Multimed Tools Appl (2016) 75:13557–13583
DOI 10.1007/s11042-015-2759-2

Steganalysis of HUGO steganography based on parameter recognition of syndrome-trellis-codes

Xiangyang Luo [1,2] · Xiaofeng Song [1,2] · Xiaolong Li [3] · Weiming Zhang [4] · Jicang Lu [1,2] · Chunfang Yang [1,5] · Fenlin Liu [1,2]

Received: 4 January 2015 / Revised: 24 April 2015 / Accepted: 17 June 2015 /
Published online: 5 July 2015
© Springer Science+Business Media New York 2015

Abstract Highly Undetectable steGO (HUGO steganography) is a well-known image steganography method proposed in recent years. The security of HUGO steganography is analyzed in this paper, and a corresponding steganalysis method is proposed based on the blind coding parameters recognition. Firstly, the principle of covert communication based on HUGO steganography and the characteristics of the Syndrome-Trellis codes (STCs) used in HUGO are analyzed; and then the potential security risk of HUGO is pointed out; Secondly, based on the idea of the blind parameters recognition for channel coding, the submatrix parameter of

✉ Xiangyang Luo
xiangyangluo@126.com

Xiaofeng Song
xiaofengsong@sina.com

Xiaolong Li
lixiaolong@pku.edu.cn

Weiming Zhang
zhangwm@ustc.edu.cn

Jicang Lu
lujicang@sina.com

Chunfang Yang
chunfangyang@126.com

Fenlin Liu
liufenlin@vip.sina.com

[1] State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China

[2] Zhengzhou Science and Technology Institute, Zhengzhou 450001, China

[3] Institute of Computer Science and Technology, Peking University, Beijing 100871, China

[4] School of Information Science and Technology, University of Science and Technology of China, Anhui 230026, China

[5] Science and Technology on Information Assurance Laboratory, Beijing 100072, China

STCs is recognized correctly, and thus the message embedded by HUGO can be extracted correctly by decode algorithm of STCs. A series of experimental results show that the proposed steganalysis method can not only detect the stego-images reliably, but also extract the embedded message correctly; these validated the existence of security flaw of HUGO steganography.

Keywords Adaptive steganography · Highly Undetectable steGO steganography · Steganalysis · Message extraction · Syndrome-Trellis codes (STCs) · Blind parameters recognition

1 Introduction

So far, many content-adaptive image steganographic methods, such as EA (Edge-Adaptive) [13], HUGO (Highly Undetectable steGO) [21] steganography and so on, are proposed. Compared with the traditional non-adaptive steganography, these content-adaptive steganographic methods can adaptively choose the changeable locations according to the image-content, so that the embedding changes are mainly made in the textured or perceptual complex image regions. By this means, a higher level of stego-security is achieved since the embedding noise is covered by the inhered noise. Among the existing content-adaptive methods, HUGO is a typical and effective spatial domain steganography method, which is designed by minimizing a suitably-defined distortion function using the Syndrome-Trellis Codes (STCs) [22]. As HUGO steganography makes embedding changes in difficult modeling cover regions and preserves high-dimensional image statistics, almost all of the traditional methods [14] of typical hard to detect the stego-images generated by HUGO steganography, and almost current steganalysis methods cannot provide satisfactory detection results even when the detection features' dimension extracted by steganalysis method is rather high. So, the steganalysis of HUGO is important and attracts many researchers' attention and interest. In this paper, we focus on the security issue and steganalysis method of HUGO steganography.

For content-adaptive steganography, a common way is to assign a cost for changing a cover element and then embed a given bit while minimizing the expected embedding costs or distortion. This can be achieved using syndrome-coding methods, such as STCs (Syndrome-Trellis Codes). HUGO is the first stego method which follows the above embedding paradigm. After HUGO steganography algorithm is proposed, many adaptive steganography algorithms followed the same embedding paradigm have been proposed [2, 4, 5, 7, 16, 23]. In [9], it is pointed out that the HUGO is unsafe when the parameter $T=90$, and this flaw can be eliminated by setting the threshold T as 255. In [4], a high security-level algorithm WOW (Wavelet Obtained Weights) steganography was proposed to change pixels of texture regions while preserving the edge of cover image. In [5], S-UNIWARD steganography was proposed using the directional high-pass filter, and obtained an advanced security. Besides the above spatial domain steganography algorithms, many adaptive JPEG steganography algorithms also were proposed. For example, a new JPEG steganography algorithm, MOD (Model Optimized Distortion) steganography was proposed in [23]. MOD can minimize the detectability by parametrizing the distortion and determining the best parameters by optimization. But the MOD has a security flaw as the distortion is optimized for incomplete cover model [8, 9]. In [2], the UED (Uniform Embedding Distortion) steganography was proposed. The distortion function of UED is defined based on the magnitude of the DCT coefficients and both their intra- and inter-block neighborhood coefficients. By the distortion metric, UED tries to modify

nonzero quantized DCT coefficients with equal probability and lead to as miminal as possible artifacts for statistics of DCT coefficients in general. In [7], a side-informed JPEG steganography algorithm is proposed, which utilizes the perturbation error, the quantization step, and the magnitude of quantized DCT coefficient to define the embedding distortion function. In addition, in literature [5], two JPEG image steganography algorithms named J-UNIWARD and SI-UNIWARD are also proposed, the former define the embedding distortion according to the sum of relative changes of the wavelet coefficients between the cover and stego JPEG image, the latter further utilize the side information obtained by compressing the higher-quality image to the cover image to define the embedding distortion. In [16], a new embedding distortion function of steganographic scheme in JPEG images is defined according to both the coefficient residual and coefficient value, and the parameters of the proposed distortion function are optimized with an exhaustive searching method. It can be seen from the above literatures, as a kind of classical adaptive steganography method, HUGO plays a very important role in the development process of adaptive steganography.

With the presentation of improved HUGO and other adaptive steganography algorithms with same framework as HUGO, the detection methods depended on traditional image features (such as PDF moment, CCF moment, Markov matrix and so on [24, 15]) has been difficult to implement effective detection. To this end, the researchers put forward some new steganalysis methods, to reliably detect the stego-images based on HUGO series steganography. Considering HUGO steganography constrains the embedding changes to texture or noise image regions, Fridrich. J et al. [1] use diverse image noise residuals according to the higher-order local models of images to form the co-occurrence matrix features and capture the vary of image statistical character after embedding change. In this method, the ensemble classifiers [10] are utilized to improve detection results, and obtain the accuracy rate of 83.90 % when relative payload is 0.4bpp (bits per pixel) and the dimension of employed features is 33,963. Gül. G et al. [3] use linear and non-linear filtering functions to get image noise residuals, and then estimate the probability density function to form detection features and further optimized the extracted features. In this method, SVM (Support Vector Machine) classifier is utilized to detection stego-images, the accuracy is 84.10 % at 0.4bpp and the feature dimension is 1273. Considering the most embedding changes occurred in the texture regions caused by HUGO steganography, Y. Q. Shi et al. [19] construct rich image residual image and extract the local binary patterns textural features as steganalysis features. The detection accuracy is 83.92 % at 0.4bpp and the employed feature dimensions are 22,593. On the basis of [1], Fridrich. J et al. present the spatial rich model features by enrich the types of residual images in [11], the detection accuracy is 86.99 % at 0.4bpp and the dimension of features is 34,671. In [6], Holub. V et al. project neighboring residual samples onto a set of random vectors, and take the first-order statistics (histogram) of the projections as the detection features. The corresponding detection accuracy is 88.28 % for the stego-images with 0.4bpp relative payload and 12,870 features are used. In the above content, we summarized the existing steganalysis methods' detection accuracy for HUGO steganography while the embedding rates are 0.4bpp, however, as the relative payload goes lower, the detect accuracy will descend obviously. For example, the detection accuracy of method in [6] will be reduced to 76.03 % when relative payload is 0.2bpp.

As mentioned above, one can see that although the existing research has made great progress on steganalysis methods of HUGO, but there are still some problems worthy of further research: 1) The existing steganalysis methods for HUGO mainly focus on detecting the existence of embedded message, moreover, and the accuracy of detection remains to be

improved significantly; 2) So far, there is no literature discussing the problem of extracting the embedded message of HUGO steganography; 3) Due to the introduction of minimum embedding distortion, almost all existing literatures believe that the HUGO has strong anti-detection capabilities and very high security, the series of HUGO steganography has become competitors of steganography and the target of steganalysis method, lack of study to discuss the security defects of HUGO steganography.

It is worth noting that, although the distortion function of the above adaptive steganography algorithms are different, the coding methods are all STCs, and these steganography methods all follow the framework of "Minimizing the distortion function + STCs coding". However, we know that main purpose of HUGO, WOW, UED and other steganography methods with superior performance adopt the STCs coding is to minimize the embedding distortion and resist the statistical detection, but ignored the risk of introducing STCs may bring. Although the STCs coding can contribute to reduce the embedding distortion, it is not designed for resistance of the embedded message extraction. The main work of this paper include: 1) the principle of HUGO steganography is introduced and analyzed firstly, including the distortion function and STCs, then point out the difference between HUGO and other existing typical steganography; 2) on the base of estimating the embedding ratio, a steganalysis method to HUGO is proposed based on blind parameters recognition of STCs; 3) the feasibility and time complexity of the proposed algorithm is discussed subsequently; 4) the effectiveness of the proposed method is verified through experiments.

The rest of this paper is organized as follows. In Section 2, HUGO steganography is described briefly and its potential security risk is analyzed; In Section 3, a steganalysis method to HUGO based on blind parameters recognition of STCs is proposed and described; The feasibility and time complexity of proposed method is analyzed in Section 4; In Section 5, the experimental results and analysis are shown, and the conclusions are given in Section 6.

2 Brief description and potential security risk analysis of HUGO steganography

HUGO provides high anti-detection ability by preserving high-dimension statistical image model and constraining embedding changes on the hard-to-model regions of cover image. The main steps of HUGO are as follows, the distortion function is defined as a weighted difference of SPAM (Subtractive Pixel Adjacency Matrix) [20] feature vectors firstly, and then the STCs is used to choose the changeable pixels and minimize the distortion function. Lastly, model correction is used to determine the values of the changed pixels and the messages are embedded. The individual steps of HUGO are depicted in Fig. 1. From Fig. 1, it can be seen that HUGO steganography includes three main process, namely distortion computation, coding and model correction. The following will mainly introduce and analyze the distortion function and coding algorithm, and then point out the security risk of HUGO steganography.

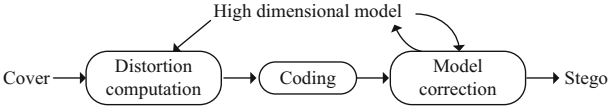


Fig. 1 High-level diagram of HUGO [21]

2.1 Embedding distortion function of HUGO steganography

HUGO steganography keeps the minimizing distortion function in touch with preserving the high dimension statistical model of cover image. The distortion function of HUGO steganography is defined according with the SPAM feature, which is widely used in the current steganalysis methods. HUGO steganography can minimize the distortion function while maintaining the high dimensional statistical SPAM feature of cover image. The distortion function $D(\mathbf{X},\mathbf{Y})$ is defined as a weighted sum of differences:

$$D(X,Y)=\sum_{d_1,d_2,d_3=-T}^{T}\left[w(d_1,d_2,d_3)\cdot\left|\sum_{k\notin\{\rightarrow,\leftarrow,\uparrow,\downarrow\}}\left(C^{X,k}_{d_1,d_2,d_3}-C^{Y,k}_{d_1,d_2,d_3}\right)\right|+w(d_1,d_2,d_3)\cdot\left|\sum_{k\in\{\searrow,\searrow,\nearrow,\swarrow\}}\left(C^{X,\,k}_{d_1,d_2,d_3}-C^{Y,\,k}_{d_1,d_2,d_3}\right)\right|\right] \tag{1}$$

where, $\mathbf{X},\mathbf{Y}$ denote cover and stego image respectively, d_1,d_2,d_3 denote the differences between adjacent pixels along k directions, $C^{x,k}_{d_1,d_2,d_3}$ and $C^{y,k}_{d_1,d_2,d_3}$ denote the elements of third-order co-occurrence matrix calculated from the cover and stego images respectively, the threshold T=255, and $w(d_1,d_2,d_3)$ is the weight function, which can be calculated as follows,

$$w(d_1,d_2,d_3)=\frac{1}{\left[\sqrt{d_1^2+d_2^2+d_3^2}+\sigma\right]^{\gamma}} \tag{2}$$

where, $\sigma,\gamma>0$ are adjustable parameters to minimize the ability of stego image to be detected.

From the definition of distortion function of HUGO, we can find that the value of distortion function will be small when the changeable pixels occur in texture regions where the values of d_1,d_2,d_3 are large. Since the texture regions are hardly modeled, HUGO can select changeable pixels adaptively according to the values of distortion function. This improves the security level of steganography markedly. At the same time, need to pay attention to is, HUGO steganography's minimizing embedding distort function is equivalent to keep a very high dimension co-occurrence matrix features, this may be relatively easy to steganography, but bring difficulty for the detection: the high dimensional feature means that machine learning algorithms among steganalysis method need higher dimensional features, more training samples, longer training times and discrimination times, which is unacceptable in practice. The above characteristics make HUGO steganography has better performance of anti-detection.

2.2 Syndrome-trellis codes (STCs)

2.2.1 Principle of STCs

STCs is one coding algorithm used in HUGO steganography, and it belongs to syndrome coding and is often used in other recent proposed steganography algorithms, such as

WOW, UED, J-UNIWARD, etc. The principle of STCs can be simply described as the following: suppose cover object and stego object are denoted as $\mathbf{x}=(x_1,\cdots,x_n)$ and $\mathbf{y}=(y_1,\cdots,y_n)$, $\mathbf{x},\mathbf{y}\in\{0,1\}^n$, respectively, where $\mathbf{m}$ is the embedded message, and the size of $\mathbf{m}$ is m. Then the embedding and extraction of message $\mathbf{m}$ can be realized by using a linear code C of length n and dimension $n-m$:

$$\mathbf{y} = \mathrm{Emb}(\mathbf{x},\mathbf{m}) = \arg\min_{\mathbf{y}\in C(\mathbf{m})} D(\mathbf{x},\mathbf{y}) \tag{3}$$

$$\mathbf{m} = \mathrm{Ext}(\mathbf{y}) = \mathbf{H}\mathbf{y} \tag{4}$$

where $D(\mathbf{x},\mathbf{y})$ is embedding distortion function, $\mathbf{H}\in\{0,1\}^{m\times n}$ is a parity-check matrix of the code C, $C(\mathbf{m})=\{\mathbf{z}\in\{0,1\}^n|\mathbf{H}\mathbf{z}=\mathbf{m}\}$ is the coset corresponding to syndrome $\mathbf{m}$, $D(\mathrm{x},\mathrm{y}) = \sum_{i=1}^{n}\rho_i|x_i-y_i|$ is embedding distort function, and all operations are in binary arithmetic.

From Eq. (3), it can be seen that the purpose of coding algorithm is to find the optimal $\mathbf{y}$ subjected Eq. (4) to minimize $D(\mathbf{x},\mathbf{y})$. To obtain a solution, STCs code represents $\mathbf{y}$ as a path of the trellis figure by constructing a special form of parity-check matrix $\mathbf{H}$, where $\mathbf{H}\mathbf{y}=\mathbf{m}$. The optimal $\mathbf{y}$ closest to $\mathbf{x}$ is then found by Viterbi algorithm. Here, the parity-check matrix $\mathbf{H}$ is composed of a small submatrix $\hat{\mathbf{H}}$ of $h\times w$ lined up along main diagonal direction. The height h of the submatrix is a design parameter that affects the coding algorithm speed and efficiency, the value of h usually is $6\le h\le 13$ [21]. w is the width of $\hat{\mathrm{H}}$, and is dictated by the relative payload α: if α is equal $1/k$ for some $k\in N$, select $w=1/\alpha=k$, otherwise, if $1/(k+1)<\alpha<1/k$, the matrix $\mathbf{H}$ will contain a mix of submatrices of width k and $k+1$, the final matrix $\mathbf{H}$ is of size $[\alpha\cdot n]\times n$, where $[\bullet]$ is the operation of rounding.

For the construction of parity-check matrix $\mathbf{H}$, an example is given as following. Suppose the size of stego object $\mathbf{y}$ is 10 and the relative payload $\alpha=0.5$, then the width w of corresponding submatrix should be $w=1/0.5=2$. So, the submatrix $\hat{\mathrm{H}} = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}$ can be selected, and the parity-check matrix $\mathbf{H}$ formed by $\hat{\mathbf{H}}$ is as follows,

$$\begin{pmatrix}
1 & 0 & & & & & & & & \\
1 & 1 & 1 & 0 & & & & & & \\
 & & 1 & 1 & 1 & 0 & & & & \\
 & & & & 1 & 1 & 1 & 0 & & \\
 & & & & & & 1 & 1 & 1 & 1
\end{pmatrix} \tag{5}$$

Furthermore, if relative payload $\alpha=0.4$, that is $1/3<\alpha<1/2$, then we need to select two submatrices with different width to generate the parity-check matrix $\mathbf{H}$. Here, if the submatrices $\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}$ and $\begin{pmatrix}1 & 0 & 1\\ 1 & 1 & 0\end{pmatrix}$ are selected to form the parity-check $\mathbf{H}$,

the final **H** will be as follows,

$$\begin{pmatrix} 1 & 0 & & & & & & & & \\ 1 & 1 & 1 & 0 & 1 & & & & & \\ & & 1 & 1 & 0 & 1 & 0 & & & \\ & & & & & 1 & 1 & 1 & 0 & 1 \end{pmatrix} \tag{6}$$

HUGO steganography uses STCs coding to embed and extract message, and the message size m and the submatrix $\hat{\mathrm{H}}_{h\times w}$ can be used as shared parameters of the sender and the receiver. At the receiver, only when the shared parameters are obtained properly, the parity-check matrix **H** can be generated correctly, and the message can be extracted correctly by **H** and the stego object **y**.

2.2.2 Coding parameters of STCs

From the above introduction of STCs coding, it can be seen that the message size m and the submatrix $\hat{\mathbf{H}}$ are two important parameters for message embedding and extraction.

1) For the message size m, it equals to the product of relative payload α and the cover object length n, and it also equals to the number of rows of matrix **H**. When cover object is embedded by HUGO, m is used as shared parameters and transmitted to the receiver, and will be used for the extraction of embedded message.
2) For the submatrix $\hat{\mathbf{H}}$, it belongs to a kind of constructive matrix, all the elements of $\hat{\mathbf{H}}$ are 0 or 1. The design of $\hat{\mathbf{H}}$ is very important because the stego object and embedding distortion are determined by $\hat{\mathbf{H}}$. In the structure of the matrix $\hat{\mathbf{H}}$, it is necessary to determine its height h, the width w, and the matrix elements $\hat{\mathbf{H}}_{ij}$, where h determines the speed and performance of the steganography. The larger h is, the more paths through the trellis can be obtained, thus we can find **y** which is closer to **x**, but the coding time will increase. The width w is determined by the embedding ratio.

The literature [22] discussed the design method of $\hat{\mathbf{H}}$ and gave some important conclusions. For a variety of different heights and widths of possible submatrix, the authors adopt the method of exhaustive search to find the design rules for optimal parity-check submatrix. They found that the parity-check submatrix has the following characteristics can get good embedding efficiency: 1) all elements in the first and last rows of the parity-check submatrix are 1, and other elements are randomly assigned 0 or 1; 2) The same column should not be in the parity-check submatrix because the syndrome trellis would contain two or more different paths with exactly the same weight, which would lead to an overall decrease in performance. By searching, the literature [22] got 3 matrix with good performance, as

shown in Eqs. (7) ~ (9).

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{7}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{8}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \tag{9}$$

For the submatrices in Eqs. (7) ~ (9), if the elements of each column is regarded as an integer's binary representation, and the first rows is the least significant bit of corresponding integer, then the decimal representation of the submatrices in (7) ~ (9) are $\hat{\mathbf{H}}_1 = [109,\ 71]$, $\hat{\mathbf{H}}_2 = [95,\ 117,\ 105,\ 97,\ 75]$, $\hat{\mathbf{H}}_3 = [181,\ 215,\ 243]$. In order to facilitate the representation, in the rest of this paper, all submatrices are using decimal integer form.

2.3 Potential security risk of HUGO steganography

From the above introduction and analysis for HUGO, it can be seen that the selection of the changeable pixels and the messages extraction in HUGO are both depended on STCs. The process of message transmission by HUGO steganography can be described as Fig. 2. It can be found that the covert communication by HUGO is completely different from those traditional steganography method which depend on the embedding key to determine the embedding position and extract the embedded message using extraction key (such as LSB replacement, LSB matching, F5 [25], OutGuess [17], and so on).

For stego-image generated by HUGO, if the size of the embedded secret message and the submatrix parameters shared by the sender and the receiver can be recognized correctly by a steganalyzer, thus the steganalyzer can construct the parity-check matrix, and will be able to extract the embedded message just like a normal receiver! Obviously, the key steps of this kind

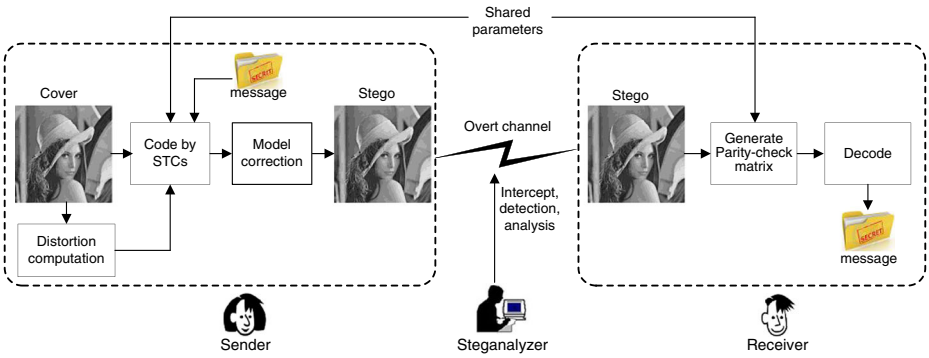


Fig. 2 Covert communication based on HUGO steganography

of steganalysis are to estimate the size of embedded message and analyze the parity-check matrix of STCs. In the next section, we will describe the steganalysis method of HUGO which can extract the embedded plaintext message accurately, and give the estimation and recognition algorithms for the above two parameters (the size of embedded message and the parity-check matrix of STCs).

3 Steganalysis method based on blind parameter recognition

From the principle of HUGO steganography, we know that the message size m and submatrix $\hat{\mathbf{H}}$ are two important parameters used in STCs coding process, which is also need to be shared between the sender and the receiver of communication. Due to the ranges of the parameters m, h and w are finite, a table of coding parameters can be constructed off-line. If the message size m can be estimated correctly and the submatrix $\hat{\mathbf{H}}$ can be looked out, we can construct the parity-check matrix $\mathbf{H}$ and obtain the correct decode sequence. On the contrary, if we can examine the correctness of the decode sequences for different parameter m and the corresponding $\hat{\mathbf{H}}$, the correct parameters m and $\hat{\mathbf{H}}$ of STCs can be obtained, and the decode sequence passed the examination just is the message embedded in $\mathbf{y}$. In the following, the principle framework of the proposed steganalysis method to HUGO will be given firstly, and then the main process of the proposed method will be described in detailed.

3.1 Principle framework of steganalysis method based on blind parameter recognition

One of the main works of this paper is to present an embedded message extraction method of HUGO steganography based on blind recognition of STCs parameters. The principle architecture of this method is shown in Fig. 3.

The main steps of the proposed steganalysis method described in Fig. 3 are as following:

Step 1: Construct the table of coding parameters. The coding parameters table needs to be constructed firstly. Since the message size m equals to the product of relative payload α and the size of cover object n, the coding parameters table can be constructed according to the possible value and range of parameters α and $\hat{\mathbf{H}}$.

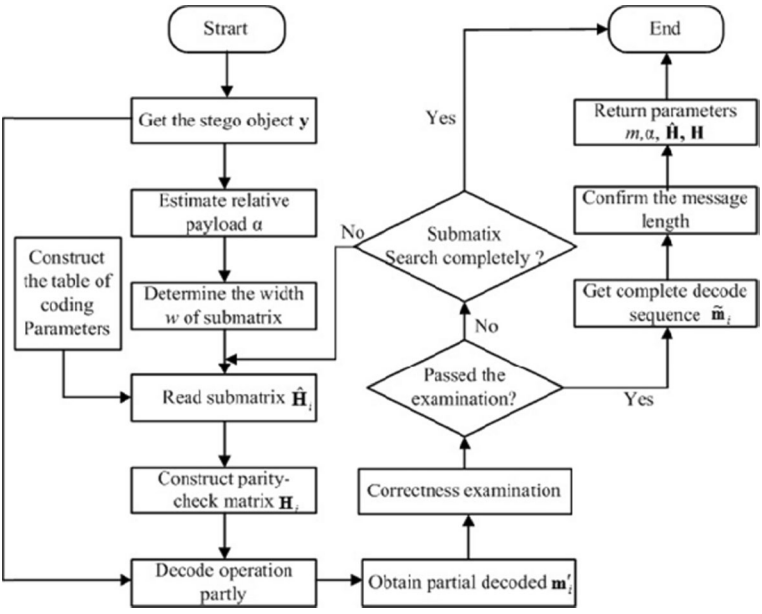


Fig. 3 Principle framework of steganalysis method based on blind recognition of STCs parameters

Step 2: Get the stego object y. Due to the message is embedded in the LSBs of cover pixels, so one can obtain the stego object **y** from the suspicious test image, where the stego object **y** should be obtained according to the same order as the order of message embedded in the cover image.

Step 3: Estimate the relative payload α. It is worth noting that, the proposed method does not need to obtain accurate relative payload α of stego image, and only need to know the approximate range of α. The relative payload α can be estimated by the quantitative steganalysis method proposed in literature [12], then the range of α can be determined by simple error processing. For example, if the estimated value of α is $\hat{\alpha}$, and the possible maximum error is δ, then the searching range of the relative payload is $[\hat{\alpha}-\delta, \hat{\alpha}+\delta]$. (For example, if the estimated value $\hat{\alpha}$ = 0.43, and the experiential maximum error δ =0.05, then the searching range of α is [0.38, 0.48]).

Step 4: Search the submatrix $\hat{\mathbf{H}}$. After the searching range of the relative payload was ascertained, the width w of submatrix $\hat{\mathbf{H}}$ can be calculated, and then we need to search for the corresponding submatrix $\hat{\mathbf{H}}$ with width w from the coding parameters table. Here, the submatrix with small height h should be searched firstly.

Step 5: Construct the parity-check matrix $\mathbf{H}_i$. According to the searching range of α, and stego object **y**, the size of message **m** can be estimated, and then the parity-check matrix $\mathbf{H}_i$ can be generated by the parameter m and the corresponding submatrix $\hat{\mathbf{H}}_i$.

Step 6: Decode and Obtain partial decode sequence $\mathbf{m}'$. According to the matrix $\mathbf{H}_i$ and stego object **y**, the partial decode sequence $\mathbf{m}_i'$ can be obtained by decoding algorithm of HUGO steganography. Here, only the partial decode sequence is needed because the correction of sequence can be determined by partial sequence.

Step 7: Examine the correctness of sequence $\mathbf{m}_i'$**.** The purpose of examining the correctness of decode sequence is to determine the correctness of the current parameters

of STCs. This can be realized by many methods, such as the randomness test or other statistical analysis methods. The advantage of partial decode is to avoid the relative payload estimation error influence the correctness of decoded sequence and reduce the time consumption of correctness examination.

Step 8: Get complete decode sequence $\tilde{\mathbf{m}}_i$. When the partial decode sequence $\mathbf{m}_i^{'}$ passes through the correctness examination, it can be considered that the current coding parameters are right. Thus, the complete decode sequence $\tilde{\mathbf{m}}_i$ can be obtained according the current parameters $\hat{\mathbf{H}}_i$ and the maximum relative payload $\hat{\alpha} + \delta$.

Step 9: Confirm the message length. For the real length of embedded message is usually less than that of the complete decode sequence $\tilde{\mathbf{m}}_i$, usually some codewords are redundant in the tail of $\tilde{\mathbf{m}}_i$. We need to confirm the length of embedded message, and get the embedded message $\mathbf{m}$. One way we can adopt is random testing on the tail codewords of $\tilde{\mathbf{m}}_i$.

Step 10: Return the coding parameters The algorithm return the message $\mathbf{m}$ and the corresponding parameters α, $\hat{\mathbf{H}}$, $\mathbf{H}$. Where, the parameter α can be calculated according to the embedded message $\mathbf{m}$ and the stego object $\mathbf{y}$.

In next several subsections, the above main steps will be described in detail.

3.2 Construct the coding parameters table

The construction of coding parameters table is a basis of blind recognition of STCs parameters. When we construct the coding parameters table, the relative payload α should be connected with the entire corresponding submatrix $\hat{\mathbf{H}}$. For example, if the parameter α=1/2, then the corresponding $\hat{\mathbf{H}}$ should be the entire submatrices with the width w=2 and a limited height h (h in a set range); if the parameter $1/2<\alpha<1/3$, the corresponding $\hat{\mathbf{H}}$ should be the entire submatrices with width w=2 or 3 and a limited height h. In addition, the different parameters α maybe correspond to the submatrices with same width, such as for the parameters α_1=0.35, α_2=0.4, the width w of corresponding submatrices are all 2 or 3.

According to the above description, we can see that the construction of STCs coding parameters table is associating the relative payload with its corresponding all possible submatrices (viz. sub parity-check matrices), thus relative payloads and the corresponding submatrices constitute the coding parameters table.

3.3 Obtain the stego objects

The blind recognition method of STCs parameters needs according to coding parameters (the relative payload and sub parity-check matrix) of the current search to generate a parity check matrix firstly, and then combines the stego sequence to obtain a corresponding decoding sequence further, finally determines whether the current encoding parameters are correct according to the correctness test result of decoding sequence. Therefore, to recognize the coding parameter of STCs, we firstly need to extract the stego sequence from the image to be detected. For HUGO steganography, the cover object $\mathbf{x}$ is formed by arranging the LSBs of image pixels in an order, so the stego object $\mathbf{y}$ should be got by the same order from the test image. Usually, the cover object $\mathbf{x}$ is formed by column (or row) scan, the stego object $\mathbf{y}$ also should be formed by column (or row) scan. Moreover, it is worth noted that this paper's main aim is exposing the security risk of using STCs coding in HUGO steganography, so how to

obtain the stego object **y** is not a sticking point, in despite of it is also an important problem worth to study.

3.4 Estimate the relative payload of stego image

The message size m is one important parameter of STCs coding. Only when the parameter m has been obtained, the parity-check matrix **H** can be constructed by combining the parameters m and submatrix $\hat{\mathbf{H}}$. To get the parameter m, we can firstly estimate the relative payload α by a quantitative steganalysis method, and then parameter m can be obtained by α multiplied by the size of stego object **y**. For estimating the relative payload, we can adopt the quantitative steganalysis method published in existing literatures (For example, by adopting the method of [12]). And based on the estimated value, we also need to consider the error range of relative payload, to determine the search scope of relative payload. That is to say, although the relative payload α can be estimated by a quantitative steganalysis method, however the estimate value is often not accurate, so the search range of parameter α should be determined by combining the possible maximum estimation error δ, which is an experience value can be obtained by a lot of experiments.

3.5 Search the submatrix

In the recognition process of STCs coding parameter, the search range of submatrix is linked with the currently selected relative payload, that is, when the search range of parameter α has been determined, for each α_i within the search range, the corresponding submatix $\hat{\mathbf{H}}$ can be searched from coding parameters table. If the current parameter $\alpha_i=1/k$ (k is an integer), then the submatrix with $w=1/\alpha_i$ should be selected; otherwise, if $1/(k+1)<\alpha_i<1/k$, then two submatrices with $w=k$ and $w=k+1$ should be selected to construct the submatrix at the same time.

For the parameters of parity-check matrix, when the width is fixed, the height h larger, the corresponding search range is bigger also. Therefore, when searching the sub parity-check matrix, we should search all possible sub parity-check matrixes from small to large height values. Namely, we should search for the submatrix $\hat{\mathbf{H}}$ with small height h firstly when the exact height h is not known.

3.6 Decoding operation

When the length n of stego object **y**, the estimation of the relative payload α, and the parity-check matrix H generated by sub parity-check matrix $\hat{\mathrm{H}}$ are obtained, we can get the corresponding decoding sequence by performing the decoding operation according to the stego object **y** and parity-check matrix H. The decoding operation of STCs is closely related to the coding process. In Fig. 4, an example of the syndrome trellis processed by the Viterbi algorithm is given (The example was shown in literature [22]), where the cover sequence x=(1,0,1,1,0,0,0,1) and the secret message m=(0,1,1,1), STCs coding can transform the stego sequence y into a path in the trellis diagram, and adopt the Viterbi algorithm to find the satisfied a stego sequence with minimum embedding distortion to

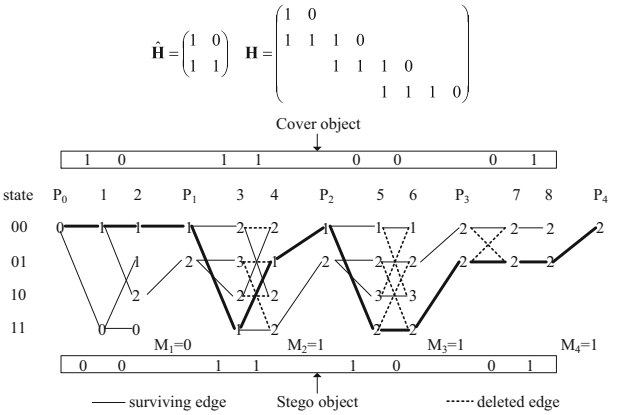


Fig. 4 An example of the syndrome trellis processed by the Viterbi algorithm [22]

satisfy the Eq. (4), finally obtains the stego object y=(0,0,1,1,1,0,0,1). The corresponding path of stego object **y** is shown with thick edge in Fig. 4, and more details about this example can be got in literature [22].

According to the principle of STCs, the corresponding path of stego object **y** can be obtained when the parity-check **H** has been got. Then, the decode sequence would be obtained by the path through the trellis. For details, from the state P_0, the each element of stego object **y** is relative to one edge of syndrome trellis, if the element equals to zero, the relevant edge is horizontal; otherwise, if the element do not equals to zero, the relevant edge is determined by the current state and the corresponding column of submatrix $\hat{\mathbf{H}}$. Finally, all the obtained edge are combined to form the path, and the least significant bit of corresponding state of the end column of each block is obtained as decode sequence. Notice that the block of syndrome trellis is corresponding to the submatrix.

In addition, because the correction examination of decode sequence can be tested by partial decoding result, to reduce the time consumption, the partial decode sequence $\mathbf{m}'$ can be obtained firstly. If the sequence $\mathbf{m}'$ can pass through correction examination, we continue to get whole decode sequence, otherwise, we should search for other coding parameters, until find out the correct parameters or traversing the entire parameter table.

3.7 Correctness examination of decode sequences

In fact, once an attacker can test the decoding sequence is correct, he/she can determine the coding parameters, and obtain the embedded information. In this paper, by embedding plaintext information as an example, we present a method of examination correctness of decoding sequence. The main objective is to validate that using STCs code to embedding message have a security risk, even though it is adaptive widely used by adaptive steganography methods in recent years, such as HUGO, WOW, S-UNIWARD, and so on.

In this paper, we adopt the randomness test to examine the correctness of decode sequence and coding parameters. As we known, when both sides of communication using the image carrier as a covert communication, the message transmitted by sender to the receiver are usually meaningful (the embedded message is often real document, picture, character sequences, etc.). In case of the message without encryption, the corresponding binary

sequences's randomness usually is very poor and different from that of a random (and Pseudo-random) sequence. So in STCs coding parameter recognition, we can determine whether the coding parameters are properly recognized according to the results of randomness examination of decoding sequences. For binary sequence, there are many methods [18] can be used for randomness examination, and the run test is selected in this paper. The processes of run test are described as following:

Step 1: For the binary sequence $\varepsilon=\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$, the proportion value ϕ is computed by $\phi = \left(\sum_{j=1}^{n} \varepsilon_j\right)/n$, and a small constant $\tau = 2/\sqrt{n}$ is given. Then, if $|\phi-0.5|\geq\tau$, the sequence ε is not random, end the test, otherwise, continue to the test.

Step 2: Compute a statistical value V_n:

$$V_n = \sum_{k=1}^{n-1} r(k) + 1 \tag{10}$$

Where

$$r(k) = \begin{cases} 0, & \varepsilon_k = \varepsilon_{k+1} \\ 1, & \varepsilon_k \neq \varepsilon_{k+1} \end{cases}.$$

Step 3: Compute the complementary error function (erfc) value P:

$$P = \mathrm{erfc}\left(\frac{|V_n - 2n\phi(1-\phi)|}{2\sqrt{2n}\phi(1-\phi)}\right) \tag{11}$$

Step 4: Compare the P with a significance level *alpha*, if $P \geq alpha$, the sequence ε can pass through the run test and is random, otherwise, ε is not a random sequence.

4 Performance analysis of proposed steganalysis method

In this section, we will discuss the feasibility of the proposed steganalysis method from the construction of STCs coding parameters table and correction examination of decode sequence. Then the search range of submatrix $\hat{\mathbf{H}}$ will be computed according to the relative payload α and the time complexity of steganalysis algorithm will be also analyzed.

4.1 Feasibility analysis

The keys of the proposed steganalysis method include the estimation of relative payload α, search of submatrix $\hat{\mathbf{H}}$, and the correctness examination of decode sequences, and the feasibility of these three aspects will be expounded respectively in the following.

(1) Estimation of relative payload α. The coding parameters table constructed for the proposed method consists of relative payload α and submatrix $\hat{\mathbf{H}}$. In the existing

literature, there have some quantitative steganalysis method, these methods can be used to estimate the information embedding ratio for a specific adaptive steganography which based on distortion function designing and STCs coding. For example, for the parameters α of HUGO steganography, in [12], J. Kodovský and J. Fridrich have proposed a quantitative steganalysis to estimate the relative payload α of HUGO. So, the range of parameter α can be determined according to the estimated value and the possible maximum error is δ.

(2) Search of submatrix $\hat{\mathbf{H}}$. As to submatrix $\hat{\mathbf{H}}$, the height h and the width w are often small, so the construction of the parameters table of $\hat{\mathbf{H}}$ is feasible according to all possible values of h and w. Furthermore, the literature [22] has discussed the construction and adoption methods of $\hat{\mathbf{H}}$, and some good submatrix $\hat{\mathbf{H}}$ was recommended, this provide much help for the construction of parameters table. Based on the coding parameters table, the blind parameters recognition method of STCs travel through all possible relative payload α and corresponding submatrix $\hat{\mathbf{H}}$, and can obtain the decode sequence by decoding algorithm of STCs.

(3) Correctness examination of decode sequence. The current parameters α and $\hat{\mathbf{H}}$ can be judged by analyzing the randomness of decode sequence. We know that, if the embedded message sequence is plaintext, which is often not completely. Therefore, if current α and $\hat{\mathbf{H}}$ are correct, then the decoded sequence is the embedded message sequence, and its randomness is weaker than that of a normal random binary sequence. On the contrary, as long as one of the current α and $\hat{\mathbf{H}}$ is error, the decoded sequence is not the embedded message, and it is extracted from the image pixels in a random sequence which randomness is often strong. Thus, by analyzing random coding sequence, one can determine the current encoding parameters are correct or not. In this paper, the run test algorithm is adopted to examine the randomness of decode sequence. The run test is a classical algorithm for randomness test, and it has powerful ability to quantify the randomness of a binary sequence.

4.2 Time complexity analysis

The time consuming of the STCs parameters recognition algorithm is mainly determined by the search process for right relative payload α and submatrix $\hat{\mathbf{H}}$. For each parameter α, the search range of $\hat{\mathbf{H}}$ consist of all possible submatrices with a variety of height h and width w. For a submatrix $\hat{\mathbf{H}}$, if the height is h and width is w, then number of all possible forms of $\hat{\mathbf{H}}$ is $2^{h\cdot w}$. Furthermore, according to the conclusions drawn in literature [22], a good submatrix should not has identical columns, and the elements should generally meet that both the first line and last line are 1. So the search range N can be reduced as follows,

$$N = 2^{(h-2)\cdot w} - 2^{h-2} \cdot \left(C_w^2 + C_w^3 + \cdots + C_w^w\right) \tag{12}$$

When the secret message is embedded into a cover image by HUGO steganography, the relative payload α is different, the width w of the submatrix is different, i.e., α is higher, the corresponding w is smaller, otherwise, w is larger. Below, we take the HUGO stego-images with three kinds of embedding ratios (α is 0.5bpp, 0.3bpp and 0.1bpp) as examples, analyze the size of submatrix $\hat{\mathbf{H}}$ parameter's search range.

(1) If the relative payload α=0.5, then the width of submatrix w=1/α=2. According to (12), the search range of corresponding submatrix should be $N=2^{(h-2)\cdot 2}-2^{h-2}$. For the submatrix $\hat{\mathbf{H}}$ may be with different height h, so all possible submatrices with width w=2 should be search from the smallest h to largest h, therefore, the search range of $\hat{\mathbf{H}}$ is $\sum_h(2^{(h-2)\cdot 2}-2^{h-2})$. Because the height, in general, has $6\le h\le 13$, hence the search range is small when relative payload α=0.5.

(2) If the relative payload α=0.3, then $1/4<\alpha<1/3$, the parity-check matrix $\mathbf{H}$ is formed by submatrices with width w=3 and w=4. So, the two submatrices with w=3 and w=4 should be searched at the same time, the search range of $\hat{\mathbf{H}}$ is

$$\sum\nolimits_h \left(2^{(h-2)\cdot 4}-2^{h-2}\cdot 11\right)\times\left(2^{(h-2)\cdot 3}-2^{h-2}\cdot 4\right) \tag{13}$$

(3) If the relative payload α=0.1, then the width w of submatrix is ten, the search range of submatrix is

$$\sum\nolimits_h 2^{(h-2)\cdot 10}-2^{h-2}\cdot\left(C_{10}^{2}+C_{10}^{3}\cdots+C_{10}^{10}\right) \tag{14}$$

In this case, the search range of submatrix is large.

For the STCs parameters recognition and message extraction, when the relative payload α is given, the corresponding submatrix is searched in ascending order of height h. The search range and time complexity of submatrix with different relative payload from 0.1bpp to 0.5bpp is given in Table 1, where $h_{\max}$ denotes the max value of height h.

From the above analyses and the results of Table 1, it can be seen that the search range of submatrix $\hat{\mathbf{H}}$ is relatively small when the relative payload α is large. So, in this case it is easy to find the right STCs parameters and extract the embedded message. However, when the relative payload α is small, the search range will become larger, and the time consuming of STCs parameters recognition and message extraction will become relatively large.

Table 1 Search range of submatrix and time complexity with different payload

Payload	Search range of submatrix	Time complexity
0.1 bbp	$\sum_h(2^{(h-2)\cdot 10}-2^{h-2}\cdot(C_{10}^{2}+C_{10}^{3}\cdots+C_{10}^{10}))$	$O(2^{(h_{\max}-2)}\cdot 10)$
0.2 bbp	$\sum_h(2^{(h-2)\cdot 5}-2^{h-2}\cdot(C_{5}^{2}+C_{5}^{3}\cdots+C_{5}^{5}))$	$O(2^{(h_{\max}-2)}\cdot 5)$
0.3 bbp	$\sum_h(2^{(h-2)\cdot 3}-2^{h-2}\cdot(C_{3}^{2}+C_{3}^{3}))\times(2^{(h-2)\cdot 4}-2^{h-2}\cdot(C_{4}^{2}+C_{4}^{3}+C_{4}^{4}))$	$O(2^{(h_{\max}-2)}\cdot 12)$
0.4 bbp	$\sum_h(2^{(h-2)\cdot 2}-2^{h-2})\times(2^{(h-2)\cdot 3}-2^{h-2}\cdot(C_{3}^{2}+C_{3}^{3}))$	$O(2^{(h_{\max}-2)}\cdot 6)$
0.5 bbp	$\sum_h(2^{(h-2)\cdot 2}-2^{h-2})$	$O(2^{(h_{\max}-2)}\cdot 2)$

5 Experimental results

To verify the validity of the proposed method, some experiments are done in the following circumstances: the operating system is Microsoft Win 7, CPU is Intel i5-2400 3.10GHz, Memory is 4GB, and the program language is Visual Studio 2008. In this paper, the significance level *alpha* is set as 0.05. The experiments mainly include three parts as the following:

1) The experimental materials and parameters;
2) Validating the effectiveness of the proposed steganalysis method;
3) The time consuming of the proposed method under different relative payload α;

5.1 Experimental setup

5.1.1 The generation of stego images

In experiments, Lena image shown in Fig. 5b is converted to binary sequence and then the binary sequence is embedded into the cover image (Fig. 5a) by HUGO. The obtained stego image is shown in Fig. 5c. Here, the cover object $\mathbf{x}$ is gotten by scanning cover image according to columns and getting the LSBs of scanned pixels. The size of cover image is 512×512, and the size of Lena image is 128×128. So the relative payload of stego image is $\alpha=1/2$, and we can choose submatrix $\hat{\mathbf{H}} = [109,\ 71]$, with the height $h=7$ and the width $w=2$.

For the embedded message obtained from Lena image is meaningful message, the return value P of the run test for embedded message is 0, so the message sequence is not random.

5.1.2 The training of relative payload estimator

The relative payload α of stego image is estimated by the quantitative steganalysis method proposed in [12]. In the experiment, 5000 images are chosen randomly from the BossBase1.01,[1] the stego images with different relative payload α are generated firstly, then 4000 images are chosen randomly from the stego images as training images, the rest 1000 images are used as validation images, lastly, the SRQ1 [11] features are extracted for training and validation images and the estimator is trained. When the estimator has been got, for the test image, the SRQ1 features are extracted firstly and then the relative payload α is estimated by the estimator.

5.2 Parameters recognition and message extraction

When the stego object $\mathbf{y}$ is got, the parameters recognition of STCs can be realized by searching the parameters table and then the current parameters are judged by randomness test of decode sequence. According to the relation between the width w and relative payload α, when the parameter α is fixed, the width w is also fixed at the same time. So,

[1] BossBase-1.01[EB/OL]. http://exile.felk.cvut.cz/boss/BOSSFinal/.2013

the relative payload α should be estimated firstly, then the submatrices with width w are searched in ascending order of height h. In the following, one example about the parameters recognition and message extraction of HUGO is given, where the stego image is shown in Fig. 5c.

5.2.1 Estimating the relative payload

For the stego image shown in Fig. 5c, the SRQ1 feature is extracted firstly, then the relative payload α is estimated by the estimator obtained in section 5.1. The estimated value of α is 0.4302bpp, and the real relative payload α is 0.5. So the parameter α can be estimated approximately by quantitative steganalysis.

After the relative payload α is estimated, the range of width w of submatrix $\hat{\mathbf{H}}$ can be fixed. For the above stego image, because the estimated value of parameter α is 0.4302bpp, the width w of submatrix $\hat{\mathbf{H}}$ would be 2 or 3. Therefore, for parameters recognition of STCs used in this stego image, we only should consider the submatrix with width w is 2 or 3.

5.2.2 Searching the submatrix and extracting the embedded message

For the parameter recognition of submatrix $\hat{\mathbf{H}}$, we should traverse all the possible submatrices, and get the decode sequence via STCs. Then the current submatrix $\hat{\mathbf{H}}$ is judged according to the result of the randomness test of decode sequence. In the following, the parameter recognition of submatrix is discussed according to different value of submatrix.

(1) The current submatrix $\hat{\mathbf{H}}$ with the right height h and the width w, but some wrong matrix elements

Suppose the test submatrices are $\hat{\mathbf{H}} = [77,\ 71]$, $\hat{\mathbf{H}} = [69,\ 71]$, $\hat{\mathbf{H}} = [69,\ 67]$ or $\hat{\mathbf{H}} = [69,\ 65]$, we can see the four submatrices all have h=7 and w=2 with the correct submatrix $\hat{H} = [109,\ 71]$, but these submatrices have one, two, three, four wrong elements respectively. In these cases, the results of randomness test for decode sequences obtained according to stego object $\mathbf{y}$ and four difference submatrices are shown in Table 2, where P is the metric of randomness obtained by run test of binary sequence randomness. If P is larger than the significance level (0.05 in this paper), this means the randomness of corresponding decode sequence is relative strong, so this decode sequence is impossible the actual embedded message.

(a) Cover image **(b)** Embedded message **(c)** Stego image

Fig. 5 Cover image, embedded message image and the stego image

Table 2 Results of Randomness test of decode sequences when submatrix has right height and width, but wrong elements

Numbers of error elements	*P* value of run test	Result of randomness test
1	0.5034	Random sequence
2	0.1706	Random sequence
3	0.5782	Random sequence
4	0.1599	Random sequence

From the experimental results shown in Table 2, when the height h and width w of submatrix are right, but some matrix elements are wrong, it can be seen that the values of P are larger than the significance level 0.05, this means the randomness of all the decode sequences are strong, so the test submatrix cannot be the correct submatrix corresponding to the stego image.

Furthermore, when the height h and the width w of submatrix are both right, the size of extracted message is the same as the size of embedded message, so we can extract an image whose size is same as the actual embedded image. For four different submatrices, the extracted images are shown in Fig. 6a–d. It can be found that the extracted messages are random when the current submatrix has wrong elements. So, a conclusion can be drawn that even if the height and width of the submatrix are correct, but there are some errors in the matrix elements, the steganalyzer still cannot extract the correct message.

(2) The current submatrix $\hat{\mathbf{H}}$ with the right width w, but the wrong height h

Suppose that the test submatrices are $\hat{\mathbf{H}} = [149,\ 247]$, $\hat{\mathbf{H}} = [339, 473]$, $\hat{\mathbf{H}} = [997, 627]$ or $\hat{\mathbf{H}} = [1475, 1597]$, respectively, we can see the four submatrices all have the same width w=2 with the correct submatrix $\hat{H} = [109,\ 71]$, but the height of these submatrices are 8, 9, 10, 11 respectively. In these cases, the results of randomness test for decode sequences obtained from stego object $\mathbf{y}$ and four difference submatrices are shown in Table 3.

From the experimental results in Table 3, it can be seen that the decode sequences are all random when the width w is right but the height h is wrong. Moreover, when the width w of submatrix is right, the size of extracted message is also same as the size of embedded message, so we can extract an image whose size is same as the actual embedded image. For four different submatrices, the extracted four images are shown in Fig. 7a–d. It can be found that the extracted message are random, this shows that if the width of the submatrix is right, but the height of the submatrix is wrong, the steganalyzer still cannot extract the correct message.

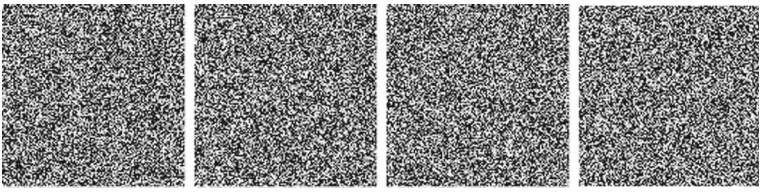


Fig. 6 Extraction results of embedded image when height and width of submatrix is right, but elements is wrong

Table 3 Results of randomness test for decode sequences when submatrix has right width, but wrong height

Height h of submatrix	P value of run test	Result of randomness test
8	0.4381	Random sequence
9	0.8954	Random sequence
10	0.4986	Random sequence
11	0.6288	Random sequence

(3) The current submatrix $\hat{\mathbf{H}}$ with the right height h, but the wrong width w

Suppose that the test submatrices are $\hat{\mathbf{H}} = [95,\ 101,\ 121]$, $\hat{\mathbf{H}} = [81,\ 107,\ 121, 95]$, $\hat{\mathbf{H}} = [95,\ 117,\ 105, 97, 75]$ or $\hat{\mathbf{H}} = [95,\ 83,\ 123, 109, 73, 103]$, respectively. The four submatrices all have the same height h=7 with the correct submatrix $\hat{\mathbf{H}} = [109,\ 71]$, but the widths of these submatrices are 3, 4, 5, and 6 respectively. In these cases, the results of randomness test for decode sequences obtained from stego object $\mathbf{y}$ and four difference submatrices are shown in Table 4.

From the experimental results in Table 4, the conclusion can be drawn that the decode sequence is random when the width w is wrong. In addition, when the width w of the current submatrix $\hat{\mathbf{H}}$ is wrong, it means that the current parameter α is not right. So the size of extracted message is not same as the size of embedded message, and an image (message) whose size is same to the actual embedded image cannot be extracted.

(4) The current submatrix $\hat{\mathbf{H}}$ is same to the correct submatrix

Suppose the test submatrix is $\hat{H} = [109,\ 71]$, that is the test submatrix is same to the correct submatrix. Then the decode sequence is obtained by STCs and the P value of run test for decode sequence is 0, so the decode sequence is not random when parameter $\hat{\mathbf{H}}$ is right.

When submatrix $\hat{\mathbf{H}}$ is right, the embedded image can be extracted correctly (See Fig. 8). Furthermore, if the submatrix is right, it means the current parameter α is also right. Therefore, only the parameters α and $\hat{\mathbf{H}}$ are both right, the embedded message can be extracted correctly.

From the above, it can be seen that the decode sequence will be a random sequence if the current parameter α and $\hat{\mathbf{H}}$ is wrong. Only when the two coding parameters of STCs are both right, the randomness of the decode sequence is weak, and the steganalyzer can recognize the correct coding parameters and can extract the embedded message correctly.

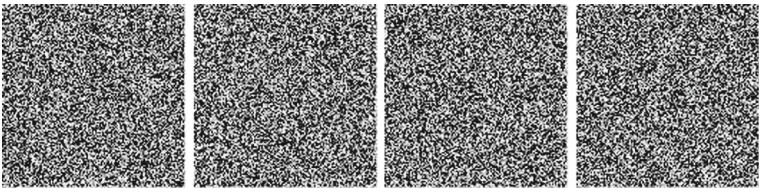


Fig. 7 Extraction results of embedded image when submatrix has right width, but wrong height

Table 4 Results of randomness test for decode sequences when submatrix has the right height, but wrong width

Width w of submatrix	P value of run test	Result of randomness test
3	0.1630	Random sequence
4	0.3829	Random sequence
5	0.3128	Random sequence
6	0.0643	Random sequence

5.3 Time consumption

For the parameters recognition of STCs, the time consumption of submatrix search varies according to the height h and width w of submatrix $\hat{\mathbf{H}}$. In the following, the time consumption of submatrix search is given for different relative payload α. Here, the height h of submatrix is 7, the submatrix has ones in first and last rows, and it has not identical columns.

In experiments, for special relative payload α, all the possible submatrices are tested until the correct submatrix is found, then the time consumption is recorded. The concrete time consumption and search range for different relative payload are given in Table 5.

From the experimental results in Table 5, it can be seen that the time consumption of submatrix search is only 8.6111×10^{-6} h when relative payload is 0.5bpp, because the search range is small in this case. But when relative payload is 0.3bpp, the time consumption is 191.1808 h. It is because that the parity-check matrix is generated by submatrix with width 3 and submatrix with width 4, and the search range of submatrix is relatively large and time consumption is also more much.

In addition, if the height h of submatrix used in HUGO is larger than seven, the search range of submatrix will become larger, and the time consumption of parameters recognition will be immense.

Fig. 8 Extraction result of embedded image when the coding parameters of STCs are right

Table 5 Time consumption of submatrix search for different relative payload when $h=7$

Relative payload	Search range of submatrix	Time consuming (unit: hours)
0.20 bpp	$\sum_{h=6}^{7}(2^{(h-2)\cdot 5}-2^{h-2}\cdot 25)$	0.1867
0.30 bpp	$\sum_{h=6}^{7}(2^{(h-2)\cdot 3}-2^{h-2}\cdot 4)\times(2^{(h-2)\cdot 4}-2^{h-2}\cdot 11)$	191.1808
0.40 bpp	$\sum_{h=6}^{7}(2^{(h-2)\cdot 2}-2^{h-2})\times(2^{(h-2)\cdot 3}-2^{h-2}\cdot 4)$	0.1809
0.50 bpp	$\sum_{h=6}^{7}(2^{(h-2)\cdot 2}-2^{h-2})$	8.6111×10^{-6}

6 Conclusions

In this paper, the steganalysis method for the classic HUGO steganography method is studied. On the basis of the principle analysis of HUGO steganography, the potential security risk of HUGO is indicated, and a steganalysis method for HUGO steganography is proposed based on blind parameters recognition of STCs. Experimental results show that the proposed analysis method not only can detect the stego-image reliably, but also can extract the embedded message correctly. This verifies the reliability of the proposed steganalysis method, and indicates the existing of security flaws of HUGO steganography at least, which may be subject to the third-party attack.

In fact, almost all steganographic methods relied on STCs code have the same potential security risk, the analyst can design a similar steganalysis methods to access the embedded message. In this paper, under the condition of embedding plaintext, the proposed method has successfully extracted the embedding message. However, if the message is embedded in the ciphertext, the steganalyzer need to adopt a new test method to determine whether the decoding result is correct or not. Next, we will study the steganalysis method of HUGO when the message is encrypted. Maybe there are other methods which can achieve the parameters of STCs more effectively. Obviously, this is a more challenging and exciting problem.

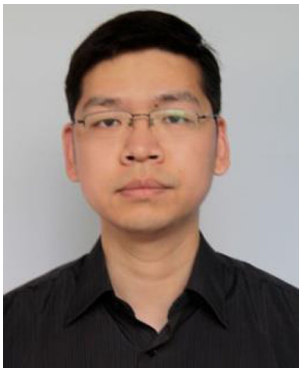
Acknowledgments This work was supported by the National Natural Science Foundation of China (No. 61379151, 61272489, 61302159, 61401512 and 61373020), the Excellent Youth Foundation of Henan Province of China (No. 144100510001), and the Foundation of Science and Technology on Information Assurance Laboratory (No. KJ-14-108).

References

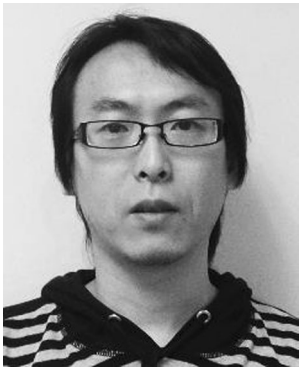
1. Fridrich J, Kodovsky J, Goljan M and Holub V (2011) Steganalysis of content-adaptive steganography in spatial domain. Proceedings of the 13th International Workshop on Information Hiding, Prague, 102–117
2. Guo LJ, Ni JQ, and Shi YQ (2012) An efficient JPEG steganographic scheme using uniform embedding. Proceedings of the 4th IEEE International Workshop on Information Forensics and Security, Tenerife, 169–174
3. Gul G, Kurugollu F. (2011) A new methodology in steganalysis: Breaking highly undetectable steganography. Proceedings of the 13th International Workshop on Information Hiding, Prague, 71–84
4. Holub V, Fridrich J (2012) Designing steganographic distortion using directional filters. Proceedings of the 4th IEEE International Workshop on Information Forensics and Security, Tenerife, 69–76
5. Holub V, Fridrich J (2013) Digital image steganography using universal distortion. Proceedings of the 1st ACM workshop on information hiding and multimedia security. Montpellier, 59–68

6. Holub V, Fridrich J (2013) Random projections of residuals as an alternative to co-occurrences in steganalysis. In: Proceedings of SPIE, electronic imaging, media watermarking, security, and forensics, California, 8665:1–11
7. Huang FJ, Huang JW, Shi YQ (2012) New channel selection rule for JPEG steganography. IEEE Trans Inf Forensc Security 7(4):1181–1191
8. Kodovsky J, Fridrich J (2009) Calibration revisited. In: Proceedings of the 11th ACMWorkshop on Multimedia and Security, Montpellier, 63–74
9. Kodovsky J, Fridrich J, Holub V (2011) On dangers of overtraining steganography to incomplete cover model. Proceedings of the 13th ACM workshop on multimedia and security, Niagara Falls, 69–76
10. Kodovsky J, Fridrich J, Holub V (2012) Ensemble classifiers for steganalysis of digital media. IEEE Trans Inf Forensc Security 7(2):432–444
11. Kodovsky J, Fridrich J (2012) Rich models for steganalysis of digital images. IEEE Trans Inf Forensc Security 7(3):868–882
12. Kodovsky J, Fridrich J (2013) Quantitative steganalysis using rich models. Proceedings of SPIE, Electronic Imaging, Security and Forensics of Multimedia XV, California, 1–11
13. Luo WQ, Huang FJ, Huang JW (2010) Edge adaptive image steganography based on LSB matching revisited. IEEE Trans Inf Forensc Security 5(2):182–193
14. Luo XY, Wang DX, Wang P, Liu FL (2008) A review on blind detection for image steganography. Signal Process 88(9):2138–2157
15. Luo XY, Liu FL, Lian SG, Yang CF (2011) On the Typical Statistic Features for Image Blind Steganalysis. IEEE J Selected AreasCommun 29(7):1404–1422
16. Li FY, Zhang XP, Yu J, Shen WF (2014) Adaptive JPEG steganography with new distortion function. Ann Telecommun 69(7–8):431–440
17. Provos N (2011) Defending against statistical steganalysis. In: Proceedings of the 10th conference on USENIX Security Symposium. Washington, 323–335
18. Rukhin A, Soto J, Nechvatal J (2008) A statistical test suite for random and pseudorandom number generators for cryptographic applications. NIST Special Publication 800–22
19. Shi Y Q, Sutthiwan P, Chen LC (2011) Textural features for steganalysis. Proceedings of the 14th International Workshop on Information Hiding. Berkeley, 63–77
20. Tomas P, Bas P, Fridrich J (2010) Steganalysis by subtractive pixel adjacency matrix. IEEE Trans Inf Forensc Security 5(2):215–224
21. Tomas P, Tomas F, Bas P (2010) Using high-dimensional image models to perform highly undetectable steganography. Proceedings of the 12th International Workshop on Information Hiding, Calgary, 161–177
22. Tomas F, Judas J, Fridrich J (2011) Minimizing additive distortion in steganography using syndrome-trellis codes. IEEE Trans Inf Forensc Security 6(3):920–935
23. Tomass F, Fridrich J (2011) Design of adaptive steganographic schemes for digital images. Proceedings of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIII, San Francisco, 2011, 1–14
24. Wang Y, Moulin P (2007) Optimized feature extraction for learning-based image steganalysis. IEEE Trans Inf Forensc Security 2(1):31–45
25. Westfeld A (2001) F5-a steganographic algorithm: high capacity despite better steganalysis. Proceedings of the 4th International Workshop on Information Hiding, Pittsburgh, 289–302

Xiangyang Luo received the M.S. degree and the Ph.D. degree from Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2004 and 2010, respectively. From 2006 to 2007, he was a visiting scholar of the Department of Computer Science and Technology of Tsinghua University. He is the author or co-author of more than 70 refereed international journal and conference papers. He is currently an associate professor of State Key Laboratory of Mathematical Engineering and Advanced Computing, and Zhengzhou Science and Technology Institute. His research interest includes multimedia security, image steganography/ steganalysis.



Xiaofeng Song received the B.S. degree from the School of Information and Technology, Zhengzhou University, Zhengzhou, China, in 2002, and M.S. degree from the School of Computer Science, Xidian University, Xi'an, China, in 2009. Now, he is a PhD Candidates of the State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou Science and Technology Institute, China. His current research interest includes steganography, steganalysis and digital image forensic.



Xiaolong Li received the B.S. degree from Peking University, Beijing, China, the M.S. degree from Ecole Polytechnique, Palaiseau, France, and the Ph.D. degree in mathematics from ENS de Cachan, Cachan, France, in 1999, 2002, and 2006, respectively. Before joining Peking University, he worked as a postdoctoral fellow at Peking University in 2007–2009. His research interests are image processing and information hiding.

Weiming Zhang received the M.S. degree and Ph.D. degree in 2002 and 2005, respectively, from the Zhengzhou Information Science and Technology Institute, Zhengzhou, China. Currently, he is an Associate Professor with the School of Information Science and Technology, University of Science and Technology of China, Hefei, China. His research interests include multimedia security, information hiding and cryptography.

Jicang Lu received the M.S. degree and Ph.D. degree from Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2010 and 2014, respectively. Currently, he is a lecturer of Zhengzhou Science and Technology Institute. His research interest includes image steganography and steganalysis technique.

Chunfang Yang received the M.S. degree and Ph.D. degree from Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2008 and 2012, respectively. Currently, he is a lecturer of State Key Laboratory of Mathematical Engineering and Advanced Computing, and Science and Technology on Information Assurance Laboratory. His research interest includes image steganography and steganalysis technique.

Fenlin Liu received the B.S. degree from Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 1986, the M.S. degree from Harbin Institute of Technology, Harbin, Heilongjiang, in 1992, and the Ph.D. degree from the Northeast University, Shenyang, Liaoning, in 1998. He is currently a professor of State Key Laboratory of Mathematical Engineering and Advanced Computing, and Zhengzhou Science and Technology Institute. His research interests include information hiding and security theory.