Efficient Least-significant-bits Steganography for VoIP

[1]Jin Liu, [1,*]Ke Zhou, [2,*]Hui Tian, [1]Cunhua Li
[1.] *School of Computer Science and Technology, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China*
[2.] *College of Computer Science and Technology, National Huaqiao University, Xiamen 361021, China*
**Corresponding Authors*
{geneleocn@gmail.com, k.zhou@hust.edu.cn, cshtian@gmail.com, li.chunhua@hust.edu.cn}

Abstract

For the instantaneity of real-time applications such as voice over Internet protocol (VoIP), the processing delay and efficiency need to be carefully considered. Matrix coding is a widely used coding scheme for steganography to improve the embedding efficiency which means the average hiding bits per covering bit change. In order to take full advantage of it, a classifying least-significant-bits (LSB) with frame window (CLFW) scheme is proposed for low bitrate speech covers in VoIP communication. The classifying method selects LSB with similar characteristics which bring consistent effect to speech quality for every bit modification. Meanwhile, the frame window strategy collects more speech frames in VoIP speech stream for matrix coding, which enhances the embedding efficiency. In CLFW, frame LSB are classified into three levels to construct separate sets of covering bits, which limits the range of embedding modification. As a result, the speech quality degradation caused by feature discrepancies of speech frames is minimized. The experimental results demonstrate that CLFW method largely increases the embedding efficiency compared with existing schemes and is much more imperceptible than direct matrix coding and simple LSB steganography.

Keywords: *Steganography, Embedding Efficiency, Matrix Coding, Least-significant-bits, VoIP*

1. Introduction

LSB substitution and its variations are the most common solutions in steganography field, which are still very infusive right now [1]. Subsequently, some improved LSB methods such as "LSB matching" and "LSB matching revisited" increased the imperceptibility of embedding algorithm. And then many stego coding schemes, especially matrix coding [2], were widely used with LSB method, which largely increased the embedding efficiency.

As the mainstream mainly concentrate on image covers, steganographic algorithms in speech or audio domain need more serious attention for the popularization of speech communication, such as VoIP and mobile phones. Due to the restriction of communication bandwidth, speech data are always compressed by various codecs for different applications. General speech compressing codecs include ITU-T G.711, ITU-T G.723.1 [3], ITU-T G.729, IETF iLBC (RFC3951) and the open source Speex. Unlike image covers, the compressed speech codecs often contain little redundancies and have some latency requirements for covert communication. Researchers have already proposed some LSB embedding scenarios of those compressed speech steganography, such as [4], which used G.711 as the covering media. Huang et al. [5] and Tian et al [6-7] individually structured the covert communication model of VoIP, which brought insights into practical applications in low bitrate speech. They both use LSB related methods as the underlying embedding algorithm, where the basic embedding model for covert speech communication were proposed. Xu et al. [8] and Liu et al. [9] adopted separately two LSB scenarios in G.723.1 and G.729a speech frames with matrix coding schemes which improved the embedding efficiency [10]. Tian et al. [11] proposed a dynamic matrix coding scenario that made embedding in VoIP streams with different performance demands more flexible. Nevertheless, they did not take more LSB into account, and different distortions introduced by separate LSB are overlooked, which may probably be limited by the performance of matrix coding. The lack of LSB affects the embedding efficiency while the introduction of more LSB in one frame causes degradation of speech quality, which is vulnerable to steganalyzers [12].

In the paper, an efficient matrix coding scheme is introduced in covert communication with low bitrate speech to increase the embedding efficiency and imperceptibility of steganography. In the matrix coding scheme more input bits (covering bits) always come with higher embedding efficiency. To decrease the speech quality distortion caused by matrix coding in LSB substitution, perceptual similar LSB in one speech frame are classified into several sets under a variance below a given value according to a given speech quality evaluation criterion. Then combined with frame window scheme in the VoIP streams more cover bits are obtained. With more perceptual similar LSB the CLFW method decreases the distortion of LSB substitution and brings higher embedding efficiency compared with previous schemes, such as method in [8], direct matrix coding with LSB and simple LSB method. The main contribution of this paper is the introduction of matrix coding in low bitrate speech steganography with high embedding efficiency and more imperceptibility, and the combination with a covert communication system for practical usage.

2. Matrix coding in G.723.1 codec

Different from image and other static covering media, matrix coding needs preprocessing for direct applying in low bitrate speech streams. And the latency requirement of real-time VoIP needs to be seriously considered.

2.1. Matrix coding and its embedding efficiency

Matrix coding algorithm was first proposed by Crandall in [13], which then becomes a general stego coding scheme in most applications for steganography in order to improve the embedding efficiency. In stego coding schemes, $COV(\rho, n, m)$ means embedding m bits of secret message in n covering bits with maximal ρ bits changed, where matrix coding is the most widely used scheme. The generally used matrix for steganography is Hamming matrix with size ($m \times 2^m$-1) in the form of $COV(1, 2^m\text{-}1, m)$ (Eq.(1), where $m = 3$). After that, many modified matrix encoding techniques emerged in the form of $COV(2, n, m)$ Hamming code modifying maximum 2 bits [14]. Other cyclic syndrome coding schemes used in steganography include BCH codes [15].

$$H_{3\times 7} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{1}$$

In LSB with Hamming codes, h_i is a ($m \times 1$) column vector of $\boldsymbol{H} = (h_1, h_2, \ldots, h_{2^m\text{-}1})$ in ($m \times 2^m$-1) binary Hamming matrix. Let $\boldsymbol{C} = (c_1, c_2, \ldots, c_{2^m\text{-}1})$ and $\boldsymbol{S} = (s_1, s_2, \ldots, s_m)$ denote respectively 2^m-1 covering bits and m secret bits. $\boldsymbol{D} = \boldsymbol{H}\bullet\boldsymbol{C}^T\oplus\boldsymbol{S}^T$. If $\boldsymbol{D} = \boldsymbol{0}$, no bit in $\boldsymbol{C}$ needs to be changed. Otherwise, there exists k satisfying $\boldsymbol{D} = h_k$. Then the stego cover $\boldsymbol{C}' = (c_1, c_2, \ldots, \tilde{c}_k, \ldots, c_{2^m\text{-}1})$ are obtained. In the retrieving procedure the extracted secret message $\boldsymbol{S}'$ is extracted by $\boldsymbol{S}'=\boldsymbol{C}'\bullet\boldsymbol{H}^{\mathrm{T}}$ (ignore the transmission error).

In matrix coding method, the embedding distortion is reduced by using more covering bits, which means “lower embedding rates, higher imperceptibility”. Hamming codes steganography with LSB reduces the average number of bit change N_c from 1/2 without stego coding method to Eq.(2) with $COV(\rho, 2^m\text{-}1, m)$. The embedding efficiency $\boldsymbol{E}$ in Eq.(3) is derived from Eq.(2) .

$$N_c = \frac{1}{2^m}\sum_{i=1}^{\rho} iC_{2^m-1}^{i} \tag{2}$$

$$E = \frac{m}{N_c} = \frac{m \cdot 2^m}{\sum_{i=1}^{\rho} iC_{2^m-1}^{i}} \tag{3}$$

In CLFW we take the binary Hamming codes scheme $COV(1,\ 2^m\text{-}1,\ m)$ as an example, where m can be adjusted in terms of practical conditions. Then the characteristics of low bitrate speech covers should be seriously considered for matrix coding.

2.2 Shortcomings of matrix coding in low bitrate speeches

Matrix coding is most widely used in image covers with a connotative hypothesis that the distortions introduced by LSB algorithm of single different covering bits have similar characteristics. Otherwise, in binary Hamming codes with matrix which embed secret bits with maximum one bit change, the distortions are largely different per bit change. That causes a quality degradation of speeches. As a consequence, evaluation criterion of embedding efficiency for stego coding with matrix coding is inappropriate. That is, the distortion or imperceptibility varies given the same embedding efficiency. When used in low bitrate speech, the LSB mainly consist of different parameter bits (Table 1) in one frame. Due to the different characteristics of parameter bits, the distortions largely varied. The existing difference decreases the speech quality and leads to the rise and fall of it, which is insecure for covert communications. Section 3 will give a detailed description of the difference.

2.3 ITU-T G.723.1 frame codec

G.723.1 is an ITU-T speech codec standard used for compressing the speech or other audio signal component of multimedia services in a very low bitrate. It is widely used in today's speech communication area. The codec contains two bitrates of 5.3 and 6.3 kbit/s [3] with similar characteristics in the main implementation algorithm. The steganography of these two bitrates codec are almost the same. In CLFW method 6.3 kbit/s coding frame is taken as an example, which has 192 bits in one frame, Table 1 illustrates the bit allocation of G.723.1 6.3 kbit/s frame. In the codec one frame consists of several separate parameters such as RF, VF, LPC and ACL.

Table 1. G.723.1 bit allocation of 6.3 kbit/s frame

***Parameters* coded**	***Sub* frame 0**	***Sub* frame 1**	***Sub* frame 2**	***Sub* frame 3**	***Sum***
Flag bits	-	-	-	-	2
Unused bit	-	-	-	-	1
LPC indices (lpc)	-	-	-	-	24
Adaptive codebook lags (acl)	7	2	7	2	18
All the gains combined (gain)	12	12	12	12	48
MSB of pulse positions (msbPos)	-	-	-	-	13
Pulse positions (pPos)	16	14	16	14	60
Pulse signs (pSig)	6	5	6	5	22
Grid index (grid)	1	1	1	1	4
Total					192

To ensure the correct transmission and reception, the 2 bits denoting bitrate (RF) and validation indication (VF) should not be modified. Then the rest 190 bits are used for steganography in terms of their noise resistant abilities by speech quality evaluation. The speech quality distortion levels of LSB for all parameters are evaluated in section 3 where appropriate bits for steganography are presented.

3. Proposed scheme

3.1 The Classifying LSB method

Xu et al. proposed a LSB with matrix coding scheme in [8], where only the lowest 5 bits of LSB vector quantization (VQ) indices (lpc parameter in Table 1) were selected for hiding. With the selected coding matrix it has a low embedding efficiency (see Table 5). In her scheme LSB even with less distortion were neglected which may due to the restriction of matrix coding.

To overcome this shortcoming, more detailed evaluation is given in the form of bit noise resistance evaluation by the commonly used perceptual evaluation of speech quality (PESQ) method regarding G.723.1 6.3 kbit/s frames. ITU-T P.862 PESQ is a speech evaluation standard which has a value between -0.5 and 4.5 [16]. Higher PESQ value always comes with more perceptual comfortable speech. PESQ takes original speech as a reference and the processed speech as the degraded version. According to previous analyzed disadvantages Figure 1 gives the mean PESQ values of each bit of G.723.1 6.3kbit/s frame, which utilized almost 20,000 seconds of speech samples. As a referred PESQ value, after an encoding and decoding process of G.723.1 6.3 kbit/s codec of non-hidden speech, the mean PESQ value is about an average of 3.57 with the same speech samples, which sounds good for human auditory system.

According to Figure 1 and Table 1, the LSB of each parameter in G.723.1 are obtained. Considering the degree of auditory distortion, the 18 bits highest PESQ value are selected for hiding, which have sustainable distortion compared to non-embedded speeches by human ears. And the PESQ values of parameter LSB in G.723.1 frame are largely varied with a highest PESQ variance. That leads to speech quality degradation during the VoIP transmission, which can easily draw eavesdroppers' attention. Otherwise, the advantage of the increase of embedding efficiency, as the key functionality of matrix coding to improve the imperceptibility, is counteracted by this drawback. For this reason, it is inappropriate to use matrix coding scheme with these 18 bits directly.

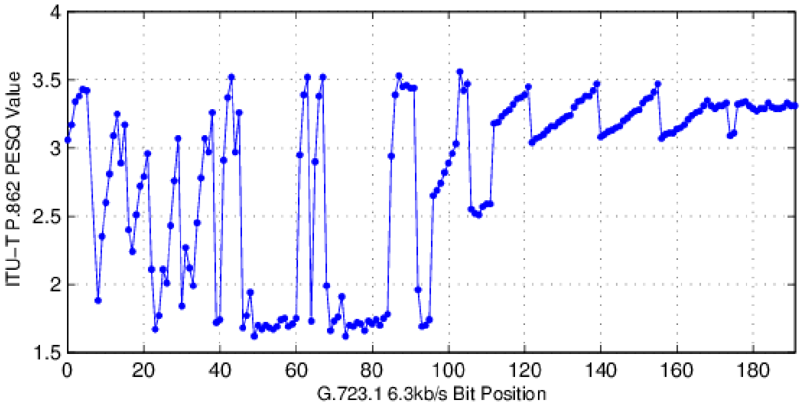

Figure 1. PESQ test of 190 bits for G.723.1 6.3 kbit/s frame

To reduce the unfavorable effect caused by discrepancy of different LSB, the selection of bits with nearest PESQ values in one or more frames is necessary. When bits with bigger difference values are divided into different sets for separate matrix, they have less mutual influence, which is beneficial to steganalysis resistance. This will be proved in the later experiment. Then the LSB are classified in terms of speech quality distortion levels in Table 2, where three levels of LSB are given. In Table 2 at level 0, 5 bits of LSB are selected for matrix coding, and then level 1 containing 6 bits as well as 7 bits in level 2 are exploited.

Table 2. Classification of LSB for G.723.1 frame

Level **label**	***Bit*** **Numbers**	***PESQ*** **(p)**
0	5	$p \geq 3.50$
1	6	$3.45 \leq p < 3.50$
2	7	$3.40 \leq p < 3.45$

Figure 2 depicts the average PESQ values of the three classification levels of frame parameter bits, in which the LSB are selected according to Table 2. Accordingly, three coding matrices should be taken into account with appropriate dimension. For Hamming coding, bits in level 0 are the best choice, and other two levels are used for practical requirements with more embedding capacity. To obtain the best embedding efficiency CLFW suggests level 0 with a window size *n=3*. Here *COV (1, 7, 3)* and *COV (1, 3, 2)* Hamming codes are used for embedding in each levels when only one frame is considered. And when more frames are introduced by a frame window mechanism, larger Hamming matrices are used with more LSB, which result higher embedding efficiency.

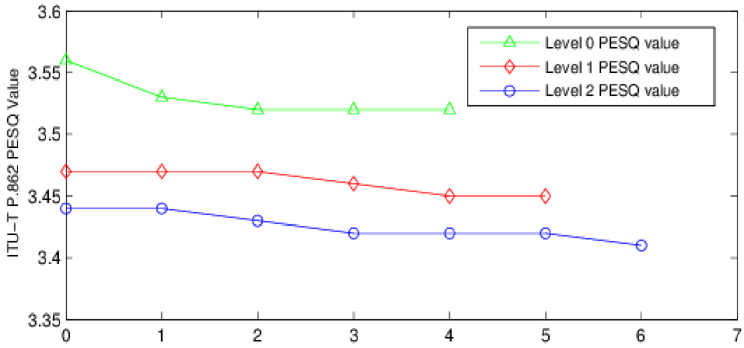

Figure 2. LSB classification according to embedding distortion

3.2 Frame window mechanism in VoIP stream

Unlike static media of image, low bitrate speech frames are used in the form of bit streams in the Internet or other private networks. Considering the shortcomings of matrix coding in low bitrate speeches, the classifying LSB and frame window (CLFW) scheme is proposed for the demand of the uniform embedding impact of LSB. Figure 3 illustrates the overview flow chart of the proposed scheme. In the scheme, the G.723.1 frame is classified into 3 levels and the frames in the speech flow are windowed. As a result, the embedding efficiency and obtained imperceptibility by exploiting matrix coding is greatly increased.

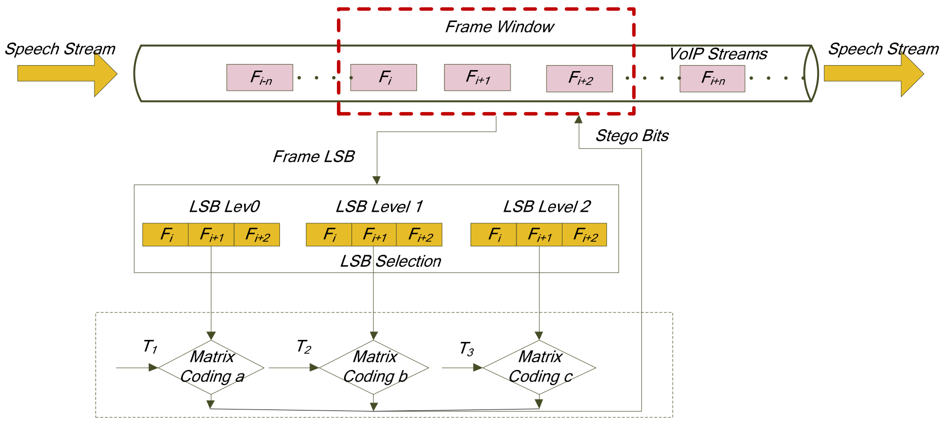

Figure 3. Classifying LSB with matrix coding scheme

The embedding procedure in sending end with CLFW mechanism is summarized into the following 4 steps.

Step 1: Window the speech stream. Fetch numbers (greater or equal to 3 frames to get better performance, yet more latency) of G.723.1 frames according to the window size *n* (here *n*=3) from the RTP stream in VoIP channel and fill them in the frame buffers. Then LSB with similar characteristics are collected.

Step 2: Assemble the classified LSB. Assemble from *n* frames of the LSB according to covert communication requirements to strike a balance between the imperceptibility and capacity.

Step 3: Embed with designated matrix coding methods. Input designated Hamming matrix (a, b or c) and secret message bits (T_1, T_2 or T_3). Then lunch the embedding procedure, where the column order of certain Hamming matrix is used as an embedding key to enhance the security.

Step 4: Write back. Insert the stego cover (frame with secret bits) back into the RTP packet. In the previous work, a covert communication VoIP system is developed to fulfill of reliable transmission.

3.3 CLFW in covert communication system

In Figure 4, CLFW method is applied to the covert communication model in which a number of stego speech frames are packetized into RTP payloads. The stego header is constructed by a reliable covert communication model to ensure correctness of sending and receiving of the secret message, in which stego synchronization signal is embedded by RTP steganographic algorithm. Besides, to fulfill the real-time requirement of communication system the latency caused by the embedding algorithm should be below the threshold between 10 ~ 15ms [17]. In experiment on an Intel Pentium Duo-Core 2.50 Hz computer with 2G DDR2 SD RAM, the processing time is less than 40μs on all levels, which is neglected compared with the provided latency threshold.

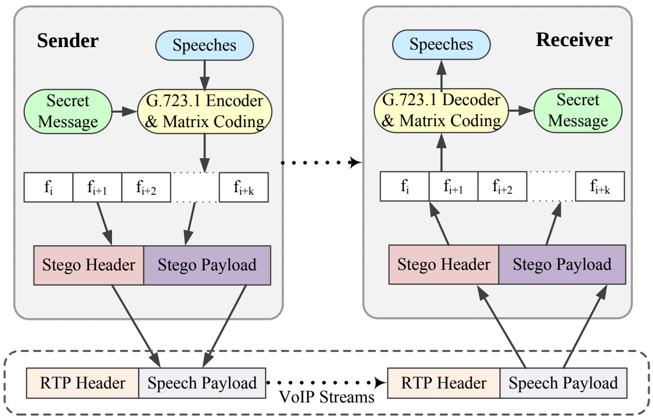

Figure 4. Covert communication model

4. Performance analysis

To evaluate of the proposed method, the imperceptibility and embedding efficiency are discussed. The test speech samples are all collected from Internet free news broadcast and other free English and Chinese language listening test. They consist of four types: English Man (EM), English Woman (EW), Chinese Man (CM) and Chinese Woman (CW). Each speech type contains 200 separate files with 16 bits 8000Hz sampling and 10 seconds length. To make full use of LSB in Table 2 of each level, matrix coding schemes varies accordingly. The matrices are listed in Table 3, where the window size is 3. For level 0 only one matrix is taken, level 1 or level 2 takes two or three matrices, respectively.

Table 3. Matrix coding scheme for each levels

Level	***0***	***1***	***2***
LSB	15 bits	18 bits	21 bits
	COV(1, 15, 4)	*COV*(1, 15, 4)	*COV*(1,15, 4)
Matrix coding	-	*COV*(1, 3, 2)	*COV*(1, 3, 2)
	-	-	*COV*(1, 3, 2)

4.1 Waveform and spectrogram comparison

In order to intuitively evaluate the speech features and the imperceptibility of proposed method, time-domain waveform and spectrogram figures comparisons of speech without hidden message and stego speech using CLFW steganography are given in Figure 5. In the figure a phrase “made in China” is embedded. And the level 0 matrix coding in Table 3 is selected. The upper group is the original speech, and the second group is the stego one. It is seen that the two groups display little distinction.

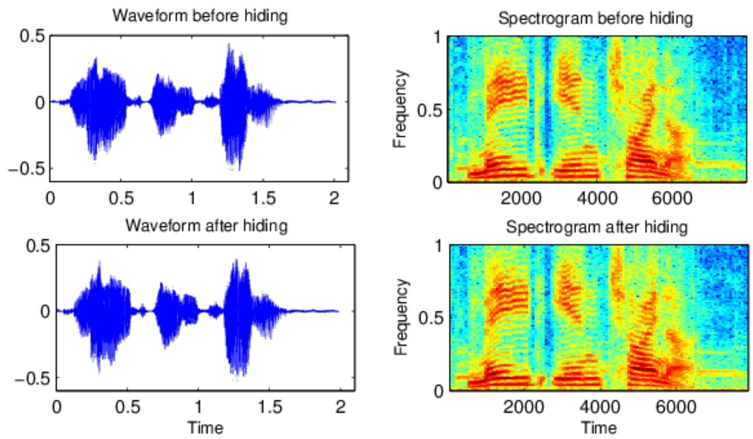

Figure 5. Waveform and spectrogram comparison

4.2 Subjective test

ABX test [18] is an effective subjective evaluation method to distinguish two similar speeches. In ABX test, A, B and X represent respectively the original speech, stego speech and an unknown speech which may be A or B. The subjects listen to X and give a conclusion whether X is A or B. Our tests include four types of speeches and 5 different subjects, the results are listed in Table 4. And it is seen that the average failure percentage of each type is approximately equal to 50%, which reveals that the subjects could hardly distinguish the speeches. As a result, it shows that CLFW gives a perceptual transparent scheme for steganography. In the table level 2 in Table 3 is selected for embedding.

Table 4. ABX test of failure percentage

Subject	***Type* 0**	***Type* 1**	***Type* 2**	***Type* 3**	***Average***
0	37.7%	82.8%	37.7%	37.7%	48.975%
1	82.8%	17.2%	94.5%	98.9%	73.35%
2	82.8%	94.5%	37.7%	82.8%	74.45%
3	62.3%	1.1%	37.7%	82.8%	45.975%
4	62.3%	37.7%	37.7%	37.7%	43.85%
Average	65.58%	46.66%	49.06%	67.98%	57.32%

4.3 PESQ test

The CLFW method is evaluated and compared with simple LSB, Xu’s LSB [8] and direct matrix coding scheme, where “direct” means using matrix coding without any other strategy. The PESQ values of the three LSB methods are evaluated and compared with G.723.1 in Figure 6, where a window with the size 3 is given. The 3 levels include 15, 18 and 21 bits, respectively. They are all hiding with binary Hamming matrix coding scheme *COV*(1, 15, 4). For “direct” method *COV*(1, 3, 2), *COV* (1, 7, 3) and *COV*(1, 15, 4) are chosen respectively for each level. Thus the rest bits of level 1 (3 bits) and level 2 (6 bits) are used for IH header (Figure 4) when RTP header is tempered by eavesdroppers.

In Figure 6, G.723.1 columns indicate the raw codec without steganography compared with the input speech for PESQ test. Then it is seen that in each level of four speech types the PESQ values of the proposed method are remarkably increased compared with other methods. Otherwise, the LSB of three levels can be combined together to gain more capacity, which gives a high embedding efficiency and sustainable speech quality. Note that the three levels combined together in CLFW are not presented in the figure as it has a lower speech quality yet higher capacity.

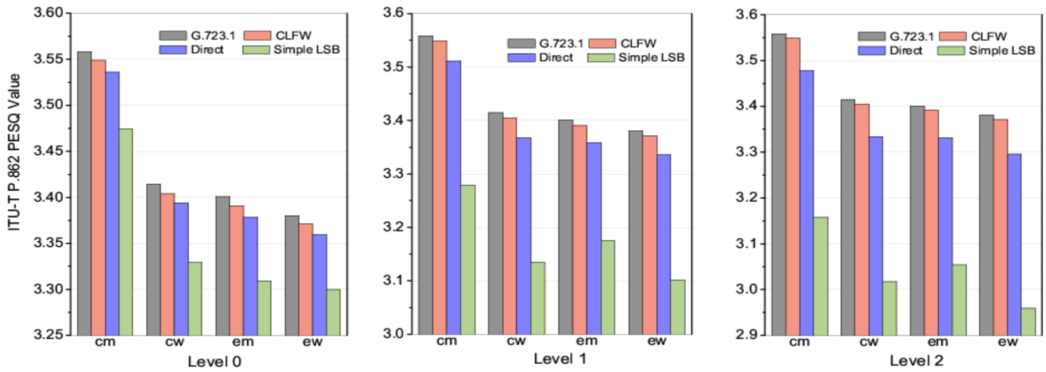

Figure 6. PESQ comparison of proposed scheme

4.4 Embedding efficiency analysis

Embedding efficiency is an important property that influences the security of steganographic scheme. The CLFW scheme is experimented under the covert communication model of Figure 4, where a reliable covert streaming media transmission model is applied. The embedding efficiency of simple LSB, Xu's LSB, direct matrix coding and CLFW are listed in Table 5. It is seen that CLFW increases the embedding efficiency largely compared with simple LSB and Xu's LSB. And it also has higher embedding efficiency than "direct" method. Note that CLFW has the same embedding efficiency with direct matrix coding with *COV*(1, 15, 4), but the former has a better and stable speech quality than the latter according to Figure 6.

Table 5. Embedding efficiency comparison

Method	***Matrix***	***Embedding* Efficiency**
Simple LSB	-	2
Xu's LSB	*COV*(*N*, 2*N*+1,2*N*)	2
Direct matrix coding	*COV*(1, 3, 2)	8/3
Direct matrix coding	*COV*(1, 7, 3)	24/7
Direct matrix coding	*COV*(1, 15, 4)	64/5
CLFW	*COV*(1, 15, 4)	64/5

5. Conclusion and future work

The main contribution of this paper is the introduction of practical matrix coding scenario CLFW in low bitrate speech, which reaches high performance in terms of embedding efficiency and imperceptibility. Otherwise, the experimentation on our covert communication system shows that CLFW has a more stable speech quality which is significant to resist steganalysis. Future work will concentrate on other cyclic syndrome codes for higher embedding efficiency and the combination of other steganographic algorithms such as QIM (Quantization Index Modulation), which is still a challenge to fulfill the practical requirements especially in real-time speech domain.

6. Acknowledgements

This work is supported in part by National Basic Research Program (973 Program) of China under Grant No.2011CB302305, National Key Technology R&D Program under Grant No.2012BAH35F03, Natural Science Foundation of Fujian Province of China under Grant No.2011J05151, and Scientific Research Foundation of National Huaqiao University under Grant No.11BS210.

7. References

[1] Hui Tian, Ke Zhou, Jing Lu, “A VoIP-based Covert Communication Scheme Using Compounded Pseudorandom Sequence”, IJACT: International Journal of Advancements in Computing Technology, vol. 4, no. 1, pp.223-230, 2012.
[2] Chao Wang, Weiming Zhang, Jiufen Liu, Nenghai Yu, “Fast Matrix Embedding by Matrix Extending”, IEEE Transactions on Information Forensics and Security, vol. 7, no. 1, pp.346-350, 2012.
[3] ITU-T. G.723.1, “Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s”, ITU, Geneva, Switzerland, 2006.
[4] Rui Miao, Yongfeng Huang, “An Approach of Covert Communication Based on the Adaptive Steganography Scheme on Voice over IP”, In Proceedings of the 2011 IEEE International Conference on Communications, pp.1-5, 2011.
[5] Yongfeng Huang, Bo Xiao, Honghua Xiao, “Implementation of covert communication based on steganography”, In Proceedings of the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp.1512-1515, 2008.
[6] Hui Tian, Hong Jiang, Ke Zhou, Dan Feng, “Adaptive partial-matching steganography for voice over IP using triple M sequences”, Computer Communications, vol. 34, no.18, pp.2236-2247, 2011.
[7] Hui Tian, Ronglie Guo, Jing Lu, Yonghong Chen, "Implementing Covert Communication over Voice Conversations with Windows Live Messenger", AISS: Advances in Information Sciences and Service Sciences, vol. 4, no. 4, pp.18 ~ 26, 2012.
[8] Tingting Xu, Zheng Yang, “Simple and effective speech steganography in G.723.1 low-rate codes”, In Proceedings of International Conference on Wireless Communications and Signal Processing, pp.1-4, 2009.
[9] Lihua Liu, Mingyu Li, Qiong Li, Yan Liang, “Perceptually transparent information hiding in G.729 bitstream”, In Proceedings of the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp.406-409, 2008.
[10] Jessica Fridrich, David Soukal, “Matrix embedding for large payloads”, IEEE Transactions on Information Forensics and Security, vol. 1, no. 3, pp.390-395, 2006.
[11] Hui Tian, Ke Zhou, Dan Feng, “Dynamic matrix encoding strategy for Voice-over-IP steganography”, Journal of Central South University of Technology, vol. 17, no. 6, pp.1285-1292, 2010.
[12] Yongfeng Huang, Yuan Zhang, Shanyu Tang, “Detection of Covert VoIP Communications using Sliding Window Based Steganalysis”, IET Communications, vol. 5, no. 2, pp.126-133, 2011.
[13] Ron Crandall, “Some notes on steganography”, Available at: “http://whitepapers.hackerjournals.com/wp-content/uploads/2010/03/Some-Notes-on-Steganography1.pdf”, 1998.
[14] Younhee Kim, Zoran Duric, Dana Richards, “Modified matrix encoding technique for minimal distortion steganography”, In Proceedings of 8th International Workshop on Information Hiding, vol. 4437, pp.314-327, 2007.
[15] Rongyue Zhang, Vasiliy Sachnev, Hyoung Joong Kim, “Fast BCH syndrome coding for steganography”, vol. 5806, pp.48-58, 2009.
[16] ITU-T 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”, ITU, Geneva, Switzerland, 2001.
[17] ITU-T G.114, “One-way Transmission Time, SERIES G: Transmission Systems and Media, Digital System and Networks”, ITU, Geneva, Switzerland, 2003.
[18] Zheming Lu, Bin Yan, Shenghe Sun, “Watermarking combined with CELP speech coding for authentication”, IEICE Transactions on Information and Systems, vol. 88, no. 2, pp.330-334, 2005.