Signal Processing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Optimal matrix embedding for Voice-over-IP steganography

Hui Tian [a,*], Jie Qin [a], Yongfeng Huang [b], Yonghong Chen [a], Tian Wang [a], Jin Liu [a], Yiqiao Cai [a]

[a] College of Computer Science and Technology, National Huaqiao University, Xiamen 361021, China
[b] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 17 November 2014
Received in revised form
6 April 2015
Accepted 2 May 2015

Keywords:
Steganography
Information hiding
Voice over IP
Adjustable matrix encoding
Optimal embedding

ABSTRACT

Divide-and-rule strategy is popularly adopted in steganography based on Voice-over-IP (VoIP), which often divides a cover into many small parts and performs embedding operation on each one to maintain the real-time requirement of VoIP services. No existing study, however, notices that the cover length of the parts may seriously affect the embedding performance. The goal of this paper is to fill in the gap and present a novel steganographic scheme to achieve the optimal embedding. To attain the goal, we first present an adjustable matrix encoding (AME) approach, which can adaptively generate a guide matrix to accommodate to the given cover length and provide the best performance while guaranteeing the desired embedding capacity. Further, the optimal embedding is modeled as a typical combinatorial optimization problem, of which the optimal solution is an AME or a combination of multiple AMEs. We evaluate the proposed scheme with ITU-T G. 711 (A-law) as the codec of the cover speech and compare it with previous methods. The experimental results demonstrate that the proposed scheme is really feasible in both theory and practice, and can provide consistently optimal embedding performance in any case.

1. Introduction

Steganography is an art and science of concealing secret messages within seemingly innocent carriers, which can be employed to achieve covert communications. It has a long history dating back to 440 BC [1,2], and has blossomed into a significant security technique with the development of information and communication technologies over the last years. Up to now, extensive researches on steganography have been carried out, and steganographic covers have been also extended from initial images to almost all multimedia [3–5]. More recently, a novel dynamic steganographic cover, Voice over Internet Protocol (VoIP, also named IP telephony), has attracted increasing attention [6–12], which possesses three main advantages over traditional storage media like images [11,12]. First, the real-time nature of VoIP provides better security for secret messages by virtue of its instantaneity, because it gives eavesdroppers almost no time to discover possible abnormity owing to concealed messages. Second, VoIP can be considered a multi-dimensional carrier in that both the packet protocol headers and the payload data can be used to conceal information. Third, the duration of a VoIP conversation is dynamic and variable, so it can offer adequate cover data according to the requirement of covert communication. In virtue of these advanced characteristics, VoIP-based steganography can provide a sophisticated solution for secure communication.

In recent years, research on VoIP-based steganography has proliferated, and various steganographic approaches have developed. For more details we refer the interested reader to an up-to-date comprehensive survey given by Mazurczyk [9]. Generally, they fall into two major categories: steganographic approaches based on protocol

* Corresponding author. Tel.: +86 18959267996; fax: +86 592 6162556.
E-mail address: cshtian@gmail.com (H. Tian).

http://dx.doi.org/10.1016/j.sigpro.2015.05.001
0165-1684/

steganography and the ones based on payload modification. The first category utilizes the specific protocols of VoIP as the carriers and embeds secret messages into unused or optional fields of protocol headers [13–15] or by modulating inter-packet times (packet rates) [16–18], while the other one employs the payloads as the carriers and conceals secret messages into the redundant parts of digital speech signals. In contrast, most researchers mainly focus on the latter category, and have proposed many successful techniques, such as Least-Significant-Bit (LSB) steganography [19–22], quantization-index-modulation (QIM) steganography [23,24], echo-based steganography [25,26], inactive-frame steganography [6] and transcoding steganography [27]. Among them, because of its relatively high capacity and low complexity, LSB steganography has gained the most attention [8,9,11]. However, it has been being faced with an important problem, how to improve its steganographic security, particularly steganographic transparency, which a lot of research effort has been devoted for. For example, Huang et al. [28] attempted to introduce LSB matching steganography to the VoIP-based covert communication; Xu et al. [29] proposed an adaptive approach to reduce the embedding distortion by randomly choosing both cover packets and embedding intervals between cover LSBs; Miao et al. [30] presented an adaptive steganography scheme to improve the security by choosing lower embedding bit rates in the flat blocks and higher embedding bit rates in the sharp blocks.

Although the approaches can reduce the distortion of the cover and thereby enhance the steganographic transparency, they inevitably lead to the decrease of the embedding capacity. That is to say, the steganographic transparency is often enhanced at the cost of decreasing the embedding capacity. In spite of this, it is perfectly possible to strike a tradeoff between them, which has been proved by our previous work [11,12]. We achieved this goal by reducing effective bit-change rate (EBCR) exploiting transparency-orientated encoding strategies.

Nevertheless, this problem needs to be further gone into, because the previous studies did not take into account the impact that would be induced by the "divide and rule" strategy popularly adopted in practice for maintaining the real-time requirement of VoIP. So-called "divide and rule" strategy means that the LSBs of the cover are divided into many small parts and the embedding process is based on each part as a unit. In such a condition, the cover length of each part may affect the steganographic performance. For instance, supposing that the length of secret messages to be embedded into each part is 4 bits, let us consider the following embedding situations. If the length of each cover part (denote by L_{cp}) is also 4 bits, simple LSB substitution, apparently, is the best and only choice. If L_{cp} is no less than 15, matrix encoding (ME) [31,32] is the best one, because it can hide secret messages with no more than one bit changed. However, if L_{cp} is between 5 and 14, there are many embedding schemes. Taking $L_{cp}=10$ for example, Table 1 shows some possible ones through combining the direct LSB substitution and the ME. In this case, there is another significant problem – which is the one with the best steganographic performance. Therefore, in this paper, we are motivated to further study VoIP-based steganography with the divide-and-rule strategy and devoted to achieving the optimal embedding.

Table 1
Some possible embedding schemes for $L_{cp}=10$.

No.	ME (2, 3)	ME (3, 7)	RLSB(x, y)
1	–	–	$1\times$RLSB(4, 10)
2	$2\times$ME (2, 3)	–	–
3	$1\times$ME (2, 3)	–	$1\times$RLSB(2, 7)
4	–	$1\times$ME (3, 7)	$1\times$RLSB(1, 3)

Note: (1) $n\times$ME (x, y) means embedding x bits of secret messages into y bits of covers with the matrix encoding (ME) and using this strategy for n times; (2) $n\times$RLSB(x, y) means embedding x bits of secret messages into y bits of cover randomly selected in each part with the direct LSB substitution and using this strategy for n times; (3) "–" means "not adopted."

To attain this goal, we first present an adjustable matrix encoding (AME), which can support the alteration of cover lengths and provide the best performance while guaranteeing the desired embedding capacity. Further, the optimal embedding is considered as a typical combinatorial optimization problem, of which the optimal solution is an AME or a combination of multiple AMEs. It has been demonstrated by theoretical analysis and experimental tests that the proposed approach can always obtain the optimal steganographic performance successfully.

The remainder of this paper is organized as follows. Section 2 first presents the AME, and Section 3 provides a general steganographic scheme using this encoding. The optimal embedding scheme based on AME is described in Section 4, followed by the evaluation of the proposed scheme and its test results that are presented in Section 5. Finally, concluding remarks are given in Section 6.

2. Adjustable matrix encoding (AME)

2.1. Motivation

The matrix encoding (ME) was first discovered by Crandall [31], and popularized by Westfeld who incorporated a typical implementation using binary Hamming codes in his F5 algorithm [32]. Further, some researchers extended the ME from Hamming codes into many other linear codes [33,34], such as, BCH [35], Reed-Solomon [36] and Trellis codes [37]. However, all the matrix encoding approaches share a common steganographic scheme as follows [33].

For an $[n, k]$ binary code $\mathcal{C}$ with a parity check matrix $\mathbf{H}$ and covering radius R, the steganographic scheme below can embedded $n-k$ bits of secret messages $\mathbf{m}=(m_1, m_2, \ldots, m_{n-k})^{\mathrm{T}}$ into n bits cover $\mathbf{c}=(c_1, c_2, \ldots, c_n)^{\mathrm{T}}$ via at most R changes

$$\varphi(\mathbf{c}, \mathbf{m}) = \mathbf{c} + e(\mathbf{m} - \mathbf{Hc}) = \mathbf{c}' \quad (1)$$

$$\begin{aligned}\psi(\mathbf{c}') &= \mathbf{Hc}' \\ &= \mathbf{Hc} + \mathbf{H}e(\mathbf{m} - \mathbf{Hc}) \\ &= \mathbf{Hc} + \mathbf{m} - \mathbf{Hc} \\ &= \mathbf{m}\end{aligned} \quad (2)$$

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

where $\varphi(\mathbf{c}, \mathbf{m})$ and $\psi(\mathbf{c}')$ are embedding and extraction functions respectively, $e(\mathbf{m}-\mathbf{Hc})$ is a coset leader of the coset $\mathcal{C}(\mathbf{m}-\mathbf{Hc})=\{\mathbf{x} \in GF^n(2) | \mathbf{Hx}=\mathbf{m}-\mathbf{Hc}\}$, and $\mathbf{c}'=(c_1', c_2', \ldots, c_n')^T$ is the steganographic form of cover $\mathbf{c}$. Let $d(\mathbf{x}, \mathbf{y})$ denote the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$, and $w(\mathbf{x})$ denote the Hamming weight of $\mathbf{x}$. Since the binary code $\mathcal{C}$ has the covering radius R, we can get that $d(\mathbf{c}, \mathbf{c}')=w(e(\mathbf{m}-\mathbf{Hc}))\le R$, which reveals that the distortion bound of the above scheme is R.

As is well known, the covering radius of Hamming codes is 1, so the ME based on Hamming codes can hide r bits into 2^r-1 cover bits with at most one bit changed, which shows that it offers the best transparency to some degree. In addition, because the sizes of the parity check matrices of Hamming codes are relatively small, it possesses the lowest complexity in all matrix encodings. Therefore, it is often regarded as the standard ME and applied most popularly in steganography [38–40].

However, the standard ME cannot always achieve the best steganographic performance. For example, if 4 bits of secret messages are needed to be embedded into 7 bits of cover messages, ME(4, 15) cannot be applied, since there are not enough cover messages; a possible alternative is to carry out ME(2, 3) twice, which, however, will leave one-bit cover message without being employed and thereby downgrade the steganographic performance. The primary reason for this phenomenon is the strict requirement in the standard ME that concealing r bits of secret messages needs 2^r-1 cover bits. In other words, the standard ME cannot adapt to arbitrary cover lengths, which limits its sphere of application. Thus, we are motivated to present an adjustable matrix encoding, which can accommodate to the various cover lengths in practice, and achieve the optimal embedding performance.

2.2. Construction of AME

Plainly, the so-called matrix encoding essentially employs the parity check matrix $\mathbf{H}$ of the given code to guide the steganographic process, so we prefer to call $\mathbf{H}$ guide matrix for short in this paper. The primary idea behind the AME is to design a way of guide matrix generation to suit the given cover length while maintaining an optimal embedding performance. To achieve this goal, we must first determine the minimum size of the guide matrix for AME. If r bits of secret messages need to be embedded into n $(n<2^r-1)$ bits of cover messages, it is inescapably clear that more distortion should be allowed. However, we must also maintain good embedding performance, so the distortion bound should be as small as possible – if "1" is unattainable, "2" is the limit. Based on the considerations, we design a construction method of the guide matrix with the distortion bound not more than two using two Hamming parity check matrices, as described in Theorem 1.

Theorem 1. *Guide matrix with the distortion bound not more than two. Let $\boldsymbol{H}_1$ be the parity check matrix of an $[n_1, n_1-r_1]$ Hamming code $\mathcal{C}_1$, and $\boldsymbol{H}_2$ be the parity check matrix of an $[n_2, n_2-r_2]$ Hamming code $\mathcal{C}_2$. The guide matrix $\boldsymbol{H}$ with the distortion bound not more than two can be obtained as follows:*

$$\mathbf{H}=\begin{bmatrix}\mathbf{H}_1 & \mathbf{0} & \mathbf{A}_1\\ \mathbf{0} & \mathbf{H}_2 & \mathbf{A}_2\end{bmatrix} \tag{3}$$

where $r_1\ge 1$ and $r_2\ge 1$; the column number of the matrix $\boldsymbol{A}=(\boldsymbol{A_1}, \boldsymbol{A_2})^T$ (denoted by k) is an arbitrary value between 0 and $n_1\times n_2$, i.e. $k \in [0, n_1\times n_2]$, and each column a in $\boldsymbol{A}$ is the result of executing the bit-wise XOR operation between one column l_1 in the matrix $\boldsymbol{L_1}=(\boldsymbol{H_1}, \boldsymbol{0})^T$ and another column l_2 in the matrix $\boldsymbol{L_2}=(\boldsymbol{0}, \boldsymbol{H_2})^T$.

Proof. Assume that (r_1+r_2) bits of secret messages $\mathbf{m}=(\mathbf{m_1}, \mathbf{m_2})^T$ need to be embedded into (n_1+n_2+k) bits of cover messages $\mathbf{c}=(\mathbf{c_1}, \mathbf{c_2}, \mathbf{c_3})^T$, then

$$\mathbf{Hc}=\begin{bmatrix}\mathbf{H}_1 & \mathbf{0} & \mathbf{A}_1\\ \mathbf{0} & \mathbf{H}_2 & \mathbf{A}_2\end{bmatrix}\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3\end{bmatrix}^T=\begin{bmatrix}\mathbf{H}_1\mathbf{c}_1{}^T\oplus\mathbf{A}_1\mathbf{c}_3{}^T\\ \mathbf{H}_2\mathbf{c}_2{}^T\oplus\mathbf{A}_2\mathbf{c}_3{}^T\end{bmatrix} \tag{4}$$

where "$\oplus$" is bit-wise XOR operation.

1) If $\mathbf{m}=\mathbf{Hc}$, namely, $\mathbf{m_1}^T=\mathbf{H_1}\mathbf{c_1}^T\oplus\mathbf{A_1}\mathbf{c_3}^T$ and $\mathbf{m_2}^T=\mathbf{H_2}\mathbf{c_2}^T\oplus\mathbf{A_2}\mathbf{c_3}^T$, no bit in $\mathbf{c}$ needs to be changed.
2) If $\mathbf{m}\ne\mathbf{Hc}$, there are four cases as follows:
   2.1) If $\mathbf{m_1}^T\ne\mathbf{H_1}\mathbf{c_1}^T\oplus\mathbf{A_1}\mathbf{c_3}^T$ and $\mathbf{m_2}^T=\mathbf{H_2}\mathbf{c_2}^T\oplus\mathbf{A_2}\mathbf{c_3}^T$, one bit in $\mathbf{c_1}$ should be changed.
   2.2) If $\mathbf{m_1}^T=\mathbf{H_1}\mathbf{c_1}^T\oplus\mathbf{A_1}\mathbf{c_3}^T$ and $\mathbf{m_2}^T\ne\mathbf{H_2}\mathbf{c_2}^T\oplus\mathbf{A_2}\mathbf{c_3}^T$, one bit in $\mathbf{c_2}$ should be changed.
   2.3) If $\mathbf{m_1}^T\ne\mathbf{H_1}\mathbf{c_1}^T\oplus\mathbf{A_1}\mathbf{c_3}^T$, $\mathbf{m_2}^T\ne\mathbf{H_2}\mathbf{c_2}^T\oplus\mathbf{A_2}\mathbf{c_3}^T$, $\mathbf{m_1}^T=\mathbf{H_1}\mathbf{c_1}^T$ and $\mathbf{m_1}^T=\mathbf{H_2}\mathbf{c_2}^T$, only one bit in $\mathbf{c_3}$ needs to be changed.
   2.4) If $\mathbf{m_1}^T\ne\mathbf{H_1}\mathbf{c_1}^T\oplus\mathbf{A_1}\mathbf{c_3}^T$, $\mathbf{m_2}^T\ne\mathbf{H_2}\mathbf{c_2}^T\oplus\mathbf{A_2}\mathbf{c_3}^T$, $\mathbf{m_1}^T\ne\mathbf{H_1}\mathbf{c_1}^T$ and $\mathbf{m_1}^T\ne\mathbf{H_2}\mathbf{c_2}^T$, two bits (one in $\mathbf{c_1}$ and another in $\mathbf{c_2}$) need to be changed.

To sum up, at most two bits in $\mathbf{c}$ needs to be changed in the embedding process. In other words, the distortion bound of the guide matrix is 2. Therefore, we can conclude that the guide matrix $\mathbf{H}$ can be obtained using formula (3). □□

Note that if r_1 (or r_2)$=1$, there are no such Hamming codes in reality. For the sake of completeness, we define [1] as the parity check matrix in this case.

Let us assume $n=n_1+n_2$ and $r=r_1+r_2$. From Theorem 1, we can learn that r bits of secret messages can be embedded into n bits of cover messages using the following guide matrix $\mathbf{H_s}$, namely,

$$\mathbf{H_s}=\begin{bmatrix}\mathbf{H}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{H}_2\end{bmatrix} \tag{5}$$

Apparently, $\mathbf{H_s}$ is the smallest one among all the guide matrices generated according to Theorem 1, but it is not necessarily the minimum one for embedding r bits of secret messages. We can make an analysis as follows. The column number of $\mathbf{H_s}$ can be considered as a function of r_1 and r_2, namely,

$$\begin{aligned} f(r_1, r_2) &= 2^{r_1}-1+2^{r_2}-1\\ &= 2^{r_1}+2^{r-r_1}-2\\ &\ge 2\sqrt{2^{r_1}\times\frac{2^r}{2^{r_1}}}-2=2^{r/2+1}-2 \end{aligned} \tag{6}$$

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

If and only if $r_1=r_2=r/2$, the equation holds. That is to say, when $r_1=r_2$, the column number of $\mathbf{H_s}$ achieves the minimum value $f_{min}(r_1, r_2)=2^{r/2+1}-2$. Considering that r_1 and r_2 are integers and r may be an odd number, we can derive a variant of the above inequality, namely,

$$\begin{aligned} f(r_1,r_2) &= 2^{r_1}-1+2^{r_2}-1 \\ &= 2^{r_1}+2^{r-r_1}-2 \\ &\geq 2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2 \end{aligned} \tag{7}$$

and if $r_1=\lfloor r/2\rfloor$ and $r_2=\lceil r/2\rceil$ (or if $r_1=\lceil r/2\rceil$ and $r_2=\lfloor r/2\rfloor$), the equation holds. Accordingly, we can get Theorem 2 below.

Theorem 2. *The minimum guide matrix with the distortion bound not more than two. Given the $r\times n$ guide matrix $\boldsymbol{H}$ constructed by the parity check matrices of an $[n_1, n_1-r_1]$ Hamming code and another $[n_2, n_2-r_2]$ Hamming code, where $r_1\geq 1$, $r_2\geq 1$, $n=n_1+n_2$ and $r=r_1+r_2$. If $r_1=\lfloor r/2\rfloor$ and $r_2=\lceil r/2\rceil$ (if $r_1=\lceil r/2\rceil$ and $r_2=\lfloor r/2\rfloor$), the column number of $\boldsymbol{H}$ achieves the minimum value $2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2$, and $\boldsymbol{H}$ is accordingly the minimum guide matrix with the distortion bound not more than two.*

According to Theorem 1, on the other hand, the maximum column number of the guide matrix is

$$\begin{aligned} n_{max} &= n_1+n_2+n_1\times n_2 \\ &= 2^{r_1}-1+2^{r_2}-1+(2^{r_1}-1)\times(2^{r_2}-1) \\ &= 2^{r_1+r_2}-1 \\ &= 2^r-1 \end{aligned} \tag{8}$$

Thus, we can get Theorem 3 below.

Theorem 3. *wThe maximum guide matrix with the distortion bound not more than two. Given the guide matrix $\boldsymbol{H}$ constructed by the parity check matrices of an $[n_1, n_1-r_1]$ Hamming code and another $[n_2, n_2-r_2]$ Hamming code, where $r_1\geq 1$, $r_2\geq 1$ and $r=r_1+r_2$. The maximum column number of $\boldsymbol{H}$ is 2^r-1. Accordingly, the $r\times(2^r-1)$ guide matrix is the maximum one with the distortion bound not more than two.*

We can deduce from Theorems 2 and 3 that, for embedding r bits of secret messages, we can construct an adjustable guide matrix, of which the column number is between $2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2$ and 2^r-1. That is to say, we can adaptively generate a guide matrix to embed r bits of secret messages into n ($n\in[2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2, 2^r-1]$) bits of cover messages, which is the very core idea of our AME for steganography.

Moreover, it can be observed that the maximum guide matrix is actually the parity check matrix of a $[2^r-1, 2^r-r-1]$ Hamming code, which suggests that the standard ME can be regarded as a special case of the AME.

3. Steganography scheme using AME

3.1. Embedding and extracting algorithms

For the convenience of implementation, this section will present the steganography scheme using AME in detail. Let us assume that r bits of secret messages $M=(m_1, m_2, \cdots, m_r)$ will be embedded into n bits of cover messages $C=(c_1, c_2, \cdots, c_n)$, where $n\in[2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2, 2^r-1]$. Note that the secret messages are often encrypted in advance. For ease of presentation, we assume that M is the cipher text of secret messages here. The embedding process using AME can be described as follows:

STEP 1: Construct an $r\times n$ guide matrix $\mathbf{H}$ using the parity check matrix $\mathbf{H_1}$ of the $[n_1, n_1-r_1]$ Hamming code and the parity check matrix $\mathbf{H_2}$ of the $[n_2, n_2-r_2]$ Hamming code according to Theorem 1, namely

$$\begin{aligned} \mathbf{H} &= \begin{bmatrix} h_{1,1} & h_{2,1} & \cdots & h_{n,1} \\ h_{1,2} & h_{2,2} & \cdots & h_{n,2} \\ \vdots & \vdots & \vdots & \vdots \\ h_{1,r} & h_{2,r} & \cdots & h_{n,r} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} & \mathbf{A}_1 \\ \mathbf{0} & \mathbf{H}_2 & \mathbf{A}_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{U}_{r\times n_1} & \mathbf{V}_{r\times n_2} & \mathbf{W}_{r\times k} \end{bmatrix} \end{aligned} \tag{9}$$

where

$$r=r_1+r_2 \tag{10}$$

$$r_1=\lfloor r/2\rfloor \ (\text{or } \lceil r/2\rceil) \tag{11}$$

$$r_2=\lceil r/2\rceil (\text{or } \lfloor r/2\rfloor) \tag{12}$$

$$n_1=2^{r_1}-1=2^{\lfloor r/2\rfloor}-1 \ \left(\text{or } 2^{\lceil r/2\rceil}-1\right) \tag{13}$$

$$n_2=2^{r_2}-1=2^{\lceil r/2\rceil}-1 \ \left(\text{or } 2^{\lfloor r/2\rfloor}-1\right) \tag{14}$$

$$k=n-n_1-n_2=n-2^{\lfloor r/2\rfloor}-2^{\lceil r/2\rceil}+2 \tag{15}$$

$$\mathbf{U}_{r\times n_1}=\begin{bmatrix} u_{1,1} & u_{2,1} & \cdots & u_{n_1,1} \\ u_{1,2} & u_{2,2} & \cdots & u_{n_1,2} \\ \vdots & \vdots & \vdots & \vdots \\ u_{1,r_1} & u_{2,r_1} & \cdots & u_{n_1,r_1} \\ \multicolumn{4}{c}{\mathbf{0}_{r_2\times n_1}} \end{bmatrix} \tag{16}$$

$$\mathbf{V}_{r\times n_2}=\begin{bmatrix} \multicolumn{4}{c}{\mathbf{0}_{r_1\times n_2}} \\ v_{1,1} & v_{2,1} & \cdots & v_{n_2,1} \\ v_{1,2} & v_{2,2} & \cdots & v_{n_2,2} \\ \vdots & \vdots & \vdots & \vdots \\ v_{1,r_2} & v_{2,r_2} & \cdots & v_{n_2,r_2} \end{bmatrix} \tag{17}$$

$$\mathbf{W}_{r\times k}=\begin{bmatrix} w_{1,1} & w_{2,1} & \cdots & w_{k,1} \\ w_{1,2} & w_{2,2} & \cdots & w_{k,2} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1,r} & w_{2,r} & \cdots & w_{k,r} \end{bmatrix} \tag{18}$$

According to the characteristic of parity check matrices for Hamming codes, we can learn that

$$i=\sum_{j=1}^{r_1} u_{i,j}\cdot 2^{j-1}, \quad u_{i,j}=0 \text{ or } 1, \quad 1\leq i\leq n_1 \tag{19}$$

$$i=\sum_{j=1}^{r_2} v_{i,j}\cdot 2^{j-1}, \quad v_{i,j}=0 \text{ or } 1, \quad 1\leq i\leq n_2 \tag{20}$$

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

Moreover, in the matrix $\mathbf{W}_{r\times k}$, we have

$$\begin{pmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,r} \end{pmatrix} = \begin{pmatrix} u_{x,1} \\ u_{x,2} \\ \vdots \\ u_{x,r_1} \\ v_{y,1} \\ v_{y,2} \\ \vdots \\ v_{y,r_2} \end{pmatrix}, \quad 1 \le i \le k, \quad 1 \le x \le n_1, \quad 1 \le y \le n_2 \tag{21}$$

which means that each column in $\mathbf{W}_{r\times k}$ is the result of executing the bit-wise XOR operation between one column in $\mathbf{U}_{r\times n1}$ and another column in $\mathbf{V}_{r\times n2}$.

It is not hard to see that there are more than one satisfactory guide matrix, if $n \in (2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2, 2^r-1)$. Generally, the number of candidate guide matrix (denoted by Q) can be determined as follows:

$$\begin{aligned} Q &= 2^{\sigma} \times \mathbf{C}^{k}_{n_1\times n_2} \\ &= 2^{\sigma} \times \mathbf{C}^{(n-2^{\lfloor r/2\rfloor}-2^{\lceil r/2\rceil}+2)}_{n_1\times n_2} \end{aligned} \tag{22}$$

where $\sigma = r \bmod 2$. Therefore, the adopted guide matrix can be also used as a key of the embedding process to further improve security.

STEP 2: Determine the weight vector $\xi=(\xi_1, \xi_2, \cdots, \xi_n)$ of $\mathbf{H}$, in which the i-th element ξ_i is the weight value of the i-th column in $\mathbf{H}$, namely

$$\xi_i = \sum_{j=1}^{r} h_{i,j} \cdot 2^{j-1}, \quad 1 \le i \le n \tag{23}$$

STEP 3: For each row in $\mathbf{H}$, calculate

$$x_j = \begin{cases} 0, & m_j = \oplus_{i=1}^{n} (c_i \cdot h_{i,j}) \\ 1, & m_j \ne \oplus_{i=1}^{n} (c_i \cdot h_{i,j}) \end{cases}, \quad 1 \le j \le r \tag{24}$$

where $\oplus_{i=1}^{n}$ represents continuous XOR operations.

STEP 4: Calculate the following expression:

$$X = \sum_{j=1}^{r} x_j \cdot 2^{j-1} \tag{25}$$

If $X=0$, then there are no bits needed to be modified in C, i.e. the steganographic form of the cover $C^*=C$; otherwise, there are two possible situations: (1) If $\exists\ X=\xi_i \in \boldsymbol{\xi}$, then the i-th bit c_i need to be flipped; (2) If $\nexists\ X=\xi_i \in \boldsymbol{\xi}$, $1 \le i \le n$, then we should choose a $\xi_\alpha \in (\xi_1, \xi_2, \cdots, \xi_{n1})$ and another $\xi_\beta \in (\xi_{n1+1}, \xi_{n1+2}, \cdots, \xi_{n1+n2})$, such that $\xi_\alpha+\xi_\beta=X$. Accordingly, we can transform C to C^* by modifying the α -th bit c_α and the β-th bit c_β in C, namely, $C^*=(c_1, c_2, \cdots, 1-c_\alpha, \cdots, 1-c_\beta, \cdots, c_n)$.

To show the above embedding process, we give two concrete examples below.

Example 1. Assume that $r=3$, $n=5$, $M=(0, 1, 0)$, $C=(0, 1, 0, 1, 0)$, then $r_1=1$ and $r_2=2$ (or $r_1=2$ and $r_2=1$). From Eq. (22), we can learn that there are six satisfactory guide matrices, of which here we take one as follows:

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

Further, we can get the weight vector $\boldsymbol{\xi}=(1, 2, 4, 6, 5)$. By Eq. (24), we can obtain $x_1=0$, $x_2=1$ and $x_3=1$, then $X=0\times 1+1\times 2+1\times 4=6=\xi_4 \in \boldsymbol{\xi}$. Therefore, c_4 needs to be flipped, i.e. $C^*=\{0, 1, 0, 0, 0\}$.

Example 2. In the above example, if $M=(1, 1, 0)$, then $x_1=1$, $x_2=1$ and $x_3=1$. Accordingly, $X=1\times 1+1\times 2+1\times 4=7 \notin \boldsymbol{\xi}$, but then we can find that, $\xi_1+\xi_4=7$, $\xi_1 \in (\xi_1)$ and $\xi_4 \in (\xi_2, \xi_3, \xi_4)$. Therefore, c_1 and c_4 need to be flipped, i.e. $C^*=\{1, 1, 0, 0, 0\}$.

The extracting process is relatively simple. The receiver first obtains C^* from the VoIP stream, and extracts each bit of embedded messages by calculating the following expression:

$$m_j = \oplus_{i=1}^{n} (c_i^* \cdot h_{i,j}), \quad 1 \le j \le r \tag{26}$$

For the given Example 1, we can easily restitute $m_1=0$, $m_2=1$ and $m_3=0$, namely, $M=(0, 1, 0)$; and for the Example 2, we can extract $m_1=1$, $m_2=1$ and $m_3=0$, namely, $M=(1, 1, 0)$.

3.2. Embedding performance analysis

Embedding rate (ER, denoted by $\mu(r, n)$) of the above steganography scheme, which indicates the embedding capability, can be calculated as follows:

$$\mu(r, n) = \frac{r}{n} \tag{27}$$

Moreover, the average number of cover bits changed (denoted by $g(r, n)$) is

$$\begin{aligned} g(r, n) &= \tfrac{1}{2^r} \times 0 + \tfrac{n}{2^r} \times 1 + \tfrac{2^r-n-1}{2^r} \times 2 \\ &= \frac{1}{2^r} \times 0 + \frac{n_1+n_2+k}{2^r} \times 1 + \frac{n_1 \times n_2 - k}{2^r} \times 2 \\ &= \frac{2\times(n_1+n_2)+2\times n_1 \times n_2 - n}{2^r} \\ &= \frac{2\times(2^{\lfloor r/2\rfloor}-1+2^{\lceil r/2\rceil}-1+(2^{\lfloor r/2\rfloor}-1)\times(2^{\lceil r/2\rceil}-1))-n}{2^r} \\ &= \frac{2\times(2^r-1)-n}{2^r} \\ &= 2-\frac{n+2}{2^r} \end{aligned} \tag{28}$$

where $(n+2)/2^r$ is invariably greater than zero. Thus, $g(r, n)$ is always less than 2, which proves once again that not more than 2 bits of cover messages need to be changed in the embedding process.

We define bit-change rate (BCR, denote by $\lambda(r, n)$) as the ratio between the number of cover bits changed and the total number of cover bits, so

$$\begin{aligned} \lambda(r, n) &= \tfrac{g(r,n)}{n} \\ &= \frac{2-\frac{n+2}{2^r}}{n} \end{aligned}$$

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

Table 2
Relation of cover length to embedding performance.

EP	$n\uparrow$	The max. value	The min. value
ER	$\mu(r, n)\downarrow$	$\frac{r}{2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2}$	$\frac{r}{2^r-1}$
BCR	$\lambda(r, n)\downarrow$	$\frac{2^{r+1}-2^{\lceil r/2\rceil}-2^{\lfloor r/2\rfloor}}{2^r\cdot(2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2)}$	$\frac{1}{2^r}$
EE	$\varpi(r, n)\uparrow$	$\frac{r\cdot 2^r}{2^r-1}$	$\frac{r\cdot 2^r}{2^{r+1}-2^{\lceil r/2\rceil}-2^{\lfloor r/2\rfloor}}$

Note: "↑" indicates "increase", and "↓" indicates "decrease".

$$=\frac{2}{n}\times(1-\frac{1}{2^r})-\frac{1}{2^r} \tag{29}$$

Further, embedding efficiency (EE, denoted by $\varpi(r, n)$), which indicates the average number of bits embedded per one bit cover change, can be determined as follows:

$$\begin{aligned}\varpi(r,n)&=\frac{\mu(r,n)}{\lambda(r,n)}\\&=\frac{\frac{r}{n}}{\frac{2}{n}\times(1-\frac{1}{2^r})-\frac{1}{2^r}}\\&=\frac{r\cdot 2^r}{2^{r+1}-n-2}\end{aligned} \tag{30}$$

Table 2 illustrates the relation of the cover length n to embedding performance (EP). Overall, both ER $\mu(r, n)$ and BCR $\lambda(r, n)$ decrease with increasing n, but EE $\varpi(r, n)$ increases with n. As mentioned above, the maximal value and the minimum of n are respectively 2^r-1 and $2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2$, we can accordingly get the maximum and the minimum of ER, BCR and EE, as shown in Table 2.

We can now make use of the above scheme to suit the cover length and achieve an optimal embedding. However, it is worthwhile to note that the cover length n should be between $2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2$ and 2^r-1 for embedding r bits of secret messages. In other words, if $n\notin[2^{\lceil r/2\rceil}+2^{\lfloor r/2\rfloor}-2,\ 2^r-1]$, it is impossible to construct a suitable guide matrix to realize the desired embedding. For example, if $r=5$ and $n=9$, no suitable guide matrix can be gotten, because the satisfactory n should be between 10 and 31. However, we can divide the messages into two or more parts, and embed them in disjoint parts of the cover using different guide matrices. In this example, we may embed the first 2 bits of secret messages into 3 bits of cover messages using a 2×3 guide matrix, and the other 3 bits of secret messages into the rest 6 bits of cover messages using a 3×6 guide matrix. Apparently, the division is often not unique. So, how many divisions are there, and which is the best one for the optimal embedding? It is another concern in this paper, which will be addressed in the following section:

4. The optimal embedding using AMEs

As mentioned above, for the given r bits of secret messages and n bits of cover messages ($n\geq r$), we can use one or more AMEs with different parameters to achieve the optimal embedding. Assume that the secret message and the cover message are respectively divided into S groups, the group length set of the secret message is $R=\{r_1, r_2, \cdots, r_S\}$, and the group length set of the cover message is $N=\{n_1, n_2, \cdots, n_S\}$, the optimal embedding model can be expressed as follows:

$$\begin{cases}\text{maximize } \varpi(r,n)=\frac{\mu(r,n)}{\lambda(r,n)}=\frac{r}{\sum_{i=1}^{S}g(r_i,n_i)}\\ s.t.\ \sum_{i=1}^{S}r_i=r\\ \sum_{i=1}^{S}n_i\leq n\\ 2^{\lceil r_i/2\rceil}+2^{\lfloor r_i/2\rfloor}-2\leq n_i\leq 2^{r_i}-1\\ r_i,\ n_i\in\mathbf{Z},\quad 1\leq i\leq S\end{cases} \tag{31}$$

The above formal description indicates that the optimal embedding problem is essentially a typical combinatorial optimization problem, which aims at optimizing all $r_i\in R$ and corresponding $n_i\in N$. Clearly, the complexity of this problem depends on the range of message group lengths. Fortunately, we just need to consider the range of r_i from 1 to $\lceil\log_2^{n+1}\rceil$ according to the given cover length n. To achieve the best embedding performance, Algorithm 1 is employed to obtain the optimal combination of R and N.

Algorithm 1. Searching algorithm of the optimal combination of R and N

Require: r and n are the lengths of secret messages and cover messages respectively.
Output: $\mathbb{R}[]$ and $\mathbb{N}[]$, which are respectively the arrays of the optimal R and N; the optimal (maximum) embedding efficiency ϖ_{opt}.

```
Determine set 𝕌 that consists of all valid R.
For each R ∈ 𝕌 Do
    Determine set 𝕋 that includes all valid N for R.
    ϖopt=0.
    For each N ∈ 𝕋 Do
        Calculate embedding efficiency ϖ(R, N).
        If (ϖ(R, N) > ϖopt) Then
            Empty ℝ[ ] and ℕ[ ].
            idx =0.
            ℝ[idx]=R, ℕ[idx]=N.
            ϖopt=ϖ(R, N).
        Else If (ϖ(R, N)=ϖopt) Then
            idx=idx+1.
            ℝ[idx]=R, ℕ[idx]=N.
        End If
    End For
End For
Return ℝ[], ℕ[] and ϖopt
```

By means of this algorithm, we can now determine an optimal embedding scheme for the given $r\geq 1$ and $n\geq r$. It is worthwhile to note that the above search process is often carried out offline, and a datasheet like Table 3 is made accordingly, in which we can conveniently look up the optimal scheme when required. Using Algorithm 1, we have collected the optimal embedding schemes with different values of r for the given $n=40$, as shown in Table 3. From these data, we can learn the following rules:

- For $r=1$ or $r=40$, the optimal embedding scheme is the direct bit substitution, and the EE is accordingly 2.0.
- For $r\in[2,5]$, we can use the matrix encoding based on Hamming codes to achieve the optimal embedding with relatively high EE. Particularly, if $r=5$, EE reaches the maximum value 5.1613.

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

Table 3
Optimal embedding schemes with different r for $n=40$.

r	ϖ_{opt}	R_{opt}	N_{opt}	Embedding Scheme	r	ϖ_{opt}	R_{opt}	N_{opt}	Embedding scheme
1	2.0000	{1}	{1}	$1\times(1, 1)$	21	2.9735	{3,4,4,5,5}	{7,6,7,10,10}	$1\times(3, 7)+1\times(4, 6)+1\times(4, 7)+2\times(5, 10)$
2	2.6667	{2}	{3}	$1\times(2, 3)$	22	2.8852	{4,4,4,5,5}	{6,6,8,10,10}	$2\times(4, 6)+1\times(4,8)+2\times(5, 10)$
3	3.4286	{3}	{7}	$1\times(3, 7)$	23	2.8308	{3,3,4,4,4,5}	{5,7,6,6,6,10}	$1\times(3, 5)+1\times(3, 7)+3\times(4, 6)+1\times(5, 10)$
4	4.2667	{4}	{15}	$1\times(4, 15)$	24	2.7826	{3,4,4,4,4,5}	{6,6,6,6,6,10}	$1\times(3, 6)+4\times(4,6)+1\times(5, 10)$
5	5.1613	{5}	{31}	$1\times(5, 31)$	25	2.7397	{4,4,4,4,4,5}	{6,6,6,6,6,10}	$5\times(4, 6)+1\times(5, 10)$
6	4.4651	{6}	{40}	$1\times(6, 40)$	26	2.6839	{2,4,4,4,4,4,4}	{3,6,6,6,6,6,7}	$1\times(2, 3)+5\times(4,6)+1\times(4, 7)$
7	4.0727	{2,5}	{3,31}	$1\times(2, 3)+1\times(5, 31)$	27	2.6341	{3,4,4,4,4,4,4}	{4,6,6,6,6,6,6}	$1\times(3, 4)+6\times(4, 6)$
8	4.3389	{3,5}	{7,31}	$1\times(3, 7)+1\times(5, 31)$	28	2.5455	{3,3,3,3,4,4,4,4}	{4,4,4,4,6,6,6,6}	$4\times(3, 4)+4\times(4, 6)$
9	4.2985	{4,5}	{15,25}	$1\times(4, 15)+1\times(5, 25)$	29	2.4681	{3,3,3,3,3,3,3,4,4}	{4,4,4,4,4,4,4,6,6}	$7\times(3, 4)+2\times(4, 6)$
10	3.9751	{4,6}	{15,25}	$1\times(4, 15)+1\times(6, 25)$	30	2.4000	{3,3,3,3,3,3,3,3,3,3}	{4,4,4,4,4,4,4,4,4,4}	$10\times(3, 4)$
11	4.0000	{3, 4, 4}	{7, 15,15}	$1\times(3, 7)+2\times(4, 15)$	31	2.3396	{4,3,3,3,3,3,3,3,3,3}	{4,4,4,4,4,4,4,4,4,4}	$1\times(4, 4)+9\times(3, 4)$
12	3.8400	{4,4,4}	{15,15,10}	$2\times(4, 15)+1\times(4, 10)$	32	2.2857	{8,3,3,3,3,3,3,3,3}	{8,4,4,4,4,4,4,4,4}	$1\times(8, 8)+8\times(3, 4)$
13	3.7143	{4,4,5}	{15,15,10}	$2\times(4, 15)+1\times(5, 10)$	33	2.2373	{12,3,3,3,3,3,3,3}	{12,4,4,4,4,4,4,4}	$1\times(12, 12)+7\times(3, 4)$
14	3.6129	{4,4,6}	{11,15,14}	$1\times(4, 11)+1\times(4, 15)+1\times(6, 14)$	34	2.1935	{16,3,3,3,3,3,3}	{16,4,4,4,4,4,4}	$1\times(16, 16)+6\times(3, 4)$
15	3.5036	{4,5,6}	{15,11,14}	$1\times(4, 15)+1\times(5, 11)+1\times(6, 14)$	35	2.1538	{20,3,3,3,3,3}	{20,4,4,4,4,4}	$1\times(20, 20)+5\times(3, 4)$
16	3.4595	{4,6,6}	{12,14,14}	$1\times(4, 12)+2\times(6, 14)$	36	2.1176	{24,3,3,3,3}	{24,4,4,4,4}	$1\times(24, 24)+4\times(3, 4)$
17	3.3580	{5,6,6}	{12,14,14}	$1\times(5, 12)+2\times(6, 14)$	37	2.0845	{28,3,3,3}	{28,4,4,4}	$1\times(28, 28)+3\times(3, 4)$
18	3.2727	{3,3,6,6}	{5,7,14,14}	$1\times(3, 5)+1\times(3, 7)+2\times(6, 14)$	38	2.0541	{32,3,3}	{32,4,4}	$1\times(32, 32)+2\times(3, 4)$
19	3.1667	{3,5,5,6}	{6,10,10,14}	$1\times(6, 5)+2\times(5, 10)+1\times(6, 14)$	39	2.0260	{36,3,}	{36,4 }	$1\times(36, 36)+1\times(3, 4)$
20	3.0769	{5,5,5,5}	{10,10,10,10}	$4\times(5, 10)$	40	2.0000	{40}	{40}	$1\times(40, 40)$

Note: (1) ϖ_{opt} is the optimal (maximum) embedding efficiency; the optimal group length set of the secret message R_{opt} is denoted as $R_{opt}=\{r_1, r_2, \cdots, r_S\}$, and the optimal group length set of the cover message N_{opt} is denoted as $N_{opt}=\{n_1, n_2, \cdots, n_S\}$, where S is the number of message groups; (2) in the embedding schemes, $k\times(x, y)$ $(x\neq y)$ means using AME with an $x\times y$ guide matrix k times, and (x, x) means using direct bit substitution; and (3) in each case, we just give one optimal scheme, although the optimal scheme is not unique in some cases.

- For $r=6$, we can employ the AME with a 6×40 guide matrix to achieve the optimal embedding with the second largest EE, i.e. 4.4651.
- In the other cases, two or more AMEs with different guide matrices should be adopted.

All in all, an optimal embedding can be realized using one or multiple AMEs in any case, which is verified once again by the above results.

5. Performance evaluation

We evaluate the proposed scheme in StegVoIP [10,21], a prototypical VoIP-based covert communication system, which supports typical speech codecs, such as ITU-T G.711, G.723.1, G.729a. It is worthwhile to emphasize that our scheme is essentially an advanced matrix encoding technique for steganography, which does not depend on specific codecs and covers. That is to say, it can be applied in the presence of all speech codecs used in VoIP, in that steganography in speech streams encoded by various codecs is feasible [9]. For realizing the scheme with diverse speech codecs, the only difference is the available cover bits in each speech frame. Without loss of generality, we typically choose widely-used ITU-T G. 711 (A-law) [41,42] as the codec of the cover speech in this study.

We compare the proposed scheme with the ME and the LSB method with the cover length per part equal to 40. Figs. 1 and 2 show the results of BCR and EE while embedding secret messages by different amounts. From them, we can learn that: (1) the proposed scheme can support different embedding capacities from 1 bit to 40 bits, while the ME can only embed at most 5 bits of secret messages, which means our scheme outperforms the ME on adaptability; and (2) although the LSB method can also adjust embedding capacities from 1 bit to 40 bits, but its BCRs increase with embedding rates, and its EEs are invariably equal with 2.0. The proposed scheme, by contrast, can achieve smaller BCRs and higher EEs. That is to say, our scheme provides better embedding performance than the LSB method.

In order to comprehensively evaluate the embedding performance of the proposed scheme, we define 8 groups of experimental modes, as shown in Table 4. Further, for the experiments, we collect 1000 ten-s speech samples, which consist of four categories: Chinese female speech, Chinese male speech, English female speech and English male speech. All samples are PCM coded files with 8 kHz sampling rate, 16 bits quantization and mono. For each sample, we perform the corresponding steganography operations on its G. 711 coded file in the 8 groups of experiments respectively. Without loss of generality, the secret messages for all the experiments are produced at random, which can be successfully embedded and extracted in any case.

For all these experimental modes, the statistical analyses are made on BCR and EE. Moreover, for assessing the embedding transparency, we employ the Perceptual Evaluation of Speech Quality (PESQ) method presented in the ITU-T P. 862 Recommendation [43,44], which compares an original

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

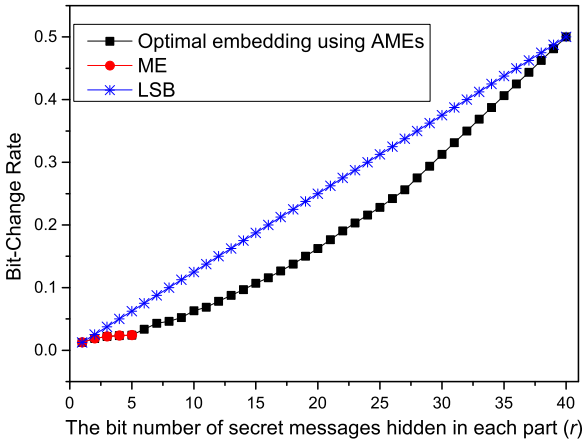

Fig. 1. BCR comparison among the proposed scheme, ME and LSB method.

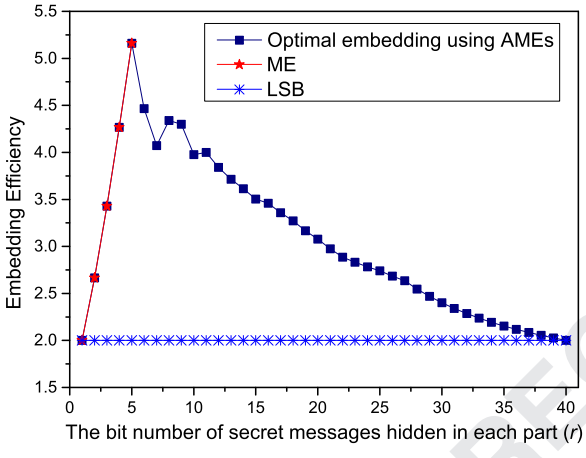

Fig. 2. EE comparison among the proposed scheme, ME and LSB method.

Table 4
Experimental modes.

Group	The cover length per part (n)	The length of secret messages per part (r)
1	5	1–5
2	10	1–10
3	15	1–15
4	20	1–20
5	25	1–25
6	30	1–30
7	35	1–35
8	40	1–40

signal with a degraded signal and outputs a PESQ score as a prediction of the perceived quality. However, Mean Opinion Score – Listening Quality Objective (MOS-LQO) converted from the PESQ score is usually used in practice, because its range is from 1.017 (worst) to 4.549 (best), which more closely matches the range of the subjective Mean Opinion Score (MOS) [45]. In the experiments, we convert all the steganographic G. 711 coded files into PCM encoded files as the degraded signals and perform the PESQ test with the original samples as the reference signals.

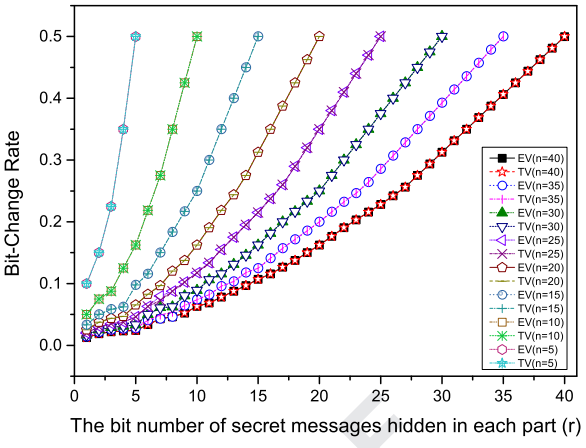

Fig. 3. Statistical results of bits-change rate (BCR), where EV($n=x$) and TV($n=x$) denote respectively the experimental value and the theoretical value of BCR when $n=x$.

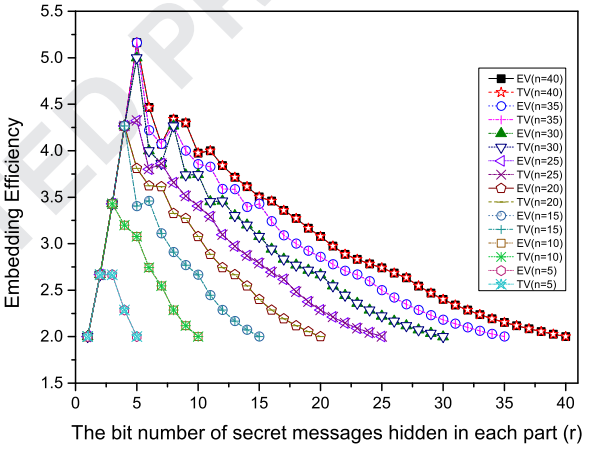

Fig. 4. Statistical results of embedding efficiency (EE), where EV($n=x$) and TV($n=x$) denote respectively the experimental value and the theoretical value of EE when $n=x$.

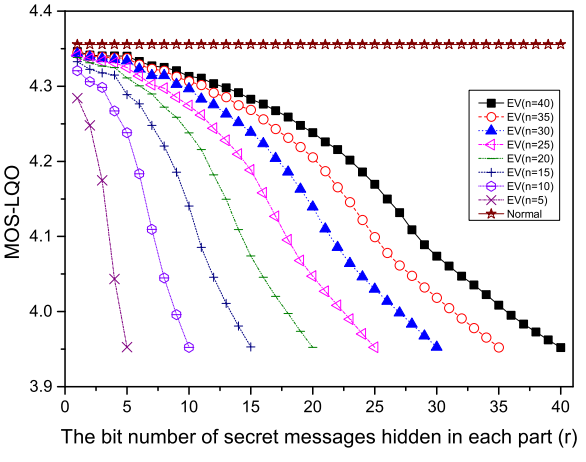

Fig. 5. Statistical results of MOS-LQO, where "Normal" denotes the MOS-LQO of speech decoded from its untouched G. 711 coded file.

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

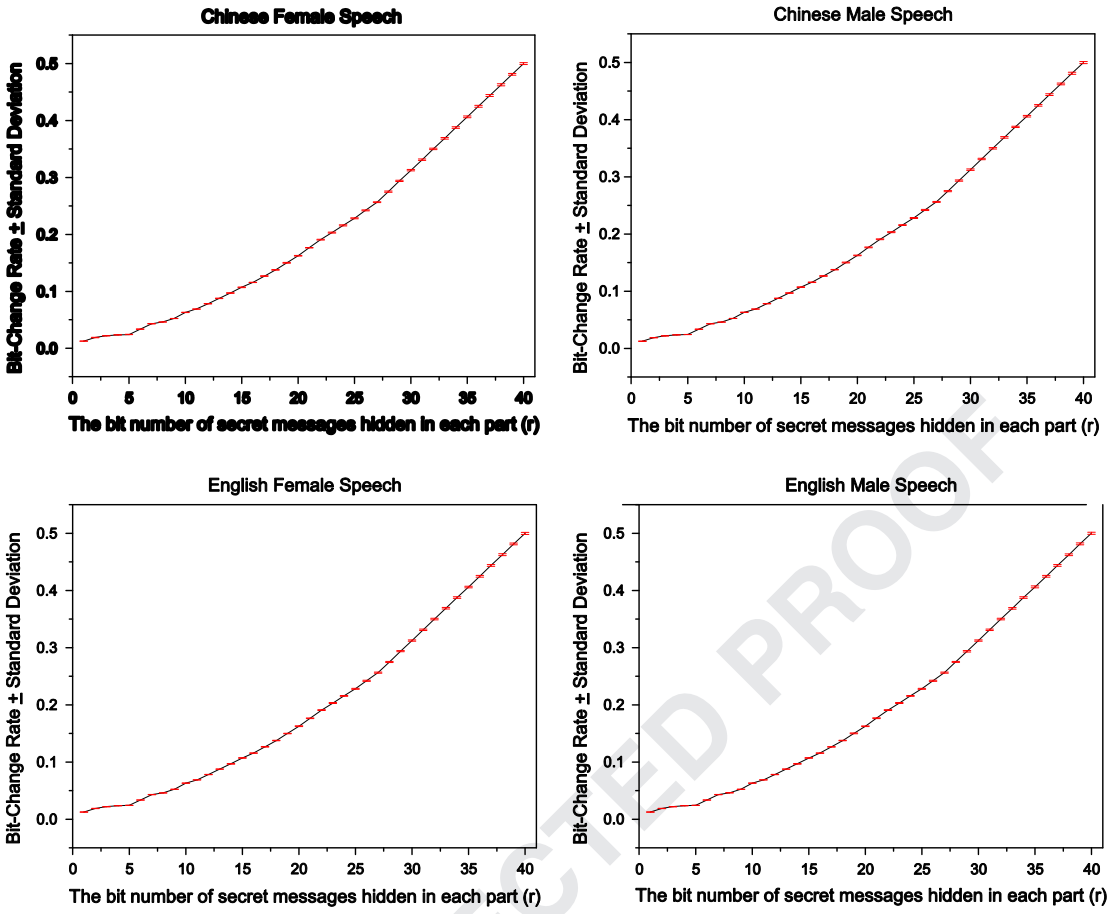

Fig. 6. Statistical results of bits-change rates with standard deviations for four speech categories when the cover length is 40.

Figs. 3–5 show the statistical results of BCR, EE and MOS-LQO respectively for all experimental modes. From these charts, we can learn that:

- The experimental values of both BCR and EE agree well with their theoretical values, which indicates the proposed scheme is really feasible in both theory and practice.
- For the given cover length n, the BCR value increases with bit number of embedded secret messages, i.e. r, and accordingly the MOS-LQO value decreases with the increase of embedded secret messages, indicating that the gain of embedding rate is at the expense of embedding transparency. Moreover, for embedding the same secret messages, the larger the cover length, the smaller the BCR, and the higher the MOS-LQO, indicating that it is advisable to optimize the use of the cover to achieve the best embedding transparency.
- If $r \le \lceil \log_2^{n+1} \rceil$, we can use a standard ME to achieve the optimal embedding with relatively high EE, which means the standard ME can be regarded as a special case of the proposed scheme. Particularly, when $r = \lceil \log_2^{n+1} \rceil$, EE reaches the maximum value. If $r > \lceil \log_2^{n+1} \rceil$, EE decreases with the increase of embedded secret messages on the whole. However, there are still a few exceptional cases (e.g. $r=9$ and $n=40$), in which EE can reach one of local maxima. The reason for this phenomenon may be that the change degrees of BCRs are unequal in different cases. This also gives us an inspiration that we prefer to choose these exceptional cases, if their embedding capacities are acceptable in practical applications.

To further verify the embedding performance of the proposed scheme, we make a detailed analysis on BCR, EE and the dispersions of their distributions when the given cover length is equal to 40. Figs. 6 and 7 show the statistical results of BCR, EE and their standard deviations on different categories of samples. From the charts, we can learn that: (1) The relation curves of BCR (EE) and the bit number of embedded secret messages on different categories of samples are exactly the same, which indicates the categories have little impact on BCR and EE. (2) The maximum standard deviations of BCR and EE are 0.0018 and 0.045 respectively and thereby negligible. Therefore, we can safely conclude that the proposed scheme can provide consistently optimal embedding performance in any case.

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

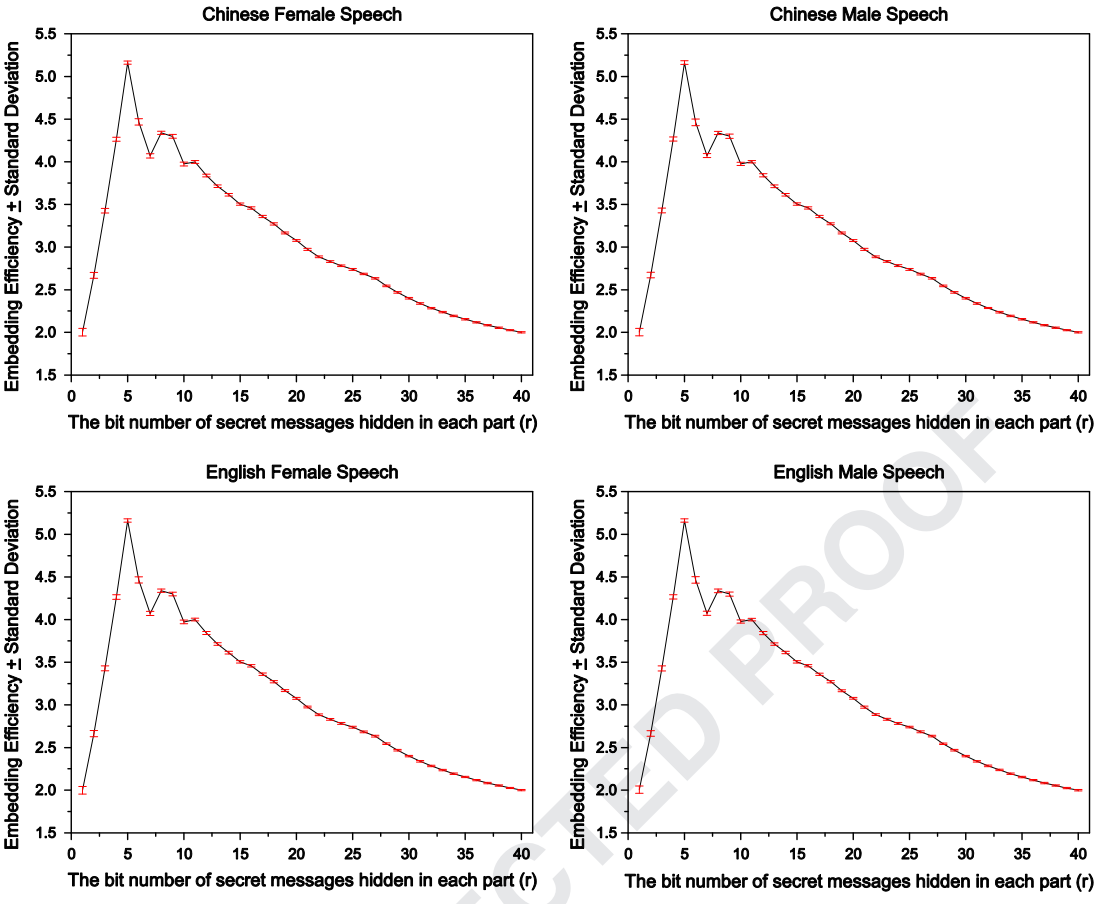

Fig. 7. Statistical results of embedding efficiencies with standard deviations for four speech categories when the cover length is 40.

6. Conclusions

In this paper, we concentrate on achieving optimal embedding for the given limited length of each cover part. For this purpose, we first present an adjustable matrix encoding, which can adaptively generate a guide matrix to make full use of each cover part compared to the previous encoding approaches. Further, we express the optimal embedding in the limited cover as a typical combinatorial optimization problem, and provide an algorithm to search the optimal solution, which can be described as an AME or a combination of multiple AMEs. The proposed scheme is evaluated with ITU-T G. 711 (A-law) as the codec of the cover speech and compared with some existing methods. The experimental results demonstrate that the proposed scheme is really feasible in both theory and practice, and can provide consistently optimal embedding performance in any case.

Besides, from the experimental results, we can learn that for the given cover length, the BCR value increases with the bit number of embedded secret messages, but the change degrees of the BCR values differ in different cases, which is the cause of existing some local maxima of EE. That also provides an inspiration for us that these exceptional cases may be exploited to achieve better embedding performance in terms of EE, if their embedding capacities are acceptable in practical applications.

At last, we would like to point out that the proposed scheme in this paper is codec-independent and cover-independent, meaning that it can be applied in the presence of any other coders used in VoIP and deployed with any other media in limited cover scenarios.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China under Grant nos. 61302094, U1405254, 61370007 and 61202468, Natural Science Foundation of Fujian Province of China under Grant no. 2014J01238, Education and Science Research Program for Young and Middle-aged Teachers of Fujian Province of China under Grant no. JA13012, Promotion Program for Young and Middle-aged Teacher in Science & Technology Research of Huaqiao University under Grant no. ZQN-PY115, and Program for Science & Technology Innovation Teams and Leading Talents of Huaqiao University under Grant no. 2014 KJT D13.

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001

References

[1] N. Provos, P. Honeyman, Hide and seek: an introduction to steganography, IEEE Secur. Priv. Mag. 1 (no. 3) (2003) 32–44.
[2] A. Cheddad, J. Condell, K. Curran, P.M. Kevitt, Digital image steganography: survey and analysis of current methods, Signal Process. 90 (3) (2010) 727–752.
[3] M.S.A. Karim, K. Wong, Universal data embedding in encrypted domain, Signal Process. 94 (2014) 174–182.
[4] M.S.A. Karim, K. Wong, Data embedding in random domain, Signal Process. 108 (2015) 56–68.
[5] E. Zielinska, W. Mazurczyk, K. Szczypiorski, Trends in steganography, Commun. ACM 57 (2) (2014) 86–95.
[6] Y. Huang, S. Tang, J. Yuan, Steganography in inactive frames of voip streams encoded by source codec, IEEE Trans. Inf. Forensics Secur. 6 (no. 2) (2011) 296–306.
[7] J. Lubacz, W. Mazurczyk, K. Szczypiorski, Vice over IP, IEEE Spectr. 47 (2) (2010) 42–47.
[8] W. Mazurczyk, K. Szczypiorski, Steganography of VoIP Streams, in: Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE, Nov. 2008, pp. 1001–1018.
[9] W. Mazurczyk, VoIP Steganography and Its Detection – A Survey, ACM Computing Surveys, vol. 46, no. 2, Nov. 2013 (article no. 20).
[10] H. Tian, K. Zhou, H. Jiang, Y. Huang, J. Liu, D. Feng, An adaptive steganography scheme for Voice over IP, in: Proceedings of the 2009 IEEE International Symposium on Circuits and Systems, May 2009, pp. 2922–2925.
[11] H. Tian, H. Jiang, K. Zhou, D. Feng, Adaptive partial-matching steganography for Voice over IP Using triple M sequences, Comput. Commun. 34 (18) (2011) 226–2247.
[12] H. Tian, H. Jiang, K. Zhou, D. Feng, Transparency-orientated encoding strategies for Voice-over-IP Steganography, Comput. J. 55 (6) (2012) 702–716.
[13] L.Y. Bai, Y. Huang, G. Hou, B. Xiao, Covert channels based on jitter field of the RTCP header, in: Proceedings of the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Aug. 2008, pp. 1388–1391.
[14] Y. Huang, J. Yuan, M. Chen, B. Xiao, Key distribution over the covert communication based on VoIP, Chin. J. Electron. 20 (2) (2011) 357–360.
[15] H. Tian, R. Guo, J. Lu, Y. Chen, Implementing covert communication over voice conversations with windows live messenger, Adv. Inf. Sci. Serv. Sci. 4 (4) (2012) 18–26.
[16] J. Lubacz, W. Mazurczyk, K. Szczypiorski, Principles and overview of network steganography, IEEE Commun. Mag. 52 (5) (2014) 225–229.
[17] W. Mazurczyk, J. Lubacz, LACK – a VoIP steganographic method, Telecommun. Syst. 45 (2010) 53–163.
[18] W. Mazurczyk, K. Cabaj, K. Szczypiorski, What are suspicious VoIP delays, Multimed. Tools Appl. 57 (1) (2012) 109–126.
[19] C. Kratzer, J. Dittmann, T. Vogel, R. Hillert, Design and evaluation of steganography for Voice-over-IP, in: Proceedings of 2006 IEEE International Symposium on Circuits and Systems, May 2006, pp. 2397–2340.
[20] J. Liu, K. Zhou, H. Tian, Least-significant-digit steganography in low bit-rate speech, in: Proceedings of the 47th IEEE International Conference on Communications (ICC), June 2012, pp. 1–5.
[21] H. Tian, K. Zhou, Y. Huang, J. Liu, D. Feng, A covert communication model based on least significant bits steganography in Voice over IP, in: Proceedings of the 9th International Conference for Young Computer Scientists, Nov. 2008, pp. 647–652.
[22] C. Wang, Q. Wu, Information hiding in real-time VoIP streams, in: Proceedings 9th IEEE International Symposium on Multimedia, Dec. 2007, pp. 255–262.
[23] B. Xiao, Y. Huang, S. Tang, An approach to information hiding in low bit-rate speech stream, in: Proceedings of the 27th IEEE Global Telecommunications Conference (GLOBECOM), Dec. 2008, pp. 1–5.
[24] H. Tian, J. Liu, S. Li, Improving security of quantization-index-modulation steganography in low bit-rate speech streams, Multimed. Syst. 20 (2) (2014) 143–154.
[25] O. Chen, W. Wu, Highly robust, secure, and perceptual-quality echo hiding scheme, IEEE Trans. Audio Speech Lang. 16 (3) (2008) 629–638.
[26] J. Rachoń, Z. Piotrowski, P. Gajewski, Authentication in VoIP telephony with use of the echo hiding method, J. Telecommun. Inf. Technol. 3 (no. 1) (2011) 17–21.
[27] W. Mazurczyk, P. Szaga, K. Szczypiorski, Using transcoding for hidden communication in IP telephony, Multimed. Tools Appl. 70 (3) (2014) 2139–2165.
[28] Y. Huang, B. Xiao, H. Xiao, Implementation of covert communication based on steganography, in: Proceedings of the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Aug. 2008, pp. 1512–1515.
[29] E. Xu, B. Liu, L. Xu, Z. Wei, B..Zhao, J. Su, Adaptive VoIP steganography for information hiding within network audio streams, in: Proceedings of the 14th International Conference on Network-Based Information Systems, Sep. 2011, pp. 612–617.
[30] R. Miao, Y. Huang, An approach of covert communication based on the adaptive steganography scheme on Voice over IP, in: Proceedings of the 46th IEEE International Conference on Communications (ICC), June 2011, pp. 1–5.
[31] R. Crandall, Some Notes on Steganography. Available: ⟨http://www2.htw-dresden.de/~westfeld/crandall.pdf⟩, 1998.
[32] A. Westfeld, F5 – a steganographic algorithm, in: Proceedings of the 4th International Workshop on Information Hiding, Apr. 2001, pp. 289–302.
[33] J. Fridrich, D. Soukal, Matrix embedding for large payloads, IEEE Trans. Inf. Secur. Forensics 1 (3) (2006) 390–394.
[34] M. Khatirinejad, P. Lisoněk, Linear codes for high payload steganography, Discret. Appl. Math. 157 (2009) 971–981.
[35] R. Zhang, V. Sachnev, H. Kim, Fast BCH syndrome coding for steganography, in: Proceedings of the 11th International workshop on Information Hiding, June 2009, pp. 48–58.
[36] C. Fontaine, F. Galand, How reed-solomon codes can improve steganographic schemes, EURASIP J. Inf. Secur. 2009 (2009). (article no. 274845).
[37] T. Filler, J. Judas, J. Fridrich, Minimizing embedding impact in steganography using trellis-coded quantization, in: Proceedings of SPIE International Symposium on Electronic Imaging, Security and Forensics of Multimedia, vol. XII, no. 754105, Jan. 2010.
[38] Y. Gao, X. Li, T. Zen, B. Yang, Improving embedding efficiency via matrix embedding: a case study, in: Proceedings of the 16th IEEE International Conference on Image Processing, Nov. 2009, pp. 109–112.
[39] G. Liu, W. Liu, Y. Dai, S. Lian, An adaptive matrix embedding for image steganography, in: Proceedings of the 3rd International Conference on Multimedia Information Networking and Security, Nov. 2011, pp. 642–646.
[40] Y. Tew, K. Wong, An overview of information hiding in H.264/AVC compressed video, IEEE Trans. Circuits Syst. Video Technol. 24 (2) (2014) 305–319.
[41] ITU-T G.711, Pulse Code Modulation (PCM) of Voice Frequencies, International Telecommunications Union, Geneva, Switzerland, 1988.
[42] ITU-T G.711, Appendix II, Pulse code modulation (PCM) of Voice Frequencies Appendix II: A Comfort Noise Payload Definition for ITU-T G.711 Use in Packet-based Multimedia Communication Systems, International Telecommunications Union, Geneva, Switzerland, 2000.
[43] P. ITU-T, 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, International Telecommunications Union, Geneva, Switzerland, 2001.
[44] ITU-T P. 862.3, Application Guide for Objective Quality Measurement Based on Recommendations P.862, P.862.1 and P. 862. 2, International Telecommunications Union, Geneva, Switzerland, 2007.
[45] P. ITU-T, 800, Methods for Subjective Determination of Transmission Quality, International Telecommunications Union, Geneva, Switzerland, 1996.

Please cite this article as: H. Tian, et al., Optimal matrix embedding for Voice-over-IP steganography, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.05.001