MATEC Web of Conferences **22**, 01012 (2015)
DOI: 10.1051/matecconf/20152201012
© Owned by the authors, published by EDP Sciences, 2015

A Study on the Spread Spectrum Steganography Based on the High-order Markov Model

Kaicheng Wu
School of Computer Science, Wuhan University, Wuhan, Hubei, China
School of Mathematics and Computer Science, Jianghan University, Wuhan, Hubei, China

ABSTRACT: Spread spectrum communication is a way of information transmission. Steganography based on the principle of spread spectrum has a strong robustness and security. This paper carries out a study on the application of the high-order Markov model of digital grayscale image in its spread spectrum steganalysis. Results suggest that the statistical estimate of the original image and the stego-image obtained by the digital image high-order Markov model is more sensitive than that of the traditional ε-secure security index and the image Markov model. Based on the fact that the high-order Markov model of carrier images has a preferable effect on SSIS steganalysis, the accuracy of steganalysis increases and False Positive decreases with the increase of the model order. The SSIS steganalysis scheme of the model is applicable not only to practical calculations but also to users of steganalysis with different requirements on analysis results.

Keywords: high-order Markov model; digital image steganography; spread spectrum steganographic algorithm; SSIS steganalysis

Retracted

1 INTRODUCTION

The age of computer came in the 1990s. We are now in the phase of further development of the computer age. With the rapid development of the communication network, transmission and communication of all kinds of information become much easier. A great challenge is thusly brought to the protection of information security in the computer field. A large number of new theories and technologies of information security emerge to face this challenge. The development of various information hiding and information security technologies, such as the digital watermark technology, the digital steganography and digital steganalysis, is the most prominent. This paper studies the spread spectrum steganalysis of digital images, aiming at exploring a scientific steganalysis algorithm and providing guidance for the improvement of the security of computer information and communication.

The communication model of the steganographic system is shown in Figure 1. The content in Figure 1 is that the sponsor of the covert communication, Alice, embeds the secret information $w \in W$ into the carrier information $c \in C$ through the map E under the control of the secret key k_e so as to obtain the hiding information $s \in S$. The hiding information s is transmitted to the receiver of the covert communication Bob through the channel monitored by the steganalysis. According to the received s and the secret key k_d, Bob extracts the secret information W with the map D. In the process of the covert communication, s is monitored by the steganalysis Eve. E carries out an analysis on the signal s transmitted on the channel with the steganalysis algorithm A so as to determine whether there are covert channels in s. If the secret information is found by Eve, it will terminate or interfere the communication and track both sides of the communication. Some advanced analysis technologies can be applied in obtaining steganographic keys and extracting embedded secret information w.

The core content of the communication model of the steganographic system is the steganalysis algorithm. Yifeng Sun et al. (2010) solved the divergence of spread spectrum steganographic system with the carrier of Gauss-Markov modeling [1]. Ling Xi et al. (2012) proposed an analytical method of Gaussian mixture model based on natural images [2]. Chunjuan Ouyang et al. (2012) conducted a research on various changes and indeterminacies of the statistical characteristic caused by image steganography and introduced the similarity measure of Vague set into the safety evaluation of the steganographic system [3]. Changyong Xu et al. (2010) studied the video steganography based on the spatial-temporal correlation [4]. Lei Xu et al. (2010) pointed out the steganographic method for data with multiple transform domains of the direct spread spectrum audio steganographic analysis method based on time-frequency domain features, the average detection accuracy of which is higher than 90% [5]. Ke Qi et al. (2013) proposed the general steganalysis algorithm of color images based on noise model and channel blending [6]. Hongzhi Zhou et al. (2014) put forward a video steganalysis method for the fine identification of spatial-temporal features [7]. Weidong Zhong et al. (2012) proposed a real-time video steganalysis method according to the patial-temporal redundancy feature of videos [8]. But certain difficulties in the practical application of these

This is an Open Access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article available at http://www.matec-conferences.org or http://dx.doi.org/10.1051/matecconf/20152201012

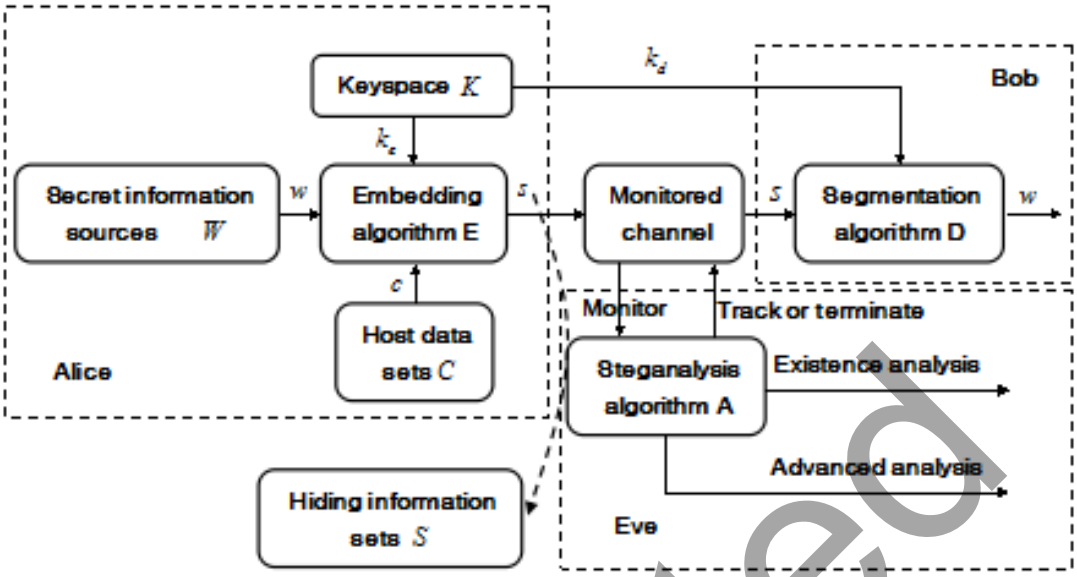

Figure 1. The communication model of the steganographic system

algorithms exist. Further studies are still needed. The spread spectrum steganalysis of the Markov chain model can avoid the difficulty of the actual operation to some extent. On this basis, this paper explores the application of the high-order Markov model in the spread spectrum steganalysis of digital images, aiming at verifying the scientific nature of the algorithm through theoretical analyses and experimental results.

2 THE CONSTRUCTION OF THE HIGH-ORDER MARKOV CHAIN MODEL OF DIGITAL IMAGES

The Markov process is an important tool for describing dynamic random phenomena, which has gained a great many achievements in the field of random data analysis [9]. The current steganalysis carries out attacks mainly from the angle of existing security. The theory of existing security is the statistical security theory of secret information, which means that the correctness of an analyst's judgment on whether the secret information is included is not higher than that of a random guess. Therefore, the study on the security of steganographic statistics requires a statistic model that is more in accord with the carrier reality so that the security performance can be reflected more accurately. Besides, this model can guide the design of the steganalysis algorithm more conveniently. The random model is constructed with the high-order Markov chain model theory in this paper.

2.1 *The high-order Markov chain model theory*

Definition: suppose $\{X(t_n), t \in T\}$ is a random process, $t_i \in T, i=1,\mathrm{L}\ ,n$ and $t_1 < t_2 < \Lambda < t_n$. As for the condition that the conditional distribution function of a random state $x_1,\mathrm{L}\ ,x_{n-1},x_n \in R, X(t_n)$ in the state space S satisfies the Formula (1), $\{X(t_n), t \in T\}$ is called the first-order Markov process.

$$P\left\{X(t_n) < x_n \middle| X(t_{n-1}) = x_{n-1},\mathrm{L}\ ,X(t_1) = x_1\right\} = P\left\{X(t_n) < x_n \middle| X(t_{n-1}) = x_{n-1}\right\} \tag{1}$$

$$P\{X(t) = x_t | X(t-1) = x_{t-1},\Lambda\ ,X(t-n) = x_{t-n},\Lambda\ ,X(t_1) = x_1\} = P\{X(t) = x_t | X(t-1) = x_{t-1},\Lambda\ ,X(t-n) = x_{t-n}\} \tag{2}$$

The core idea of the first-order Markov process is to correlate the future t_n only with the current t_{n-1}, which is unrelated to the past $t_1 \sim t_{n-2}$. This characteristic is called non-aftereffect property. The right part of Formula (1) is called the transition probability distribution of the Markov process. The core idea of the high-order Markov process is the change of variables that is not necessarily reflected in the next variable immediately. The future state is not only related to the current state but also influenced by one or more past states. The first-order non-aftereffect property of the Markov model is expanded to the high-order non-aftereffect property.

The high-order professional probability matrix model, proposed by Pegram (1975), transits the high-order state to the next state as a transitional element [10]. Suppose the spatial state of the Markov chain of the n-order discrete parameter is $S=\{1,2,\mathrm{L}\ ,m\}$, the high-order transition probability matrix Q is designed in the form shown in Formula (3).

Retracted

$$Q=\begin{array}{c} \\ 1\cdots11 \\ 1\cdots12 \\ \vdots \\ 1\cdots1m \\ 1\cdots21 \\ \vdots \\ \vdots \\ m\cdots mm \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & m \end{array} \\ \begin{pmatrix} p_{1\cdots111} & p_{1\cdots112} & \cdots & p_{1\cdots11m} \\ p_{1\cdots121} & p_{1\cdots122} & \cdots & p_{1\cdots12m} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1\cdots1m1} & p_{1\cdots1m2} & \cdots & p_{1\cdots1mm} \\ p_{1\cdots211} & p_{1\cdots212} & \cdots & p_{1\cdots21m} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ p_{m\cdots mm1} & p_{m\cdots mm2} & \cdots & p_{m\cdots mmm} \end{pmatrix} \end{array} \tag{3}$$

Formula (3) is neither a square matrix nor a random matrix. The estimated number m^{n+1} of a great deal of parameters in Q restricts the development of the model. For this reason, according to the high-order feature that time dynamic variables are only related to the previous continuous n states, Raftery (1985) presumed that transition probabilities of n orders have the relation of linear self-regression and established a high-order Markov model with simplified parameters [11]. Suppose a random process $\{X(t),t\in N\}$ is a Markov chain with discrete parameters and the state space is $S=\{1,2,\mathrm{L},m\}$, $\{X(t),t\in N\}$ can be regarded as a MTD high-order Markov chain model if a random state $x_{t-n},\Lambda,x_{t-1},x_t$ in the state space S satisfies the relation shown in Formula (4).

$$P\{X(t)=x_t|X(t-1)=x_{t-1},\Lambda,X(t-n)=x_{t-n}\}=\sum_{i=1}^{n}\lambda_i q_{x_{t-i}x_t} \tag{4}$$

In Formula (4), $\sum_{i=1}^{n}\lambda_i=1$ and $Q=(q_{ij})_{m\times m}$ is a random matrix. The sum of each row of $Q=(q_{ij})_{m\times m}$ is 1, so $0\le\sum_{i=1}^{n}\lambda_i q_{x_{t-i}x_i}\le 1$.

2.2 *Security indexes of the steganography Markov chain of digital images*

At present, the widely used steganographic security indexes are ε-secure indexes established by Cachin [12] in line with the security theory of statistics and the K-L (Kullback – Leibler) distance. The theory models the attack steganalysis as a hypothesis testing problem of the theory of statistics. The hypothesis testing method is used to determine whether the information intercepted by the monitored information channel comes from the original carrier information collection or the hiding information collection. Suppose the pixel value of the carrier image x represents independent identically distributed random variables $(\mathrm{i.i.d.})$ and G is the collection of all possible values of x (If it is a grayscale image of 8 bits, values range between 0-255), $x\in G$. Suppose C and S are the original carrier image and the secret-carrying image after steganographic embedding, $P_C(x)$ and $P_S(x)$ are probabilities of the pixel value x in C and S. The statistical distribution K-L distance of the carrier image and the secret-carrying image is shown in Formula (5):

$$D(P_C\|P_S)=\sum_{x\in G}\left\{P_C(x)\log\left[\frac{P_C(x)}{P_S(x)}\right]\right\} \tag{5}$$

The steganographic system is absolutely safe when $D=0$ and it is ε safe when $D\le\varepsilon$. The ε-secure security indexes are based on the hypothesis of the carrier i.i.d. model and changes of the first-order statistical distribution of the carrier are also compared. However, there is generally a strong correlation between carrier signals no matter the steganographic carrier is a digital image or a digital audio/video. Besides, many steganalysis technologies are now inclined to conduct analyses with the high-order statistical distribution features of carrier signals [13]. Thus, if ε-secure indexes are directly used in the security test of the steganographic system, the effect of steganalysis schemes conducted in line with the carrier signal correlation and various high-order statistical features is frequently underestimated, which also means that the security of the tested steganographic system is overestimated.

According to the defects of ε-secure security indexes, many researchers have put forward various kinds of statistic models that are in accord with the reality of steganographic carrier signals for the improvement. As for digital images, it is more in line with the practical situation of the carrier if the statistical distribution of natural image pixels is modeled as a Markov Random Field (MRF). But the calculation amount of the MRF model is tremendous, so this modeling is not convenient for the steganographic security analysis. Besides, the MRF model is not convenient for guiding the design of steganalysis and high-security steganographic algorithm as well with regard to the calculation ability of steganographic analysts. On this basis, Sullivan et al. [15] simplified the MRF model and proposed that the safety testing indexes of the Markov Chain (MC) model can be adopted. Suppose i and j are image pixels, G is the collection of all possible values of i and j, and $i,j\in G$. Firstly, pixels of the carrier image are arranged in a data chain X. Suppose x_t belongs to X, t stands for the position of x_t in X and $P(x_t)$ is the probability of x_t. According to the first-order Markov model of image steganography proposed by Sullivan et al., the current value of pixel x_t in the data chain X is only related to the former pixel value x_{t-1} in the chain. The empirical matrix of the first-order Markov model of the carrier image can be thusly obtained by recording all pixel values i, namely the occurrence number j. Suppose C and S are respectively the original carrier image and the se-

Retracted

cret-carrying image, M^C and M^S are respectively defined as the first-order Markov empirical matrix of the original carrier image and the secret-carrying image, m^C and m^S are elements in M^C and M^S. The distance measure of the statistical distribution of the original image C and the secret-carrying image S can be obtained in accordance with the divergence distance formula shown in Formula (6).

$$D(M^C, M^S) = \sum_{i,j \in G} \left[m_{ij}^C \log \left(\frac{m_{ij}^C}{\sum_j m_{ij}^C} \cdot \frac{\sum_j m_{ij}^S}{m_{ij}^S} \right) \right] \quad (6)$$

As for a steganographic system, $D(M^C, M^S)$ provides the inherent difference measure of statistical distribution features between the original carrier image and the secret-carrying image. The statistical measure of the first-order Markov model of the digital image steganography $D(M^C, M^S)$ contains certain correlation information of the carrier and pays attention to the second-order statistical distribution features of the carrier. Therefore, the assessment of this statistical measure on the statistical security of the steganographic algorithm is more accurate than traditional ε-secure security indexes. But each pixel of a natural image normally has a strong correlation with at least 8 neighboring pixels, so a great deal of relevant information of carrier will be lost if the statistical measure of the first-order Markov model is applied to the security assessment of the steganographic system. This paper introduces the high-order Markov model so as to provide a more accurate model for the study on spread spectrum steganography of digital images.

2.3 *The construction of the high-order Markov chain model of digital images*

With the grayscale image *A* as an example, this paper illustrates the construction method of the high-order Markov model of digital images and the empirical matrix. First, the n-order Markov chain $X = \{x_1, x_2, \mathrm{L}, x_t, \mathrm{L}, x_{L-1}, x_L\}$ can be obtained by scanning the spatial domian pixels of *A* in a certain way. *L* is the length of the data chain *X*. According to the definition of the high-order Markov model, the element x_t in the data chain *X* satisfies $P(x_t | x_{t-1}, x_{t-2}, \mathrm{L}, x_1) = P(x_t | x_{t-1}, x_{t-2}, \mathrm{L}, x_{t-n})$. It means that the value of any element x_t in *X* is only related to $x_{t-1} \sim x_{t-n}$. Suppose *G* is the collection of all possible values of x_t (If it is a grayscale image of 8 bits, values range between 0-255), $x_t \in G$.

Definition: $\eta_{i_1, i_2, \mathrm{L}, i_n, i_{n+1}, (X)}$ is the occurrence number of state transition that all data of n-order Markov chain *X* passes from i_1 to i_2, i_3 and finally reaches i_{n+1}, which is also the occurrence number of $x_t = i_{n+1}, x_{t-1} = i_n, \mathrm{L}, x_{t-(n+1)} = i_2, x_{t-n} = i_1$ in *X*.

$$M^X \triangleq \left\{ m_{i_1, i_2, \Lambda, i_k, \Lambda, i_{n+1}} = \frac{\eta_{i_1, i_2, \Lambda, i_k, \Lambda, i_{n+1}}(X)}{L-n}, i_k \in G \right\} \quad (7)$$

In Formula (7), $\frac{\eta_{i_1, i_2, \mathrm{L}, i_k, \mathrm{L}, i_{n+1}}(X)}{L-n}$ indicates the proportion of the pixel value changes of the *X* chain grayscale pixel value passing from i_1 to i_2, i_3 and finally reaching i_{n+1} in the total pixel value change. It also provides the estimation of the joint distribution probability in *X*. As shown in Figure 2, the 2-order Markov model of a binary image is taken as an example to illustrate the construction method of the n-order Markov model and the empirical matrix of digital images.

In Figure 2, the method of column scanning is adopted in the binary image scanning. First, the data chain $X = 001011100$ can be obtained through the column scanning of the binary image. If the data chain *X* is set as a 2-order Markov chain, the 2-order Markov empirical matrix of *X* is $M_{2\times2\times2}^X$. The element of $M_{2\times2\times2}^X$ is $m_{i_1, i_2, i_3}^X = \frac{\eta_{i_1, i_2, i_3}}{L-n}$. *L* is the length of the chain *X* and *n* is the order of the Markov chain. *X* is a 2-order Markov with a chain length of 9, so $m_{i_1, i_2, i_3}^X = \frac{\eta_{i_1, i_2, i_3}}{7}$. The value of $M_{2\times2\times2}^X$ is 0 with elements of (0,0,0), 1/7 with elements of (0,0,1), 1/7 with elements of (0,1,0), 1/7 with elements of (0,1,1), 1/7 with elements of (1,0,0), 1/7 with elements of (1,0,1), 1/7 with elements of (1,1,0), and 1/7 with elements of (1,1,1). Element values in the empirical matrix are shown in Figure 2.

3 SPREAD SPECTRUM STEGANALYSIS BASED ON THE IMAGE HIGH-ORDER MARKOV MODEL

Spread spectrum embedding is an important embedding mode of steganography, which was first proposed as a digital watermarking embedding mode by Cox [16]. The embedding domain mainly involves the DCT coefficient domain of images. Marvel [17] then applied this embedding mode to the steganography of digital images, expanded the embedding domain to the carrier

Retracted

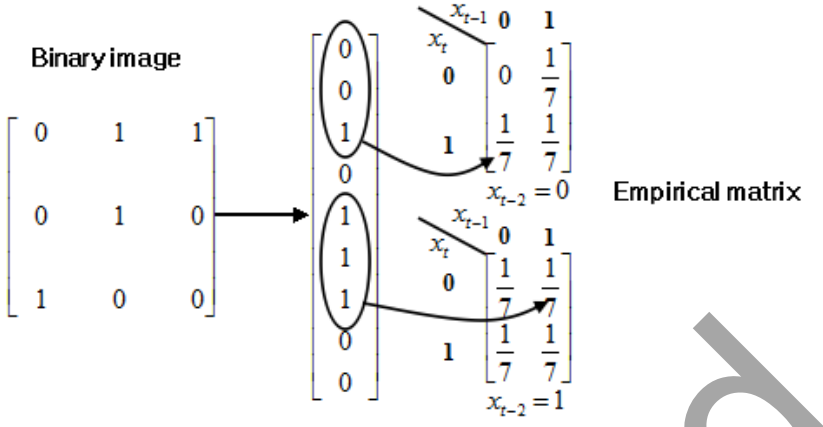

Figure 2. The construction diagram of the 2-order Markov model and the empirical matrix of digital images

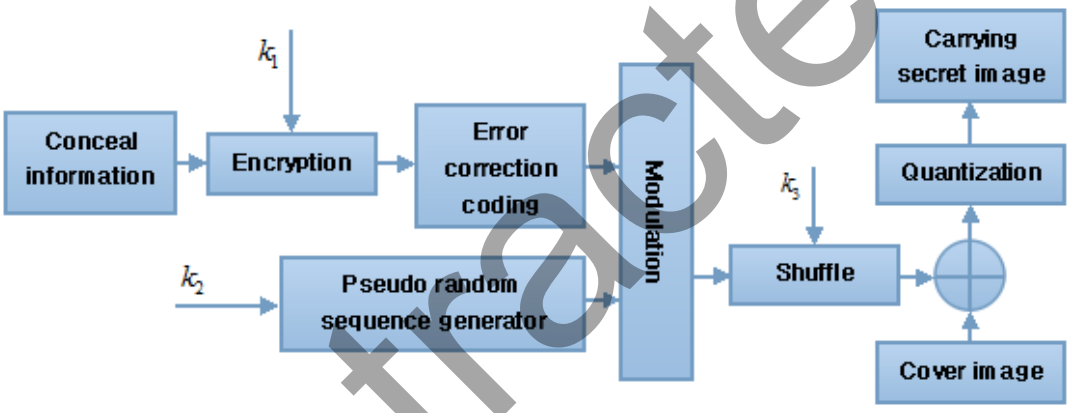

Figure 3. The embedding process of SSIS

spatial domain pixel, and put forward the Spread Spectrum Image Steganography (SSIS) algorithm. Fridrich [18] further improved SSIS soon afterwards and put forward the random modulation steganography algorithm with a larger actual embedding capacity. The SSIS algorithm is one of the important embeddin modes of steganography, so it is necessary to carry out studies on the steganalysis. The method of steganalysis of spread spectrum image steganography is studied in this section with help of the high-order Markov model of digital images.

3.1 *The spread spectrum steganography algorithm of digital images*

SSIS is a steganographic algorithm that embeds the zero-mean Gaussian white noise of unit variance after the modulation of secret information into the image spatial domain or the transform domain. The influence on carrier signals is equivalent to the random noise superposition in the embedding domain. The embedding process of SSIS is shown in Figure 3:

It can be known from Figure 3 that embedded secret information should be first encrypted with the secret key k_1 so as to obtain signals needing to be embedded through the error correction of coding. Meanwhile, the secret key k_2 is used to generate a pseudo-random zero-mean real-number sequence of unit variance. Signals need to be embedded after the error correction of coding and the pseudo-random real-number sequence are modulated together and superposed to the embedding domain of the carrier image after the shuffle of the secret key k_3 .If the embedding domain is a spatial domain pixel value of the image, the newly-superposed carrier image is quantized to obtain the secret-carrying image. If the embedding domain is a transform domain, the secret-carrying image can be obtained through the inverse transformation and quantization. The extraction process of SSIS is shown in Figure 4.

Retracted

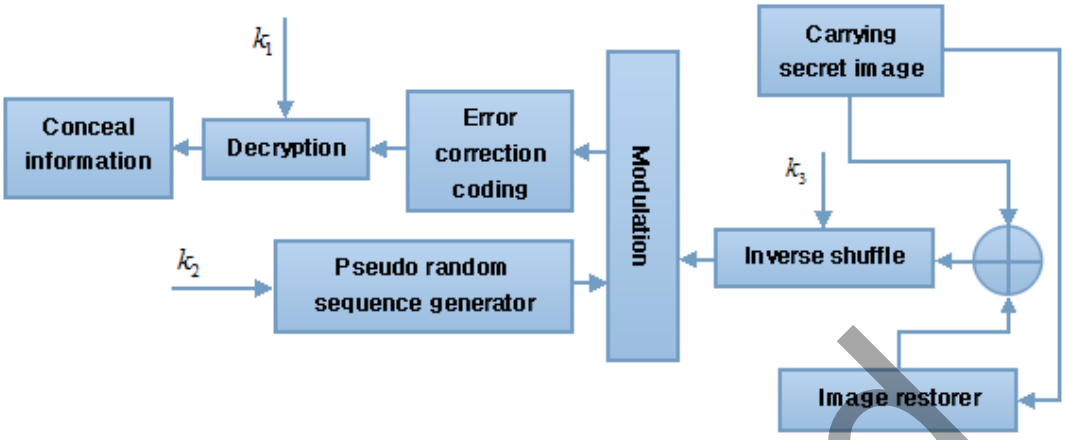

Figure 4. The extraction process of SSIS

In Figure 4, the secret-carrying image is first restored with the image recovery technique and the difference of the secret-carrying image and the recovered image is then solved and shuffled inversely with the secret key k_3. The difference is modulated by the pseudo-random real-number sequence generated by the secret key k_2 and is finally coded by error correction with the secret key k_1 so as to obtain the secret information. The image recovery technique used in the extraction cannot recover the original carrier image accurately, so the extracted secret information is inevitably different from the originally embedded secret information. The error correction coding mechanism is introduced in order to avoid communication errors. But the actual steganographic capacity is lost at the same time. SSIS loads the noise signal modulated with the secret information to the embedding domain of the carrier generally through additive and multiplicative embedding modes. SSIS frequently carries out embedding operations in the image spatial domain pixel or the sub-block DCT coefficient domain, so there are mainly four kinds of common SSIS, namely additive SSIS of spatial domain, multiplicative SSIS of spatial domain, additive SSIS of sub-block DCT domain and multiplicative SSIS of sub-block DCT domain.

3.2 *Spread spectrum steganalysis based on the high-order Markov model*

The statistical measure $D_n\left(M^X, M^S\right)$ of the high-order Markov model of the carrier image reflects the influence of steganography on the high-order statistical distribution of the carrier. The calculation is shown in Formula (8).

$$D_n\left(M^X, M^S\right) = \sum_{i_1 \in G}\left[p_{i_1}^X \log\left(\frac{p_{i_1}^X}{p_{i_1}^S}\right)\right] = D\left(P_X \| P_S\right) \quad (8)$$

In Formula (8), p_i stands for the probability of pixel values in the image. The calculation of this measure requires the original image without embedded secret information and the steganographic analyst usually does not have conditions of mastering the original carrier image. Therefore, $D_n\left(M^X, M^S\right)$ cannot be directly used in the steganalysis of the high-order Markov model. If the empirical matrix of the high-order Markov model of the carrier image is taken as features of the steganalysis, the dimensionality of the eigenvector reaches up to $\left(2^r\right)^{n+1}$ (r is the image bit, n is the model order). Thus, the calculation scale of the steganographic analyst is excessively large. Considering that the empirical matrix of the carrier image disperses around the leading diagonal after steganography, the phenomenon can be taken advantage of in the dimensionality reduction of the feature extraction of the empirical matrix of the secret-carrying image's high-order Markov model so as to meet the requirements of the computational complexity. The digital image discussed in this paper is a 2-order Markon model of an 8-bit grayscale image. The process of SSIS steganaysis algorithm based on the high-order Markov model is presented below:

Part1. Part steps of the classifier training

Step1. N carrier images of enough number are used as the classifier training set, N/2 images are randomly selected and embedded in the secret information through SSIS as the secret-carrying image set, and the rest is used as the original image set. The sample size adopted in this paper is N=1000.

Step2. As for each image of the training set, the 2-order Markov empirical matrix is constructed with the method shown in Figure 2. The 374-dimensional eigenvector of each image can be obtained through the method mentioned in Literature [19].

Step3. Eigenvectors of all images in the training set are used to carry out trainings on the SVM (Support

Retracted

Vector Machine) classifier so as to get the SVM classifier after trainings.

Part2. Part steps of the analysis and detection of the SVM classifier after trainings

Step1. Construct the 2-order Markov empirical matrix of the image to be analyzed with the method shown in Figure 2 and obtain the eigenvector of the image through the method mentioned in Literature [19].

Step2. Analyze the eigenvector of the image to be analyzed with the SVM classifier after trainings so as to determine whether the image belongs to SSIS steganographic secret-carrying images.

4 EXPERIMENTAL RESULTS ANALYSIS

Image source: 1338 uncompressed images from the UCID.V2 image library and scanned 662 uncompressed images are selected as the image library for the training and testing of the classification analysis.

Model selection: The 2-order Markon model of an 8-bit grayscale image is taken as the example for the experiment.

Experimental objective: Make a comparison with the SSIS steganalysis of the 1-order model in Literature [15] on the basis of introducing the image spatial domain pixel, the additive and multiplicative SSIS steganography of sub-block DCT coefficient and the experimental situations of four kinds of steganographic algorithms so as to highlight the superiority of the high-order Markov model.

Experimental process: (1) Convert all images in the training and testing image library to 8-bit grayscale BMP images. 1000 images are randomly selected as the training set and the rest is the testing set. 500 images are respectively selected from the training set and the testing set as the secret-carrying image set and the rest images are used as the original image set. (2) Apply the additive SSIS (embedded energy coefficient $\alpha = 0.375$) and the multiplicative SSIS (embedded energy coefficient $\alpha = 0.05$) to secret-carrying images in the training set and the testing set. The even-distributed secret information {0, 1} is embedded in the spatial domain pixel and the 8×8 sub-block DCT coefficient of the carrier image with the embedding rate of 0.91bpp. Thus, embedded secret-carrying images after the additive and the multiplicative SSIS steganography of the spatial domain and the DCT coefficient domain can be obtained respectively. (3) Scan images of the training set and the testing set respectively to construct a 2-order Markov model and a 1-order Markov model. The 2-order Markov model is scanned in the Hibert [21] way while the 1-order Markov model is scanned in the way mentioned in Literature [15]. (4) Extract the image features of the scanned empirical matrix of the 2-order Markov model in the way mentioned in Literature [19] and extract the image features of the empirical matrix of the 1-order Markov model in the way mentioned in Literature [15]. Trainings and tests of the SVM classifier are carried out by the extracted eigenvectors.

Mean results of the repeated 20 experiments are presented in Table 1 and Figure 5.

In Figure 5, 1-MC stands for the 1-order Markov model and 2-MC stands for the 2-order Markov model. It can be known from Table 1 and Figure 5 that, as for the additive SSIS image steganography of the spatial domain pixel and the transform domain coefficient, the analytical accuracy of the spread spectrum steganalysis based on the high-order Markov model improves with the increase of the model order. It can be seen from the experimental results of the 2-order and the 1-order model, and the analytical accuracy of the 2-order model increases by nearly 200 percent compared to that of the 1-order model. Besides, False Positive decreases obviously. As for the multiplicative SSIS image steganography of the spatial domain pixel and the transform domain coefficient, the analytical accuracy also improves greatly with the increase of the model order. This can be seen from the 2-order and the 1-order experimental results. Besides, False Positive also decreases significantly.

Table 1. Experimental results of the SSIS steganalysis

Embedding domain		DCT coefficient		Spatial domain pixel	
Embedding method		Additive SSIS	Multiplicative SSIS	Additive SSIS	Multiplicative SSIS
1-order Markov model	True Positive	97.63%	76.92%	97.04%	85.21%
	False Positive	8.88%	21.30%	5.92%	23.67%
	Accuracy	94.38%	77.81%	95.56%	80.77%
2-order Markov model	True Positive	94.08%	85.21%	96.45%	85.80%
	False Positive	0.59%	16.77%	1.18%	16.57%
	Accuracy	96.75%	84.32	97.63%	84.62%

Note: True Positive means that the ratio of secret-carrying carriers is determined correctly; False Positive means that the ratio of secret-carrying carriers is determined improperly, namely the false alarm probability; True Negative means that the ratio of non-secret-carrying carriers is determined correctly; Accuracy means that the value of accuracy equals to (True Positive + True Negative) / 2.

Retracted

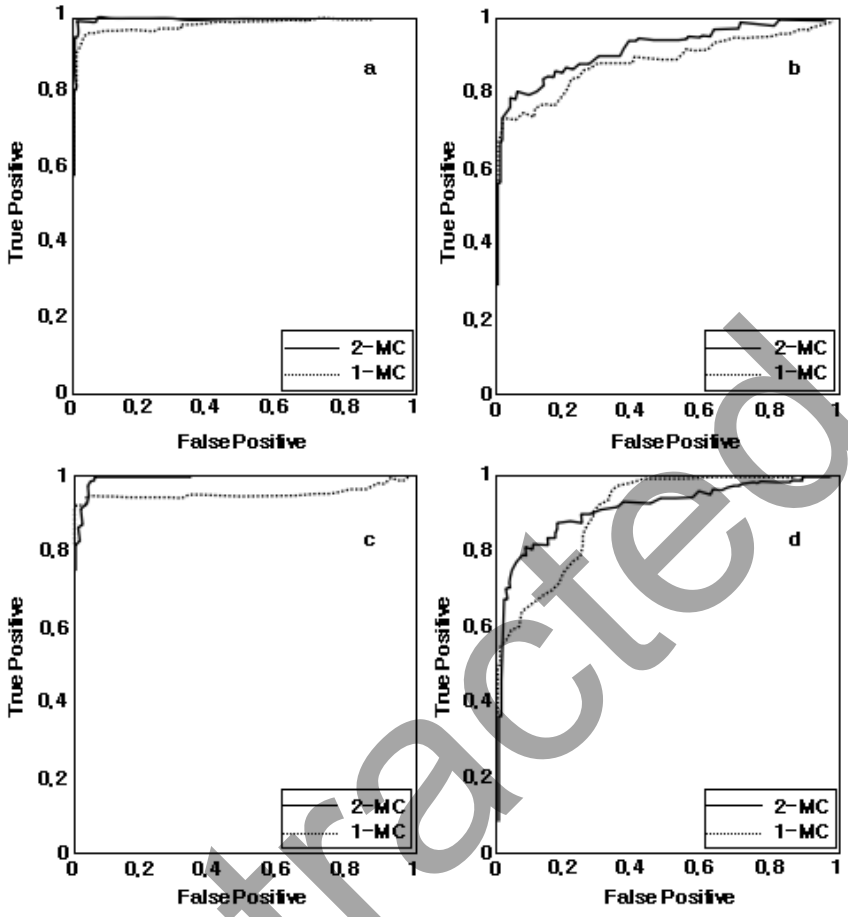

Figure 5. ROC curve of the SSIS steganalysis experiment of carrier images (spatial domain pixels, a-additive, b-multiplicative; 88 sub-block DCT coefficient, c-additive, d-multiplicative)

Retracted

5 CONCLUSION

This paper constructs the high-order Markov model of digital images based on the study of the high-order Markov chain model theory and the steganographic Markov chain security indexes of digital images. It mainly studies the spread spectrum steganalysis based on the high-order Markov model of images, carries out an experiment with the 2-order Markov model of 8-bit grayscale carrier images as the example, and analyzes ROC curves of the additive SSIS steganography and the multiplicative SSIS steganography of the DCT coefficient and the spatial domain pixel. Conclusions are drawn as follows:

1) The high-order Markov statistical distribution model of digital image steganography is proposed in this paper. The scanning mode of the Hilbert space filling curve can be used to construct the high-order Markov model and its empirical matrix.

2) The statistical measure of the high-order Markov model is more sensitive than that of the traditional ε-secure security indexes and the image Markov model. It can fully reflect changes of the statistical distribution of the carrier image high-order Markov model caused by steganographic embedding.

3) With the example of the 2-order Markov model, steps of the SSIS steganalysis algorithm based on the image high-order Markov model can be divided into two parts. One part is the training of the classifier and the other part is the analytical testing.

4) The high-order Markov model based on carrier images has a preferable effect on the SSIS steganalysis. In addition, the accuracy of steganalysis improves False Positive decreases while with the increase of model order within certain limits.

To sum up: the SSIS steganalysis scheme based on the high-order Markov model is applicable not only to practical calculations but also to users of steganalysis

with different requirements on analysis results.

REFERENCES

[1] Sun, Y.F. & Liu, F.L. 2010. A study on the anti-detection property of the spread spectrum steganography of Gauss-Markov carriers, *Data Acquisition and Processing*, 25(4): 462-468.

[2] Xi, L., Ping, X.J. & Zhang, H. 2012. Security analysis of the self-adaptive spread spectrum steganography based on the GMM model, *Computer Engineering*, 38(1): 137-139.

[3] Ouyang, C.J., Li, B., Li, X. & Wang, N. 2012. Security measure of the image steganographic system based on the similarity measure of Vague set, *Chinese Journal of Computers*, 35(7): 1510-1521.

[4] Xu, C.Y. & Ping, X.J. 2010. An analysis on the video steganography based on the spatial-temporal correlation, *Journal of Image and Graphics*, 15(9): 1331-1337.

[5] Xu, L., Guo, L., Wang, C.P., Wang, Y.J. & Yang, F. 2010. Direct spread spectrum audio steganalysis based on video features, *Communication Technologies*, 43(1): 78-81.

[6] Qi, K. & Xie, D.Q. 2013. Steganalysis of colored images based on noise model and channel integration, *Computer Research and Progress*, 50(2): 307-318.

[7] Zhou, H.Z. & Wang, D.M. 2014. Video steganalysis of fine identified spatial-temporal characteristics, *Computer Engineering*, 40(1): 149-152.

[8] Zhong, W.D., Wu, J.Q., Wu, G.R. & Yang, H.B. 2012. A video steganalysis method based on the invisibility of spatial-temporal redundancy statistics, *Application Research Of Computers*, 29(10): 3846-3850

[9] Rong, T.Z. 2012. Study on the Prediction Method Based on Higher Order Period Markov Chain Model, a thesis of the Doctor's degree. *Chongqing University*.

[10] Pegram G. G. S. 1975. A multinomial model for transition probability matrices, *Journal of Applied Probability*, 12(3): 498-506.

[11] Raftery A.E. 1985. A model for high-order Markov chains, *Journal of the Royal Statistical Society*, 47(3): 528-539.

[12] Cachin C. 2004. An information-theoretic model for steganography, Information and Computation, 192(1): 41-56.

[13] Wang, S.Z., Zhang, X.P. & Zhang, W.M. 2009. Research progress of the steganalysis with digital images as carriers, *Chinese Journal of Computers*, 32(7): 1247-1263

[14] Luo, X.Y., Wang, D.S. & Wang, P. 2008. A review on blind detection for image steganography, *Signal Processing*, 88(9): 2138-2157.

[15] Sullivan K, Madhow U. & Chandrasekaran S, et al. 2006. Steganalysis for Markov cover data with applications to images, *IEEE Transaction on Information Forensics and Security*, 1(2): 275-287.

[16] Cox I J, Kilian J, Leighton T, et al. 1997. Secure spread spectrum watermarking for multimedia, *IEEE Transaction on Image Processing*, 6(12): 1673-1687.

[17] Marvel L M, Boncelet C G. & Retter C T. 1999. Spread spectrum image steganography, *IEEE Transaction on Image Processing,* 8(8): 1075-1073.

[18] Fridrich J. & Goljan M. 2003. Digital image steganography using tochastic modulation, in Proceedings of SPIE: Security, Steganograhy, and Watemarking of Multimedia Contents V, SPIE: Santa Clara, Cuba, 121(2): 191-202.

[19] Zhang, Z. 2010. Research on High-order Statistical Secure Steganography Algorithm, a thesis of the Doctor's degree. *Nanjing: Nanjing University of Science and Technology*.

[20] Schaefer G. & Stich M. 2004. *UCID-An Uncompressed Colour Image Database*.

[21] Zhang, Z., Liu, G.J., Wang, J.W., Dai, Y.W. & Wang, Z.Q. 2010. Spread spectrum steganalysis based on the image high-order MARKOV chain model, *Chinese Journal of Electronics*, 38(11): 2578-2585.

Retracted