16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

A Blind Steganalysis Scheme using Estimation Techniques

Albert Ortiz [2], *Balakumar Ragunathan* [1], *Li Bai* [1], *Saroj Biswas* [1] *and Don Dalessandro* [2]

1 - Electrical and Computer Engineering Department, Temple University, Philadelphia PA 19122
2 - Department of the Navy, Naval Surface Warfare Center Carderock Division, Philadelphia PA

Abstract*— Distributed image steganography (DIS) [11] is a method for hiding a secret image inside n different innocuous cover images. The method is secure and reliable because it leaves very little traces of the secret image in the stego image. However, it is possible that criminals could use it for unchallenged, covert communication of illegal secret information over the Internet. In such cases, a suitable countermeasure (also known as steganalysis), to reverse the DIS is needed which can detect and disclose this kind of secret information. In this paper, the challenge of disclosing the secret image was handled by adopting an estimation theoretic approach [5] using techniques such as Kalman filtering [1]. Another unique feature is the blind steganalysis approach which assumes no knowledge of the cover images. The results indicate the potential for its usage by law enforcement agencies in detecting and extracting secret information hidden by the DIS.

Keywords: Distributed image steganography, steganalysis, kalman filtering, estimation approach, stego images.

Abbreviations: Distributed Image Steganography (DIS), Blind Steganalysis Scheme (BSS).

I. Introduction

Steganography [7] is the science of hiding a secret (text or image) inside an innocuous text or image, making the secret invisible to the naked eye. Steganalysis is the process of reversing the steganographic scheme to detect and disclose the secret. The purpose of steganography is to have a covert communication channel between two parties whose existence is unknown to a possible intruder; and successful intrusion relies on detecting the existence of the communication. Steganography, which hides the existence of the secret message, is different from cryptography where the existence of the secret message is not disguised, but the content is obscured. Cryptography is used in situations where the possibility of jamming a communication system, to extract secret information, is quite difficult. For example, military applications have a sophisticated communication system which cannot be jammed easily, and their secret information can be transmitted safely using cryptography. On the other hand, steganography is used in places where it is easy to jam into the communication system which makes it quite popular among criminals to use it for illegitimate purposes.

Older steganographic schemes focus on hiding the secret inside a single host image. They can either be *spatial domain embedding* like *least significant bit embedding* [4] or *transfer domain embedding* like *discrete cosine transform (DCT)* or *wavelet transform (WT)* [10]. These methods are simple and straightforward but lack in the amount of information that can be hidden inside the host image. Moreover these methods are subject to single-point-failure since the information is hidden as just one secret text or image, and the chances of an intruder getting hold of it is quite high. An emerging steganographic scheme which overcomes these disadvantages is called *distributed image steganography (DIS)* [11]. DIS concentrates on image steganography where a secret image is split into n secret shares using a secret sharing scheme [9]. Out of these n secret shares, any k $(k \leq n)$ shares can be used for reconstructing the secret image but $k-1$ or fewer cannot do the same. This makes it very robust and reliable, and a successful countermeasure to the DIS (*i.e.* steganalysis) is to detect and disclose the secret image. The open literature offers blind detection schemes [8] which can detect the presence of the secret image without the knowledge of the cover image into which they are hidden, however this alone is not sufficient because there may be chances of misdetection and disclosing the secret would actually defeat the very purpose of using DIS for illegitimate purposes.

This paper focuses on designing a blind steganalysis scheme which can disclose the secret by using the estimation theoretic approach. The estimation approach constructs a *process model* for the parameter (secret) to be estimated, and an *observation model* for the information (DIS stego image) that is available [5]. Estimation techniques, such as particle filter or Kalman filter [1] can then be used to estimate the parameter. The process model and the observation model can be linear/non-linear with Gaussian/non-Gaussian inputs based on which a suitable estimation technique is used. This paper presents the basic concept of the proposed method and illustrates it with simulation results.

The rest of the paper is organized as follows: Section 2 details some background research, Section 3 explains the blind steganalysis scheme, Section 4 shows some results obtained through the scheme, and Section 5 presents the conclusion and some ideas for future work.

II. Background research

DIS makes use of a secret sharing scheme to split the secret image into multiple secret shares, and hide them inside the cover images. The basis of this method is a (k, n) threshold based *secret sharing scheme (SSS)* suggested by Shamir [6] which was used later by Thien and Lin [9] on images.

A. Secret Sharing and Image secret sharing

Blakely [3] and Shammir [6] developed the concept of threshold based secret sharing technique. A polynomial function of power $(k-1)$ is used to construct n shares from a secret value. The polynomial function is of the form:

$$f(x) = (d_0 + d_1 x + d_2 x^2 + \cdots + d_{k-1} x^{k-1}) (\text{mod } p) \quad (1)$$

where d_0 is the secret value, the n shares are $(x_1,\ f(x_1))$, $(x_2,\ f(x_2))$, $\cdots$, $(x_n,\ f(x_n))$ and p is the prime modulus used in the cryptographic computation. If k shares are available, one can construct k polynomial equations which are then solved to reveal the secret d_0. The technique allows any k out of the n shares to reconstruct the secret value, and any $k-1$ or fewer shares do not reveal any information about the secret value d_0.

Thien and Lin [9] proposed an image secret sharing scheme by using the Shamir's secret sharing scheme to generate image shares, also known as *shadows*. Suppose I is an $m \times m$ pixel image with intensity $I(i,j)$ where $1 \leq i \leq m$, $1 \leq j \leq n$. According to Shamir's secret sharing scheme, the polynomial function of degree $(k-1)$ for I can be computed as

$$\begin{aligned} S_x(i,j) \quad = \quad & I(i \times k + 1, j) + I(i \times k + 2, j)x + \ldots + \\ & I(i \times k + k, j)x^{k-1} (\text{mod } p) \end{aligned} \quad (2)$$

This method reduces the size of each secret share (or shadow) to $1/k$ of size of the secret image. Any k shadows are sufficient to reconstruct the secret image.

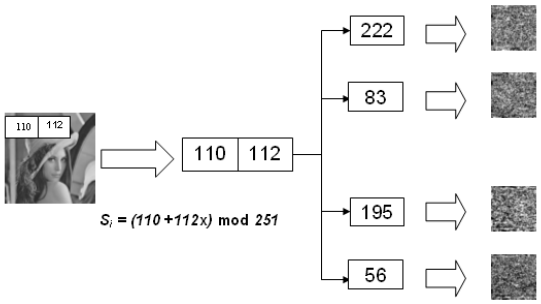

Fig. 1. Example of a $(2,4)$ image secret sharing.

An example of a $(2,4)$ image secret sharing scheme is illustrated in Figure 1 where $k = 2$ and $n = 4$. The first order polynomial function can be constructed as

$$S_x(1,1) = (110 + 112x) (\text{mod } 251) \quad (3)$$

where 110 and 112 are the first two pixel intensities in the Lena image. For the four participants, four random x values are chosen, and substituted into the polynomial function. The value of p is set as 251, which is the largest prime number lesser that 255 (maximum gray image value). Four shares are computed as (1,222) , (2,83) , (3,195) and (4,56), which form the four image shares for the first pixel. This process is repeated until all pixels are encoded. Four image shares, each half the size of secret image, are obtained as shown in the right side of Figure 1.

B. DIS

Wu [11] introduced Distributed image steganography which combines Shammir's and Thien and Lin's method. It produces a robust and reliable communication system to hide a secret image inside multiple cover images. Figure 2 shows the block

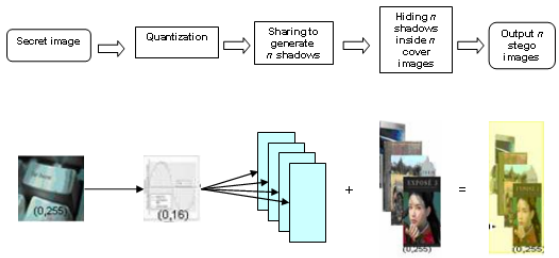

Fig. 2. Block diagram and example for DIS.

diagram for the DIS. There are three main processes, namely: *quantization, sharing and hiding*. *Quantization* reduces a secret image of 256 pixel intensity levels (0 to 255 since it is a gray scale image) into 17 pixel intensities. Then *sharing* splits this quantized secret image into secret shares each having pixel intensities distributed between -8 and 8. Finally, *hiding* embeds the secret shares inside the cover images. The secret shares, due to their small pixel intensity level, don't alter the pixel intensity levels of the cover images that much, and so the stego images look very similar to the cover images. DIS can be compared to a wireless multiple input multiple output (MIMO) communication system as shown by Figure 3.

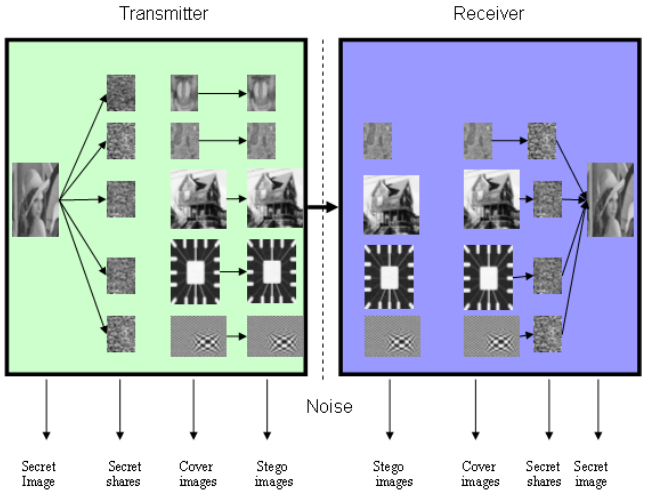

Fig. 3. DIS and Wireless MIMO Communication system.

DIS uses diversity to increase the channel capacity (*i.e.* information hiding capacity) by sending the signal (*i.e.* secret image) projected along a different axis (*i.e.* different segments of the secret image) through each channel (*i.e.* host image). Figure 3 shows the example of a $(4,5)$ DIS system. In this system, an attacker can monitor or disrupt the steganographic images transmitted to the receiver. However, any four stego

images can be used to reconstruct the secret image. This illustrates the advantages of the DIS system of being more reliable than conventional steganographic techniques. The stego-images produced by the DIS offer similar benefits to the transmitted secret information as observed in MIMO systems for communication. However, DIS attracts attention for using it for illegitimate purposes because of its inherent robustness and security. In such cases, a suitable steganalysis technique needs to be developed in order to reverse the DIS, which is described in section 3.

III. Blind steganalysis scheme

This section explains the Blind Steganalysis Scheme (BSS) using the estimation theoretic approach, which assumes no knowledge of the cover images. Figure 4 shows the block diagram for the BSS. It is based on the following assumptions:

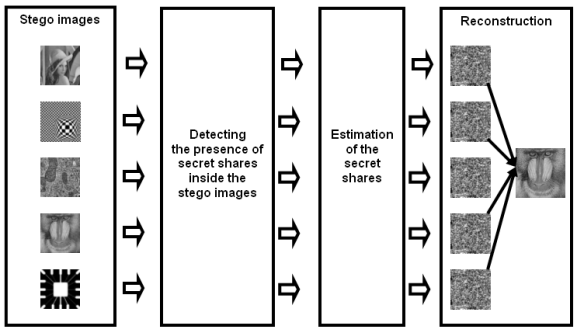

Fig. 4. Block diagram for the BSS

i) The number of images detected for the presence of hidden content is equal to number of secret shares (*i.e* k) required for reconstructing the secret image.
ii) All secret shares hidden inside the detected stego images are derived from the same image.

There are a number of blind detection schemes that are available in the literature, and this research made use of *image quality matrix (IQM)* [2]. The main focus of this paper is on the estimation module which attempts to estimate (*i.e.* extract) the secret shares so as to hamper the communication established by the DIS. There are three main steps involved here:

i) Consider Secret share as process model
ii) Consider Stego image as observation model
iii) Use Kalman filtering to obtain the secret shares

A. Secret share as process model

Each secret share hidden inside the detected image can be represented as a random sequence. The secret shares are converted into one dimensional column vectors represented by $s(n)$. Each secret share can then be modeled as a dynamic system equation similar to a dynamic equation used in the estimation process [1]. The general equation can be represented as:

$$s_k = A_k s_{k-1} + B_k c_k + u_k \tag{4}$$

Here A_k represents the state transition matrix, B_k represents the control input matrix, and u_k represents the noise. So each pixel $s(n)$ in the secret share can be represented in a similar equation with the assumption that there is no control input c_k, and the state transition matrix A_k is unity. Thus the process model representing the secret share can be written as

$$s(n) = s(n-1) + u(n) \tag{5}$$

where $-8 \leq s(n) \leq 8$, for any n, and $u(n)$ is the process noise associated with the model with zero mean and non-zero covariance Q. The process noise basically represents the difference between adjacent pixels of the secret share which needs to be determined to define the process model completely. For using the Thien and Lin's image secret sharing method, the equation representing the process noise is taken as

$$u(n) = [s(n) - s(n-1)] \bmod 17 \tag{6}$$

Figure 5 shows the histogram of the process noise $u(n)$ and is distributed between -8 and 8 as evident from the Thien and Lin's image secret sharing scheme. The reason for this could be attributed to the fact that the quantization process reduces the intensity level of a gray scale image from 256 (0 to 255) to a quantized image of 17 (-8 to 8) levels.

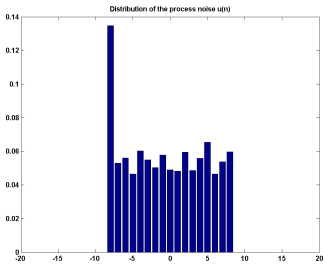

Fig. 5. Distribution of the process noise.

B. Stego image as the observation model

Stego images are obtained by embedding each secret share onto the cover image. Since the secret shares have a low pixel intensity level (*i.e* 17 levels), they do not alter the pixel intensity level of the cover image significantly so that the stego image and the cover image look almost identical. This particular feature of the stego image is used to create an observation model for the same. Since this is a *blind steganalysis scheme* (*i.e.* no cover images) with the only data available being the stego images, the observation model is determined using a two-step process:

i) Predict the pixels intensities of every cover image from the corresponding stego image
ii) Construct Observation equation

1) Predict the pixel intensities of the cover image: Most of the pixels of a natural image have some correlation with their neighboring pixels. Considering this fact, every pixel of the cover image is predicted using the neighboring pixels

of the corresponding stego image. The stego images and the cover images look identical, and this is reflected by their pixel intensities which are close to each other. So this idea is used to create a model where each pixel of the cover image is predicted by taking the mean of the neighboring pixels in the corresponding stego image. Figure 6 illustrates this model.

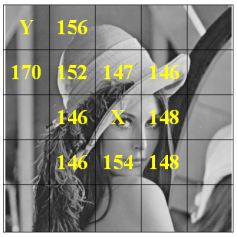

Fig. 6. Predicting the pixel intensities of the cover image from the stego image.

The figure shows two cases: Case I: the pixel is in the middle (X), and Case II: the pixel is at the edge (Y). The pixel values X and Y are calculated by averaging pixel intensities of neighboring pixels as shown in the following equations

$$X = \frac{152 + 147 + 146 + 148 + 148 + 154 + 146 + 146}{8} \tag{7}$$

$$Y = \frac{170 + 152 + 156}{3} \tag{8}$$

Let $h(n)$ denote the pixel intensities of a cover image and $\hat{h(n)}$ denote the pixel intensities of the cover image predicted from the corresponding stego image $x(n)$. Then the equation representing the prediction process described above can be written as

$$\hat{h(n)} = f(x(n)) \tag{9}$$

2) Observation equation: Each stego image $x(n)$ is obtained by embedding the secret share $s(n)$ inside the cover image $h(n)$. Note that $x(n)$ is already available, and the pixels of the cover image were predicted form the stego image (*i.e* $\hat{h(n)}$) as explained earlier. Therefore the equation representing the stego image as the observation model can be written as

$$x(n) = \hat{h(n)} + s(n) + w(n) \tag{10}$$

The observation noise $w(n)$ associated with this equation is due to the prediction process explained in the previous section. It is assumed to be Gaussian $N(0, R)$, and the covariance R needs to be determined. For that purpose, the observation equation can be written as

$$x(n) - \hat{h(n)} = z(n) = s(n) + w(n) \tag{11}$$

Since $z(n)$, $s(n)$ and $w(n)$ are independent of each other, the covariance of the observation noise R is equal to the difference between the covariances of $z(n)$ and the secret share $s(n)$ (Q). The equation describing this can be represented as

$$R = cov(z(n)) - Q \tag{12}$$

After determining the the observation noise covariance R, the observation equation is defined completely. The next step is to use Kalman filtering to estimate the secret shares.

C. Kalman filtering

Kalman filtering uses the process model and the observation model created for each *secret share-stego image* pair to obtain an estimate of each secret share. It aims at estimating the *a posteriori* state of a process, using the *a priori* estimate of the process (*i.e* process model) and the measurements (*i.e* observation model) [1]. Figure 7 illustrates the flowchart representing the Kalman filtering process used to estimate each secret share.

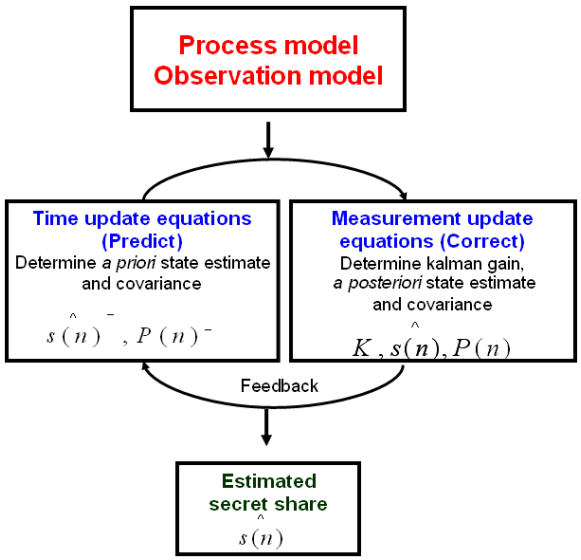

Fig. 7. Flowchart representing the Kalman filtering process to estimate the secret share.

IV. Results

Since a $(4, 5)$-DIS system is assumed here, any 4 out of the 5 estimated secret shares, as described from the previous section, can be used for reconstructing the secret image. The reconstruction module reverses the Thien and Lin's secret sharing method to get the quantized image and decodes the quantized image back to the original secret image. Figures 8, 9 and 10 shows the results obtained through the blind steganalysis scheme. The secret image considered here is *Baboon* (256×256 *pixels)*, and this is spilt into 5 secret shares which are hidden inside 5 cover images - *Lena, Cell, Neal, Baboon and Board* (128×128 *pixels)*. Figure 8 shows the accuracy graph which compares the estimated secret shares (obtained through the blind steganalysis scheme) to the original secret shares (obtained from the Thien and Lin's secret sharing scheme). The figure shows that the blind steganalysis scheme performs really well for secret shares embedded inside Lena, Baboon and Board images with an accuracy close to 90 percent.

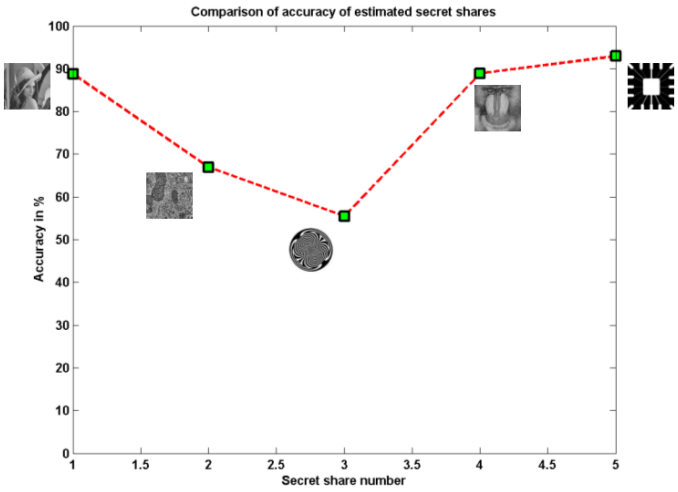

Fig. 8. Accuracy of Estimated secret shares - Baboon as the secret image

The combination of the four best secret shares (*i.e* the secret shares inside Lena, Cell, Baboon and Board) and the four worst secret shares (*i.e* secret shares inside Lena, Cell, Neal and Baboon) are used to reconstruct the estimated secret images which are shown in Figures 9 and 10. The steganalysis scheme clearly reveals the presence of a hidden secret in the host images, and recovered few significant features including the nose, eyes, face etc. The steganalysis scheme also performs fairly for the worst estimated secret image as there is considerable amount of information retrieved as opposed to none.

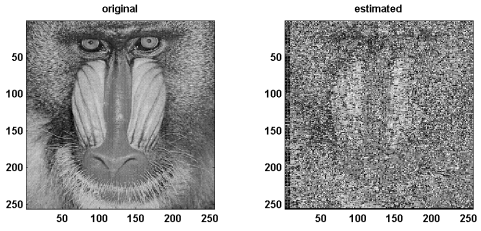

Fig. 9. Comparison of Original and Estimated secret image - Best performance with Baboon as the secret image

V. Conclusions

In this paper, a blind steganalysis scheme (BSS) is presented for disclosing a secret image hidden inside multiple cover images (*i.e.* DIS). The proposed method is based on estimation theoretic approach. DIS is a secure and reliable system, but attracts the attention of using it for illegitimate purposes.

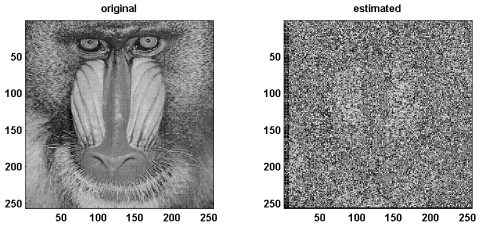

Fig. 10. Comparison of Original and Estimated secret image - Worst performance with Baboon as the secret image

The proposed BSS reverses the DIS and discloses important details about the secret information hidden in the host images. This could be used as an excellent intelligence gathering tool for law enforcement agencies to crack down cases where criminals use DIS for illegitimate purposes. Also the BSS could serve as a general steganalysis technique for other image steganographic schemes. Accuracy of the estimation could be improved using other estimation techniques, such as particle filters or extended Kalman filters.

References

[1] M. Arulampalam, S. Maskell, and N. Gordon, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Trans. Signal Process*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[2] L. Bai, S. Biswas, and E. Blasch, "An estimation approach to extract multimedia information in distributed steganographic images," presented at the 10th International Conference on Information Fusion (ICIF), Quebec city, Canada, July 2007.

[3] G. R. Blakley and C. Meadows, "Security of ramp schemes," presented at the Advances in Cryptology – Crypto '84, G. R. Blakley and D. Chaum, Eds., vol. 196, Santa Barbara, California, USA, Aug. 1984, pp. 242–269.

[4] F. J, G. M, and R. Du, "Detecting lsb steganography in color, and gray-scale images," *IEEE Multimedia*, vol. 8, no. 4, 2001.

[5] S. M. Kay, *Fundamentals of statistical signal processing:estimation theory*. Upper saddle river,New Jersey 07458: Prentice Hall, 1993.

[6] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979.

[7] G. Simmons, "Prisoners problem and the subliminal channel," presented at the Advances in Cryptology: Proceedings of CRYPTO 83, 1983.

[8] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, "Steganalysis for markov cover data with applications to images," *IEEE Trans. Information Forensics and Security*, vol. 1, no. 2, pp. 275–287, 2006.

[9] C.-C. Thien and J.-C. Lin, "Secret image sharing," *Computers & Graphics*, vol. 26, no. 5, pp. 765–770, 2002.

[10] Y. Wang and P. Moulin, "Steganalysis of block-dct image steganography," *IEEE workshop on Statistical Signal Processing*, Sept. 2003.

[11] Y.-S. Wu, C.-C. Thien, and J.-C. Lin, "Sharing and hiding secret images with size constraint," *Pattern Recognition*, vol. 37, no. 7, pp. 1277–1385, 2004.