OPTIMIZATION OF VARIOUS SPECIFIC AND GENERIC STEGANALYSIS SCHEMES

Rajy Xavier
Assistant Professor/ECE
MET'S School of Engineering
Mala,Thrissur.
rajyxavier@gmail.com

Sinu Mathew K.
P.G Scholar
Hindusthan College of
Engineering,Coimbatore.
sinumathew@yahoo.com

Abhijith Augustine
Assistant Professor/EEE
MET'S School of Engineering
Mala,Thrissur.
abhijith04@gmail.com

ABSTRACT
The objective of steganalysis is to detect messages hidden in cover objects, such as digital images, video and audio. The ultimate goal is to extract and decipher the secret message. In a few recent cases, images are used as carrier medium by some unauthorized users as it is less suspected for criminal purposes. There are two types of image steganalysis techniques referred as Specific and Generic steganalysis schemes. The Specific approach represents a class of image steganalysis techniques that very much depend on the underlying steganographic algorithm used and have a high success rate for detecting the presence of the secret message if the message is hidden with the algorithm for which the techniques are meant for. The Generic approach represents a class of image steganalysis techniques that are independent of the underlying steganography algorithm used to hide the message and produces good results for detecting the presence of a secrete message hidden using new and/or unconventional steganographic algorithms. Various steganalysis techniques are implemented and analyzed to find out the most efficient one.

General Terms
Histogram, Stegoimage

Keywords
Steganography, Steganalysis, LSB Embedding, Closest Color Pair

1. INTRODUCTION
We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace the content with your own material. Steganography is the art of invisible communication. Its purpose is to hide the very presence of communication by embedding messages into innocuous-looking cover objects. In today's digital world, invisible ink and paper have been replaced by much more versatile and practical covers for hiding messages – digital documents, images, video, and audio files. As long as an electronic document contains perceptually irrelevant or redundant information, it can be used as a "cover" for hiding secret messages. In this paper, we deal solely with covers that are digital images stored in the JPEG format.

Each steganographic communication system consists of an embedding algorithm and an extraction algorithm. To accommodate a secret message, the original image, also called the cover-image, is slightly modified by the embedding algorithm. As a result, the stego-image is obtained.

Steganalysis is the art of discovering hidden data in cover objects. As in cryptanalysis, we assume that the steganographic method is publicly known with the exception of a secret key. The method is secure if the stego-images do not contain any detectable artifacts due to message embedding. In other words, the set of stego images should have the same statistical properties as the set of cover-images. If there exists an algorithm that can guess whether or not a given image contains a secret message with a success rate better than random guessing, the steganographic system is considered broken.

Medical records of patients are extremely sensitive information, requiring uncompromising security during both storage and transmission. In addition, these records often have to be traceable to patient medical data such as X-ray or scan (CAT, MRI etc.) images. Nowadays medical images are suspected for unauthorized communication by using steganographic techniques. As the medical images are less suspected, even terrorists are using it as a safer cover media.

2. STEGANOGRAPHY METHODS
In this section five data hiding methods that are used to generate stegoimages are described.

2.1 LSB embedding Method

LSB embedding method is one of the most popular steganography methods due to its simplicity, high embedding capacity and high imperceptibility of secret message. In the LSB embedding method, image is decomposed in bits planes (8 bits planes for 8 bits gray scale images and 24 bits planes for color images), and its least significant bits (LSB) plane is replaced by secret message. Generally secret message is encrypted by any encryption algorithm before its embedding. Here two public domain steganographic tools based on LSB embedding method: Stools and Invisible Secret which are described in .

2.2 DCT Domain Embedding Method

In the DCT Domain embedding method, firstly the cover image is transformed by DCT and then the embedding process is performed in the DCT coefficients instead of in the image pixels. The principal advantage of this method is that it is more secure than the LSB embedding method against many statistic analyses; however the embedding capacity of the secret message is limited by imperceptibility constrains. In this paper two data hiding methods, Huang's data hiding

method and Piva's data hiding method, are used to generate stegoimages.

2.3 Bit-Plane Complexity Segmentation Steganography

Bit-Plane Complexity Segmentation Steganography (BPCS) is similar to the LSB embedding method. In BPCS, the image is segmented in blocks of 8x8 pixels, and each block is decomposed in bit-planes (for 8 bits gray scale image, 8 bit-planes are decomposed). In each segmented bit-plane its complexity is analyzed. If the complexity of a segmented bit-plane is higher than a predetermined threshold, this segmented bit-plane is replaced by the secret message. Therefore in BPCS, depending on its complexity and the secret message is embedded in any bit planes, not only in the LSB plane.

Here Quickstego is used for the purpose of steganography. The text to be hidden and the cover image is given as input to the Quickstgo software. The output will be the stegoimage that contains the secret message hidden inside the cover image, without any visible change in the cover image. Here medical images are considered as the application since it is extensively used for steganalysis for nowadays.

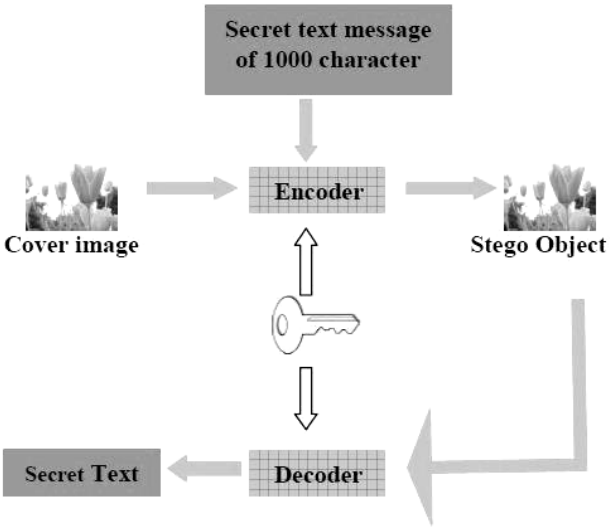

Fig 1: Stegnography Process

3. STEGANALYSIS METHODS

In this section, three steganalysis methods: difference image histogram method, closest color pair method and features extraction method, are described.

3.1 Difference Image Histogram Method

The histogram distribution may be used to discriminate stegoimages from natural images. However the variation of the distribution between different types of images is bigger than the variation between a natural image and its stegoimage, therefore the histogram distribution only can not be used for this purpose. Zhang et al. proposed the difference image histogram method, that generates a difference image D, calculating the difference value between two adjacent pixels of the image as given by,

$$D(i,j) = I(i+1,j) - I(i,j) \qquad (1)$$

Three difference images *Dh, Df, Dg* of a suspicious image, the image with flipped LSB and the image with LSB replaced by zero are calculated by (7). Histograms of these difference images are calculated and demoted by

$H = \{hi \setminus i = -255 \ldots 255\}$, $F = \{fi \setminus i = -255 \ldots 255\}$ and

$G = \{gi \setminus i = -255 \ldots 255\}$respectively. Zhang et al. observe d that these histograms are related with each other when the image doesn't contain any hidden message, however if the image contains some secret message in its LSB plane, the relationship of these three histogram is broken. The relationships between these three histograms are described by Figure 2.

In the Figure 2, $a2i,2i-1,a2i,2i$,$a2i,2i+1$ are the values of transition from histogram G to H, and from G to F. Using these values, following three values are calculated as follows.

$$\alpha_i = (a_{2i+2,2i+1}) / (a_{2i,2i+1})$$
$$\beta_i = (a_{2i+2,2i+3}) / (a_{2i,2i-1}) \qquad (2)$$
$$\gamma_i = (g_{2i}) / (g_{2i+2})$$

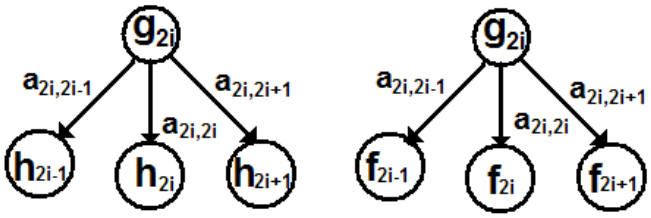

Fig. 2 Transition values *a2i,2i−1,a2i,2i ,a2i,2i+1* from G to H and F.

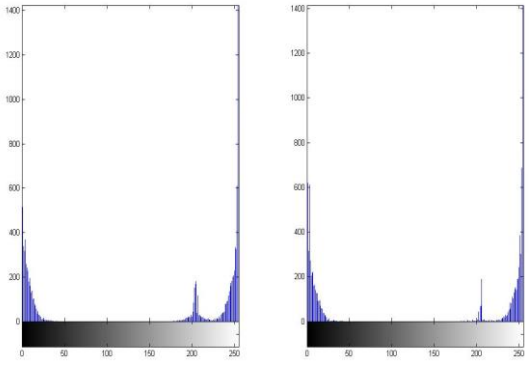

Fig. 3 Comparison of histogram of a stegoimage and ordinary image

If the image contains some secret message, (3) is satisfied for all values of i =-255…255.

$$\alpha_i \approx 1 \qquad (3)$$

and for natural images, (4) is satisfied.

$$\alpha_i \approx \gamma_i \qquad (4)$$

This method can be applied directly to grayscale images, but for color images, some considerations must be taken account

3.2 Closest Color Pair Method

Fridrich et al. observed that the number of close color pairs is increased considerably when an image has a secret message embedded in its LSB plane, and they proposed a steganalysis method for LSB embedding technique using a ratio between the number of closest color pair and all pairs of colors of the image. In this method, the number of all color pairs existing in an image and the number of close color pairs within all existing color pairs are computed. Then on purpose, a LSB embedding steganography algorithm is applied to the image, and also number of existing color pairs and close color pairs are computed.

The condition of close color pair for (C_1 = $[R_1,G_1,B_1]$,C_2=$[R_2,G_2,B_2]$) is given by

$$(R_1-R_2)^2+(G_1-G_2)^2+(B_1-B_2)^2\le 3 \qquad (5)$$

The ratios *R* and *R'* between the number of close color pair and the number of all color pairs is calculated for the image under analysis and its steganography version as

$$R=\frac{P}{\binom{U}{2}} \qquad (6)$$

$$R'=\frac{P'}{\binom{U'}{2}} \qquad (7)$$

where *P* and *P'* are the number of close color pairs, and *U* and *U'* are number of all color pairs in the image and its steganography version, respectively. If (8) is satisfied, the image can be considered as natural image, otherwise the image contains some secret message.

$$\frac{R}{R'}\ge Th \qquad (8)$$

Fridrich et al. proposed, after exhaustive proves, that a suitable threshold value *Th* is 1.1.

3.3 Features Extraction Method

Recently the DCT domain embedding methods were proposed instead of LSB embedding method. The efficiency of two steganalysis methods mentioned above for LSB embedding method is shown, but there are not information about the efficiency to detect the DCT domainbased embedding method. The feature extraction method proposed by [6] extracts 39 features, which are used to classify images as natural images or stegoimages. In this section, we describe the features based on statistical moments of wavelets characteristic function for the steganalysis.

3.3.1 De-correlation of wavelet transform

The histograms of all wavelet subbands only reflect the statistical distribution of coefficients in the subband, but it doesn't reflect the correlation of the coefficients within this subband. The wavelet transform is well known for its capability of multi-resolution decomposition and coefficients de-correlation. It is known that for discrete wavelet transform, different high frequency subbands within one level will be uncorrelated to each other. The features extracted from one high frequency subband are thus uncorrelated to that extracted from another high frequency subband at the same level. Therefore, features from different dimensions most likely uncorrelated to each other. From this point of view, this multi-dimensional feature vector will be suitable to represent the image for steganalysis purpose.

3.3.2 Characteristic Function and Its Statistical Moments

The data hiding process can be modelled as an additive signal, which is independent to the cover-image; this signal is added to the cover media. It is well known that the effect of the additive signal on the image is equivalent to a convolution of two probability density functions (PDFs). According to one interpretation of the characteristic function (CF) is that it is complex conjugate of Fourier transform of the PDF. We can consider the PDF as the normalized version of a histogram, in this case the image histogram and the subbands coefficients histogram. The CF is defined as:

$$H(f_1)=DFT(h[x])=\sum_{x=0}^{N=1}h[x]e^{\frac{2\pi f_x f_1}{N}} \qquad (9)$$

where H(fi) is the equal to CF, DFT(h[x]) is the Discrete Fourier Transform of the histogram, N is the total number of points in the horizontal axis of the histogram, fi is frequency component. Owing to the de-correlation capability of the discrete wavelet transform (DWT), the coefficients of different subbands at the level, we can assume independence to each other. Therefore, the features generated from different wavelet subbands at the same level are independent to each other as well. This property is desirable for steganalysis. Therefore, it is proposed to use the statistical moments of the characteristic functions of wavelet subbands as futures for steganalysis. The n-th statistical moment of a CF is defined as follows

$$M_n=\frac{\sum_{j=1}^{N/2} f_j^n\left|H(f_j)\right|}{\sum_{j=1}^{N/2}\left|H(f_j)\right|} \qquad (10)$$

where $|H(fi)|$ is the magnitude of the CF component at frequency *fi*, *N* is the total number of points in the horizontal axis of the histogram. The zero frequency component of the CF is excluded from the calculation of moments because it represents only the summation of all components in the discrete histogram. For an image, the zero frequency component is the total number of pixel, while for a wavelet subband, it is the total number of the coefficients in the subband. In either case, it does not change during the data hiding process. In order to get the feature, the subbands decomposed by Haar wavelets until three levels were used. Therefore, there are 12 subbands, denoted by LL1,HL1, LH1, HH1, LL2, HL2, LH2, HH2, LL3, HL3, LH3, HH3. The first three moments for each of subbands and the test image, denoted by LL0, result a vector with 39 features.

4. CONCLUSION

When the cover-images are stored in the JPEG format, the detection method must be modified to accommodate the effects of double JPEG compression produced by the embedding. The F5 always decompresses the cover-image and recompresses it using a user-defined quality factor. This leads

to artifacts in coefficient histograms (jaggedness) that may introduce quite large detection errors. Fortunately, the previous JPEG compression can be estimated from the stego-image and the same compression/decompression that occurred prior to applying the F5 can be carried out for the cropped stego-image before deriving the estimated histograms for comparison. This small modification of the detection algorithm dramatically improves the performance and makes the accuracy and reliability of our results independent of the cover-image format. The method for obtaining the cover-image histogram by cropping and low-pass filtering can in fact be used for designing detection mechanisms for other steganographic schemes that manipulate quantized DCT coefficients. We can use different statistical quantities rather than first-order statistics in the frequency domain to obtain their baseline values. For example, the increase of "blockiness" (the sum of spatial discontinuities at block boundaries) during embedding can be used as the distinguishing quantity for OutGuess. Using this measure, we have been able to successfully attack OutGuess . The blockiness measure increases with embedding for most steganographic schemes for JPEGs independently of their inner mechanisms. This opens up a new direction in steganalysis of JPEG images that yet needs to be further explored.

5. REFERENCES

[1] Natarajan Meghanathan and Lopamudra Nayak,Steganalysis algorithms for detecting the hidden information in image,audio, and audio cover media. International Journal of Network Security & Its Application (IJNSA), Vol.2, No.1, January 2010.

[2] Xiaozhong Pan, BoTao Yan, Ke Niu,Multiclass Detect of Current Steganographic Methods for JPEG Format Based Re-stegnography,2010.

[3] Hafiz Malik, K.P.Subbalakshmi, R.Chandramouli, Steganalysis of QIM-Based Data Hiding using Kernel Density Estimation.

[4] Jessica Fridrich, Miroslav Goljan, Dorin Hogea,Steganalysis of JPEG Images: Breaking the F5 Algorithm.

[5] Yeshwanth Srinivasan, Brian Nutter, Sunanda Mitra,Benny Phillips, and Daron Ferris,Secure Transmission of Medical Records Using High Capacity Steganography.

[6] Johann Barbier, _Eric Filiol, and Kichenakoumar Mayoura, New features for speci_c JPEG Steganalysis.

[7] A. D. Ker, "Steganalysis of LSB Matching in Grayscale Images", IEEE Signal Processing Letters, vol. 12, no. 6, pp. 441-444, 2006.

[8] W. Lie and L. Chang, "Data Hiding in images with adaptive numbers of least significant bits based on human visual system", in *Proc.,IEEE Int. Conf. Image Processing*, pp. 286-290, 1999.

[9] J. Fridrich, M. Goljan and R. Du, "Detecting LSB Steganography in Color and Gray Scale Image*", IEEE Multimedia*, vol.8, no. 4, pp. 22- 28, 2001.

[10] J. Fridrich, R. Du and M. Long, "Steganalysis of LSB Encoding in Color Images" , in *IEEE Int. Conf. on Multimedia and Expo*, 2000, 1279-1282.