Least-significant-digit Steganography in Low Bitrate Speech

Jin Liu and Ke Zhou*
School of Computer,
Wuhan National Laboratory for Optoelectronics,
Huazhong University of Science and Technology, Wuhan, China
Email: geneleocn@gmail.com, k.zhou@hust.edu.cn

Hui Tian
College of Computer Science and Technology,
National Huaqiao University, Xiamen, China
Email: cshtian@gmail.com

***Abstract*—For steganography over speech frames, least-significant-bit (LSB) approach has been one of the most important alternatives. However, its embedded capacity is quite limited, due to the low redundancy characteristic of low bitrate speech. Thus, we suggest a novel least significant digit (LSD) method in this paper, which makes full use of the frame bits to hide secret messages and provides a larger embedding capacity than the LSB method. The LSD method exploits the multiple adjacent states of frame parameters, which are produced by multiple modifications (e.g. $+1, -1, +2, -2$) and encoded as LSDs using a multi-ary numeration system. Beyond providing a considerable embedding capacity, the LSD method further extends the key space for key-based steganography, which can enhance the security of covert communication. The LSD method is implemented and evaluated with G.723.1 as the codec for speeches. The experiment results show that, compared with the previous LSB method, the LSD method increases around 30% of embedding capacity, and induces less distortion given the same embedding capacity.**

***Index Terms*—Information Hiding, Least Significant Digit, Low Bitrate Speech, Multi-ary Numeration, Steganography.**

I. Introduction

Among all applications of information hiding, steganography in speech streams is one of the most promising development branches, which could be widely used in today's communication area such as VoIP. However, due to communication bandwidth limits, the speech frames always contain little redundancies. That gives difficulties for covert communication, especially the speech covers. Prior hiding algorithms mainly concentrate on covert communication models [1], [2], other feasible and effective algorithm [3] uses quantization index modulation method with little distortion, and [4] gives a spectrum speech hiding scenario. Nevertheless, due to the confinement of communication latency, speech quality distortion and hiding rates (covert transmission rates), the LSB method is still the most effective and popular one [5].

Ito et al. [6] proposes a novel LSB substitution method to gain a high hiding rate. The stego cover is G.711 speech which also has high redundancies. However, more low bitrate speech codec with less redundancies such as G.729, G.723.1, GSM, which are more widely used in VoIP and wireless phones, need to be considered. Liu [7] provides an algorithm for steganography in G.729 LSBs yet with little embedding capacity. In [5] a silent frame hiding approach is proposed for G.723.1 codec, which takes account of a large number of silent frames in normal speech communication over network. It obtains 101 bits suitable for embedding per frame with the fulfillment of little distortion. Other stego schemes [8], [9] focus on stego coding scenario in order to improve the imperceptibility, which would further decrease the embedding capacity. In other words, we could provide more embedding bits to exchange for more imperceptibility.

In LSB steganography of low bitrate speech frames, distortions caused by embedding modification vary from every frame bits. In LSB substitution method only the operations of $0 \rightarrow 1$ and $1 \rightarrow 0$ of the LSB are applied. More adjacent states of the parameter bits are not considered which could gain more hiding possibilities. LSB matching method [10] uses the "$+1$" , "-1" and "0" operations to decrease the statistic characterize of stego object, which could be less vulnerable to steganalysts [11], [12]. Zhang [13] proposes a (2n+1)-ary steganograhpy method in order to get high embedding efficiency in images with the "extra" states. They both use the bits states as an alternative modification to increase the imperceptibility, while all the states can also be utilized to enhance the embedding capacity.

In the paper we propose a least significant digit (LSD) method in common used G.723.1 speech frames. In LSD steganography not only the LSBs but also the least significant states are exploited. The frame parameters and secret messages are both encoded with the multi-ary numeration systems (e.g. ternary, 5-ary numeration), then the LSDs of the multi-ary numbers are obtained. The method takes full advantage of the bit states of each frame parameters to gain more hiding bits with lowest distortion, where ternary and 5-ary numerations are presented to embed $\lfloor log_2^{(2n+1)} \rfloor$ bits with n bits LSBs. Otherwise, more embedding bits give a chance to utilize a larger key space for the map of embedding function in key-based stego systems to enhance the embedding security.

We give a brief introduction of G.723.1 codec and multi-ary

*Corresponding author

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

978-1-4577-2053-6/12/$31.00 ©2012 IEEE

coding in section II, followed by LSB analysis of parameters in G.723.1 frames in section III. Then section IV describes details of the LSD method. The experiment evaluates the LSD method by means of speech quality evaluation criterion in section V. Section VI concludes the whole paper in the end.

II. Multi-ary numeration and G.723.1 codec

A. Multi-ary method

Binary numeration is the most widely used coding method in computer science. It has two states representing binary digits 0 and 1 respectively. Other multi-ary coding methods (including 3-ary (ternary), 5-ary and etc.) may be exploited for some practical use. In the paper we introduce a multi-ary encoding method to increase the performance of steganography. In an x-ary numeration system, one x-ary digit has x states which can be described by a basic digit set $B = \{0, 1, 2, \ldots, x-1\}$. Let $(T)_x$ represents an x-ary number T. Then we can easily translate a multi-ary number into a general decimal number according to Equation (1), where $(T)_x = (T_{m-1}, T_{m-2}, \cdots, T_0)$ and $T_i \in B$ represent an m digits of x-ary number and one x-ary digit respectively.

$$(T)_{10} = T_{m-1} \times x^{m-1} + \cdots + T_i \times x^i + \cdots + T_0 \times x^0 \quad (1)$$

Table I gives a description of the relationships among kinds of multi-ary numbers, where the value scope of an x-ary number with $(m)_{10}$ digits belongs to $[0, (x^{m-1})_{10}]$.

TABLE I: Translation among mulit-ary numbers

Decimal	Binary	Ternary	5-ary
0	0000	000	00
1	0001	001	01
2	0010	002	02
3	0011	010	03
4	0100	011	04
5	0101	012	10
6	0110	020	11
7	0111	021	12
8	1000	022	13
9	1001	100	14

We can then easily deduce from Equation (1) that an $(m)_{10}$ digits of x-ary number can be translated into a y-ary number with at most $(n)_{10}$ digits according to Equation (2). Typically an $(m)_{10}$ bits of binary number can be translated into a ternary number with maximal $\lceil m \log_3^2 \rceil$ digits.

$$(n)_{10} = \lceil m \log_y^x \rceil \quad (2)$$

B. G.723.1 codec

G.723.1 [14] is an ITU-T 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 (e.g. VoIP). This paper takes the widely used low bitrate speech codec G.723.1 with 6.3 kbit/s bitrate as an example. In VoIP communication speech with G.723.1 coding is transmitting in the form of frames with 24 bytes. And each frame is composed by several parameters listed in Table II. The two bits represented separately for bitrate (RF) and validation indication (VF) are not presented in order to ensure the correct transmission and reception for communication, and should not be used for steganography. The four sub-frames of the parameter (e.g. lpc) are noted as lpc0, lpc1, lpc2 and lpc3 respectively.

TABLE II: G.723.1 bit allocation with 6.3kbit/s

Frame parameters	**Sub-frame 0**	**Sub-frame 1**	**Sub-frame 2**	**Sub-frame 3**	**Total**
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					189

G.723.1 is a parameter coding speech codec. It contains six main parameters noted as lpc, acl, gain, pos, sig and grid. And they have different noise resistant features, which give different LSB characteristics of themselves. For a better performance of LSD multi-ary embedding method in the speech frame, only a part of them are selected.

III. LSBs of G.723.1 frame

In order to test the noise resistant features of all parameter bits in G.723.1 frame and get the LSBs or LSDs of speech frame, we flip the 189 frame bits (exclude the 2 flag bits) one by one and evaluate the speech quality. ITU-T PESQ [15] method is more close to human perception than other evaluation criteria for objective speech quality evaluation. Then we adopt the PESQ criterion to evaluate the G.723.1 6.3 kbit/s frame bits in Fig. 1, where 800 of Chinese man, Chinese woman, English man and English woman speeches are processed respectively. And each speech has a duration of 10 seconds.

According to Fig. 1, we obtain the LSDs (LSBs for binary) of G.723.1 speech frame which are distributed in the frame parameters. And LSBs (include 14 bits) with nice noise resistant (high speech quality after embedding) features in Table III are selected for the LSD method, which have PESQ values greater or equal to 3.4.

TABLE III: LSB distribution of speech frame

Parameter	LSB (bits)	Parameter	LSB (bits)
lpc	2	grid (0∼3)	1
gain (0∼3)	1	pPos (0∼3)	1

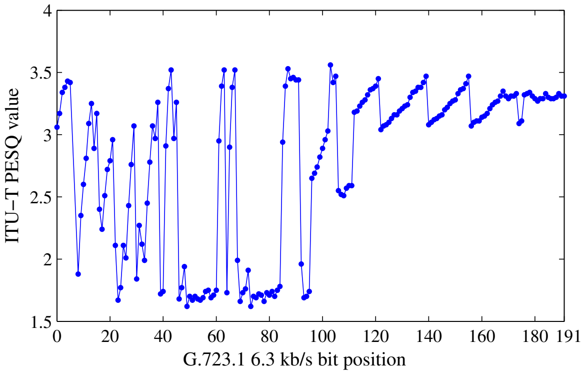


Fig. 1: G.723.1 steganography PESQ evaluation

For the application of ternary embedding two bits are minimally needed. The four grid parameters contain only one bit respectively, which are already fully used for hiding (no extra state). Consequently, they are not capable for multi-ary embedding (available for LSB substitution only). The appropriate parameters include lpc, gain0, gain1, gain2, gain3, pPos0, pPos1, pPos2 and pPos3. For the demand of more embedding capacity, other bits with PESQ values below 3.4 in the figure would be included.

IV. LSD STEGANOGRAPHY

A. Bits modification of LSB substitution

Let $C = \{c_{n-1}, \dots, c_i, \dots, c_0 \mid c_i = 0 \ or \ 1\}$ denotes an n bits of frame parameter cover in binary numeration, and $M = \{m_{k-1}, \dots, m_i, \dots, m_0 \mid m_i = 0 \ or \ 1\}$ denotes a k bits of secret message. In primitive LSB substitution method, the embedding operation is depicted by Equation (3). And for k bits LSB substitution, the embedding function is represented by Equation (4).

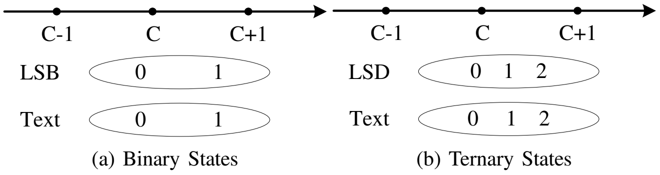
$$C' = \sum_{i=1}^{n-1} c_i \cdot 2^i + m_j \tag{3}$$

$$C'' = \sum_{i=k}^{n-1} c_i \cdot 2^i + \sum_{j=0}^{k-1} m_j \cdot 2^j \tag{4}$$

where C' and C'' denote the stego frame parameters.

In above, the translation of $1 \rightarrow 0$ or $0 \rightarrow 1$ is applied. When only one bit of LSB is considered, the cover parameter may be modified into its two adjacent states $C-1$ (e.g. $1 \rightarrow 0$) and $C+1$ (e.g. $0 \rightarrow 1$) (See Fig. 2a). In binary mode, the LSB of C' has only two states $\{0, 1\}$, that is $\text{LSB}(C') \in \{0, 1\}$. The embedding and extracting functions are bijections from set $\{0, 1\}$ to set $\{0, 1\}$. And the LSBs of $C+1$ and $C-1$ are both equal to 1 or 0. As a result, the modification of k bits of LSB substitution only takes two digit states into account and brings k bits of capacity.

When we take ternary numeration into account, the three states $\{C, C+1, C-1\}$ are mapped to ternary $\{0, 1, 2\}$ by a bijection. The three states represent a ternary digit denoted as LSD, and $\text{LSD}(C') \in \{0, 1, 2\}$. Fig. 2b depicts this condition.



(a) Binary States

(b) Ternary States

Fig. 2: State mapping of LSB substitution

Then the embedding capacity for modification are extended to three states, aka a ternary digit. The embedding and extracting functions are converted to maps from set $\{0, 1, 2\}$ to set $\{0, 1, 2\}$ with one modification of LSD of the frame parameter. And the maximum capacity are increased from 1 bit to $\log_2^3$ bits.

B. Multi-ary embedding of LSD

From above analysis, we can deduce that for a k bits LSB substitution, the embedding and extracting functions are bijections from $\{0, 1, \dots, 2^k - 1\}$ to themselves. And when we take all the states into account, the embedding and retrieving functions can be extended to bijections from $\{0, 1, \dots, 2^k - 1, 2^k\}$ to themselves. Consequently, we use the $2^k + 1$ basic digits in a $(2^k + 1)$-ary numeration system to represent $\{0, 1, \dots, 2^k - 1, 2^k\}$. And the embedding capacity can be extended from k bits to $\log_2^{2^k+1}$ bits. In the paper we make use of three kinds of numerations for different parameters listed in Table IV. The binary embedding of "gain" parameter is due to the length limits of these parameters.

TABLE IV: Numeration of parameters

Parameter	Numeration	LSD (bits)
lpc	5-ary	$\log_2^5$
gain (0∼3)	ternary	$\log_2^3$
grid (0∼3)	binary	1
pPos (0∼3)	ternary	$\log_2^3$

Fig. 3 gives an evaluation of four modification operations of the selected bits in Table III of G.723.1 frame, where all bits reveal nearly equal features of the four operations except for the "-1" and "-2" operation of four "gain" parameter bits. Alternatively, the rest 2 plus operations upon "gain" bits also show good quality. As a result, we could choose multiple states (including the original state) of each frame parameter listed in Table V for a multi-ary LSD embedding, where one 5-ary and 8 ternary embedding operations are applied, and the "grid" operation is LSB substitution.

Then we can embed and extract the secret message in speech frames according to Equation (5).

$$(C')_x = \sum_{i=1}^{n'-1} d_i \cdot x^i + o(l_j) \tag{5}$$

Where $(C')_x = \{d_{n'-1}, \dots, d_1, d_0 \mid d_i \in \{0, 1, \dots, x - 1\}\}$denotes the x-ary format of stego parameter, and $(M)_3 =$

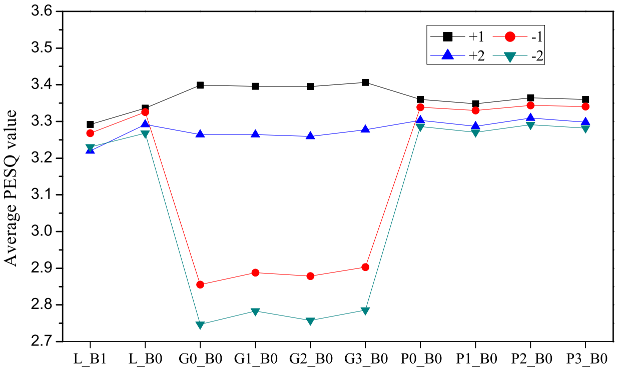


Fig. 3: Modification evaluation of selected frame bits

TABLE V: The states of frame parameters

State	lpc	gain (0∼3)	grid (0∼3)	pPos (0∼3)
0	0	0	0	0
1	+1	+1	+/− 1	+1
2	−1	+2	-	−1
3	+2	-	-	-
4	−2	-	-	-

$\{l_{k'-1}, \ldots, l_1, l_0 \mid l_j \in \{0, 1, \ldots, x-1\}\}$ denotes the secret message in x-ary numeration mode. $o(x)$ is an operation in Table V decided by the secret LSD states and the predefined maps.

C. The embedding and extracting procedure

In the embedding procedure of LSD method, the frame parameters should be translated into multi-ary numbers first respectively. The translations are independent from each parameter. Consequently, the secret message are translated into different numerations correspondingly. In Fig. 4, binary bits are firstly embedded by LSB substitution method. Then a number of designated bits are taken out for ternary embedding after a binary-to-ternary translation. And next a 5-ary embedding operation is done in the same way. After that, all stego parameters are translated back into binary numeration for speech transmission.

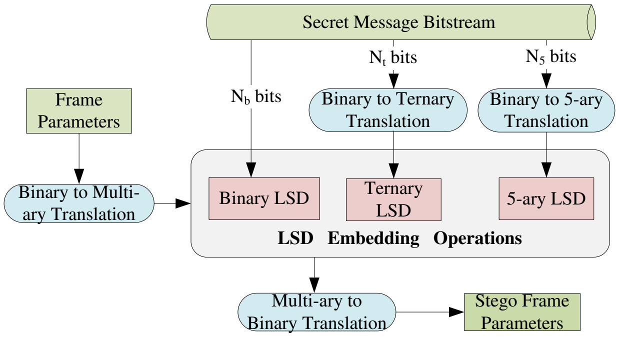


Fig. 4: Embedding procedure of LSD steganography

Note that the LSD embedding is not belong to a substitution method which introduces more distortions than nearest arithmetic $\pm n$ operation. In the figure, binary to multi-ary translation are short for kinds of binary to ternary or binary to 5-ary translation, and so is the multi-ary to binary translation. N_b, N_t and N_5 represent the embedded bits of binary, ternary and 5-ary LSD embedding respectively. They must be confirmed before the binary to multi-ary translation of secret message. According to Table IV, the LSDs of binary, ternary and 5-ary include 4, 8 and 1 corresponding multi-ary digits respectively. Then we can figure out N_b, N_t and N_5 according to Equation (2).

$$N_b = \lfloor 4 \times \log_2^2 \rfloor = 4\ (bits)$$
$$N_t = \lfloor 8 \times \log_2^3 \rfloor = 12\ (bits)$$
$$N_5 = \lfloor 1 \times \log_2^5 \rfloor = 2\ (bits)$$

The extracting procedure is relatively easy, we only extract the LSDs of the stego frame parameters according to the numeration adopted in the embedding algorithm. And then translate it into binary bits, Fig. 5 is the flow chart for extracting.

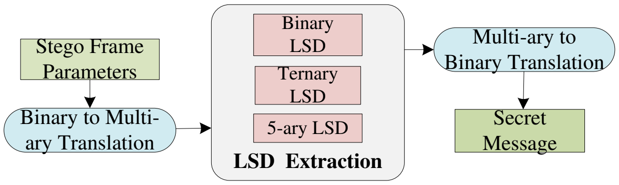


Fig. 5: Extracting procedure of LSD steganography

V. Performance analysis

According to previous analysis, the ternary and 5-ary LSD method could embed $\log_2^3$ and $\log_2^5$ bits respectively. Considering the selected 14 cover bits of G.723.1 frame, the embedding capacity increased about 28% to 18 bits per frame. In order to increase the embedding capacity, we generally increase the selected cover bits in one frame, which leads to a worse speech quality compared with LSD method due to the less noise resistance bits adopted. Otherwise, for the introduction of encryption mechanism in steganography, the key space increase from 2^{14} of LSB method to $(2^4 \cdot (P_3^2)^8 \cdot (P_5^4) \approx 1.50 \times 2^{31})$ of LSD method, which enhances the hiding security largely for speech transmission.

A. Spectrogram comparison

Compared with LSB method with same selected cover bits (14 bits LSB), ternary and 5-ary embedding in LSD method increase the operations of modification. And it also brings more capacity as well as more distortions. Nevertheless, the increased capacity are based on bits with high noise resistance (high PESQ value). Consequently, LSD embedding show more similarity with original speech and better speech quality than LSB substitution with same embedding capacity (18 bits LSB). Fig. 6 shows the speech spectrogram comparison of original speech (G.723.1), 14 bits LSB substitution, LSD method and 18 bits LSB substitution. The figure describes the spectrogram of sentence ”Made in China” of a man’s voice with 100% embedding rate. It reveals that 14 bits LSB and LSD method both have better similarities with G.723.1 than 18 bits LSB embedding. And the LSD spectrogram shows little visible difference with 14 bits LSB method, which means the introduction of little extra distortion.

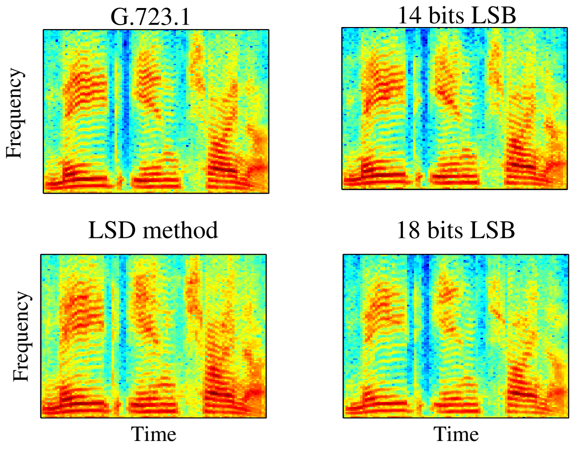


Fig. 6: Spectrogram comparison

B. PESQ evaluation

In order to evaluate the speech quality of LSD embedding objectively and compare it with previous LSB substitution method, we test the PESQ values of them in Fig. 7. In the figure a PESQ value of G.723.1 coding is given for comparison, and CM, CW, EM, EW are short for Chinese woman, Chinese man, English man and English woman speeches respectively with a duration of 10s among 200 repeated tests. It is obviously that the LSD method makes little degradation of PESQ value compared with 14 bits LSB method, and the two embedding methods both have better speech qualities than 18 bits LSB method. Note that there is a big degradation of all these three methods compared with the G.723.1 coding, that’s because of the 100% embedding rate adopted with all the selected bits. We could also use a lower embedding rate with a part of LSDs or embed with a stego coding method to gain more imperceptibility.

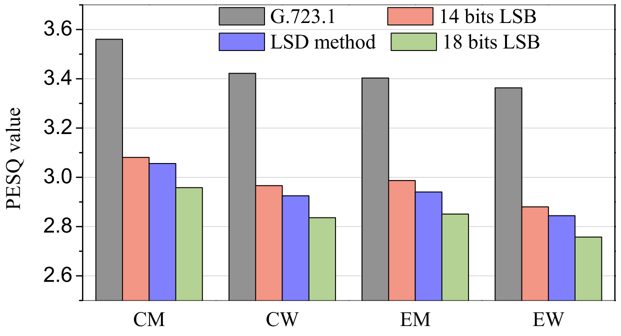


Fig. 7: PESQ value comparison

VI. Conclusion

This paper gave a new LSD method to take full advantage of low bitrate speech frame parameter bits. The LSD method extended the traditional LSB substitution steganography by the introduction more embedding states. Then a multi-ary numeration mechanism was adopted for the translation of the states to LSDs. All the arithmetic operations of embedding modification were applied using frame bits with high noise resistant characteristic, which introduce less distortion than LSB substitution under a given embedding capacity. Otherwise, the enlarged embedding capacity in one frame could also be used to exchange with imperceptibility and gave a larger key space for key-based steganographic method in order to increase the hiding security. The LSD method was independent from a specific cover, it also can be extended to other stego schemes.

References

[1] W. Mazurczyk and K. Szczypiorski, “Steganography of voip streams,” in *Proceedings of Lecture Notes in Computer Science*, vol. 5332, no. 2, 2008, pp. 1001 – 1018.

[2] H. Tian, K. Zhou, H. Jiang, J. Liu, Y. Huang, and D. Feng, “An m-sequence based steganography model for voice over ip,” in *Proceedings of IEEE International Conference on Communications*, Piscaatway, NJ, USA, 2009, pp. 1–5.

[3] B. Xiao, Y. Huang, and S. Tang, “An approach to information hiding in low bit-rate speech stream,” in *Proceedings of 2008 IEEE Global Telecommunications Conference*, Piscataway, NJ, USA, 2008, pp. 1940–1944.

[4] F. Djebbar, H. Hamam, K. Abed-Meraim, and D. Guerchi, “Controlled distortion for high capacity data-in-speech spectrum steganography,” in *Proceedings of Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Los Alamitos, CA, USA, 2010, pp. 212–215.

[5] Y. 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. 96, no. 99, pp. 296–306, 2011.

[6] A. Ito, S. Abe, and Y. Suzuki, “Information hiding for g.711 speech based on substitution of least significant bits and estimation of tolerable distortion,” *Ieice Transactions on Fundamentals of Electronics Communications and Computer Sciences*, vol. 93, no. 7, pp. 1279–1286, 2010.

[7] L. Liu, M. Li, Q. Li, and Y. Liang, “Perceptually transparent information hiding in g.729 bitstream,” in *Proceedings of International Conference on Intelligent Information Hiding and Multiedia Signal Processing*, 2008, pp. 406–409.

[8] R. Zhang, V. Sachnev, and H. J. Kim, “Fast bch syndrome coding for steganography,” vol. 5806, Darmstadt, Germany, 2009, pp. 48–58.

[9] W. Zhang, X. Zhang, and S. Wang, “Maximizing steganographic embedding efficiency by combining hamming codes and wet paper codes,” in *Proceedings of 10th International Workshop on Information Hiding*, vol. 5284, 2008, pp. 60–71.

[10] J. Mielikainen, “Lsb matching revisited,” *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 285–287, 2006.

[11] M. Goljan, J. Fridrich, and T. Holotyak, “New blind steganalysis and its implications,” in *Security, Steganography, and Watermarking of Multimedia Contents VIII*, ser. Proceedings of SPIE - The International Society for Optical Engineering, vol. 6072. SPIE, 2006.

[12] Y. Huang, S. Tang, C. Bao, and J. Yip, “Steganalysis of compressed speech to detect covert voip channels,” *IET Information Security, IEE/IEEE Journal*, vol. 5, no. 1, pp. 26–32, 2011.

[13] X. Zhang and S. Wang, “Efficient steganographic embedding by exploiting modification direction,” *IEEE Communications Letters*, vol. 10, no. 11, pp. 781–783, 2006.

[14] *G.723.1 Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*, ITU-T Std., 2006.

[15] *P.862 Perceptual evaluation of speech quality (PESQ)*, ITU-T Std., 2001.