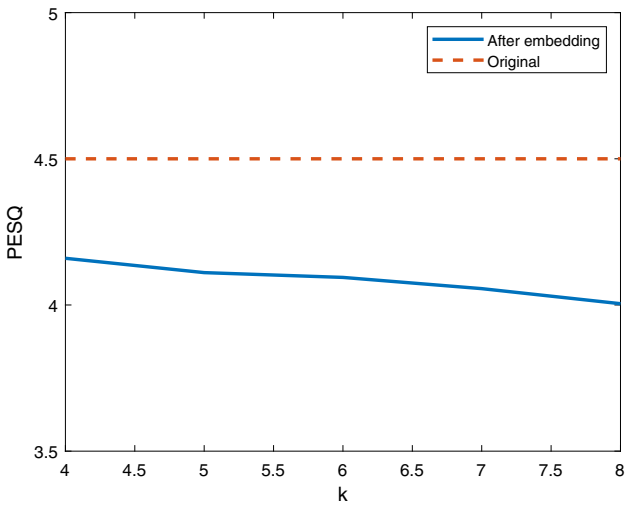

Fig. 4 Perceptual Evaluation of Voice Quality (PESQ) results for different values of k

4 Conclusion

We propose a scheme to enhance the reliability of any VoIP steganographic method. In our scheme, one packet loss in each group of k packets is compensated and the lost bits of the secret message are reconstructed using parity vectors transmitted to the receiver. In this scenario, secret message bits are lost with a probability of less than 0.001. We compared our scheme with RelACK which is the only previously proposed method capable of packet loss compensation. ReLACK is only applicable to LACK whereas our scheme can be applied to any embedding algorithm. Also, at different values of k, the sacrificed steganographic bandwidth in our method is less than ReLACK. In this work, in order to compensate for the packet loss and lost bits reconstruction, parity bits are used. Our scheme leads to the fact that using error correction codes is suitable to enhance the reliability of VoIP steganography.

References

1. Huang, Y. F., Tang, S., & Yuan, J. (2011). Steganography in inactive frames of voip streams encoded by source codec. *Transactions on Information Forensics and Security*, *6*(2), 296–306.
2. Zhou, K., Feng, D., Tian, H., & Jiang, H. (2012). Transparency-orientated encoding strategies for voice-over-ip steganography. *The Computer Journal*, *55*(6), 702–716.
3. Lubacz, J., Mazurczyk, W., & Szczypiorski, K. (2010). Vice over ip. *Spectrum*, *47*(2), 42–47. IEEE.
4. Komaki, N., Aoki, N., & Yamamoto, T. (2003). A packet loss concealment technique for voip using steganography. *Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *86*(8), 2069–2072. IEICE.
5. Ito, A., Suzuki, Y., et al. (2010). Information hiding for G.711 speech based on substitution of least significant bits and estimation of tolerable distortion. *Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *93*(7), 1279–1286. IEICE.
6. Wu, Z., & Yang, W. (2006). G. 711-based adaptive speech information hiding approach. In *International conference on intelligent computing*, (pp. 1139–1144). New York: Springer.
7. Takahashi, T., & Lee, W. (2007). An assessment of voip covert channel threats. In *International conference on security and privacy in communications networks and the workshops-secure comm*, (pp. 371–380). IEEE.

8. Xu, T., & Yang, Z. (2009). Simple and effective speech steganography in g.723.1 low-rate codes. In *International conference on wireless communications and signal processing*, (pp. 1–4). IEEE.
9. Crandall, R. (1998). Some notes on steganography. *Posted on steganography mailing list*, pp. 1-6.
10. Westfeld, A. (2001) F5a steganographic algorithm. In*International workshop on information hiding*, (pp. 289–302)New York: Springer.
11. Tian, H., Qin, J., Huang, Y., Chen, Y., Wang, T., Liu, J., et al. (2015). Optimal matrix embedding for voice-over-ip steganography. *Signal Processing*, *117*, 33–43.
12. Mehic, M., lachta, J. & Voznak, M.(2015). Hiding data in sip session. In *International conference on telecommunications and signal processing (TSP)*, (pp. 1–5) IEEE.
13. Arackaparambil, C., Yan, G., Bratus, S., & Caglayan, A. (2012). On tuning the knobs of distribution-based methods for detecting voip covert channels. In *Hawaii international conference on system sciences*, (pp. 2431–2440) IEEE.
14. Mazurczyk, W., & Szczypiorski, K. (2008). Covert channels in sip for voip signalling. In *International Conference on Global e-Security*, (pp. 65–72). New York: Springer.
15. Zhang, X., Tan, Y., Liang, C., Li, Y., & Li, J. (2018). A covert channel over volte via adjusting silence periods. *Access*, *6*, 9292–9302. IEEE.
16. Kundur, D., & Ahsan, K. (2003). Practical internet steganography: data hiding in IP. In *Proc. Texas wksp. Security of information systems*.
17. Servetto, S. D., & Vetterli, M. (2001). Communication using phantoms:covert channels in the internet. In *International symposium on information theory (IEEE Cat. No. 01CH37252)*, (p. 229). IEEE.
18. Mazurczyk, W., & Lubacz, J. (2010). Lacka voip steganographic method. *Telecommunication Systems*, *45*(2–3), 153–163.
19. Neal, H., & ElAarag, H. (2015). A reliable covert communication scheme based on VoIP steganography. In *Transactions on data hiding and multimedia security*, (pp. 55–68) New York: Springer.
20. Neal, H., & ElAarag, H. (2012). A packet loss tolerant algorithm for information hiding in voice over IP. In *Proceedings of IEEE Southeastcon*,(pp. 1–6). IEEE.
21. Jiang, Y., Tang, S., Zhang, L., Xiong, M., & Yip, Y. J. (2016). Covert voice over Internet protocol communications with packet loss based on fractal interpolation. *Transactions on Multimedia Computing, Communications and Applications (TOMM)*, *12*(4), 54. ACM.
22. Mazurczyk, W. (2013). Voip steganography and its detectiona survey. *Computing Surveys (CSUR)*, *46*(2), 20. ACM.
23. Huang, Y., Liu, C., Tang, S., & Bai, S. (2012). Steganography integration into a low-bit rate speech codec. *Transactions on Information Forensics and Security*, *7*(6), 1865–1875. IEEE.
24. Hamdaqa, M., & Tahvildari, L. (2011). Relack: a reliable voip steganography approach. In *International conference on secure software integration and reliability improvement (SSIRI)*, (pp. 189–197). IEEE.
25. Mazurczyk, W. (2012). Lost audio packets steganography: the first practical evaluation. *Security and Communication Networks*, *5*(12), 1394–1403.
26. Fridrich, J. (2009). *Steganography in digital media: principles, algorithms, and applications*. Cambridge: Cambridge University Press.
27. "Linphone open source voip project," http://www.linphone.org/, Accessed: (2019).
28. "TIMIT acoustic-phonetic continuous speech corpus," https://catalog.ldc.upenn.edu/LDC93S1.
29. "ITU-T recommendation P.862 perceptual evaluation of speech quality (PESQ)," ITU-T, Feb. (2001).