Statistical Steganalysis for Content-Adaptive Steganography

V. Gokula Krishnan\(^1\), M. Deepak\(^2\), S. Praveen Kumar\(^3\), B. Vinoth Kumar\(^4\)
Assistant Professor\(^1\), UG Scholar\(^2\,3\,4\)
Department of CSE
Panimalar Institute of Technology, Chennai, India
gokul_kris143@yahoo.com\(^1\), deepakatcrime@gmail.com\(^2\), pravinhawk6@gmail.com\(^3\), vinothbabu2426@gmail.com\(^4\)

Abstract:
Network is a group of two or more computer systems linked together. There are many types of computer networks, including the following: local-area networks (LANs), wide-area networks (WANS) metropolitan area networks (MANS). Network security consists of the policies adopted to prevent and monitor unauthorized access, misuse, modification, or denial of a computer network and network-accessible resources. Network security involves the authorization of access to data in a network, which is controlled by the network administrator. In this project, we transfer the data from sender to receiver securely by using Steganography method. Steganography means hiding the data inside the image. In existing system, they will hide all data inside the image even the data is not that much of sensitive, so it takes more time and also they will not compress the large data. But it proposed system, user can share the data without hiding the data into the image and also in the same way user can share the image also which he wants to share securely. Before sharing all these data, the user can compress the data so the transmission speed from sender to receiver will increase. Finally the user can view the transmission path of the data between the sender and the receiver.

Keywords: Steganography, Networks, Authorization, Security, Transmission.

INTRODUCTION
Information hiding has received an increasing interest in the last decades driven by the numerous possible applications such as water mark identification and tampering detection. Unfortunately malicious usage of information hiding like Steganography, Steganography and steganalysis are a cat and mouse game: steganographers embed a secret message in a cover medium while steganalysts try to detect the presence of this hidden message. This paper focuses on the simple but popular LSB replacement. Surely, much better algorithms are nowadays available. However, the proposed methodology can be applied to other schemes providing that a statistical model of steganographic impact is available. With many tools available in the public domain, Steganography is within reach of anyone, for legitimate or malicious usage. It is thus crucial for security forces to reliably detect steganographic content among a set of media; many methods have been proposed for this purpose. Even though some steganalyzers are very efficient, detection rate is not the only performance criterion is some circumstances. For instance, when carrying an investigation, steganalysis results will hardly be accepted without an analytically predictable and warranted false alarm rate. In this situation, supervised learning based method can hardly be used. This justifies the statistical study of a Steganography detection scheme to which this paper is devoted. The study of steganalysis as a hypothesis test requires an accurate image model; only few works in the literature explicitly use such models.

In the distribution of Discrete Cosine Transform (DCT) coefficients is used to detect Steganography in JPG images. In a similar fashion, the distribution of Discrete Fourier Transform (DFT) coefficients is used. An independent and identically distributed (i.i.d) pixels model is exploited in to derive a statistical hypothesis test.

Our motivation for solving the problem of the cost-function design comes from the HUGO algorithm that assigns the costs of individual changes based on the pixel neighborhood. Unfortunately, this approach does not easily generalize to other cover sources, such as JPEG or color bitmap images, neither is it clear how to optimize the design. Here we open the question of the cost-function design and strive for a robust approach that generalizes well to unseen Cover images and unseen steganalysis features to avoid over fitting to a particular cover source and feature space. For example, the Feature Correction Method, which is a heuristic approach to embed while approximately preserving the cover image feature vector, is known to be overly sensitive to the chosen feature set and does not generalize or scale well. The work in has an alternate feature preservation approach and also empirically considers the dynamics between steganographers and steganalyzers.

Empirical Design of Cost Functions
We focus on designing adaptive embedding schemes for the payload-limited sender subjected to sequential steganalysis.
In this regime, the sender decides on the number of bits he wants to hide in a given cover object, embeds his payload, and sends the stego object through a passively Monitored channel. In sequential steganalysis, the warden has to decide whether a given image is cover or stego solely based on a single object. We deliberately omit the possibility of intentionally spreading the payload into a group of cover images – a technique known as the batch Steganography. This mode can improve the security of the scheme; however, it should no longer be tested with sequential steganalysis. A common way of testing steganographic schemes is to report a chosen detection metric (ROC curve, accuracy, minimum error probability under equal priors PE, etc.) empirically estimated from a database of cover and stego images where each stego image carries a fixed relative payload. Whenever possible, we report results obtained from cover images of roughly the same size to reduce the effect of the square root law. Our goal is to design a set of functions $i, i \in \{1, \ldots, n\}$, which, given the original cover image, assign the cost of changing individual cover elements to their new values. For digital images, the dependence between two cover pixels rapidly decreases with their distance. In case of gray-scale spatial-domain digital images, the cost of changing a single pixel should mainly depend on its immediate neighborhood. For this reason, we constrain to be a real-valued function $\rho$ with small support, $\rho(x, y) = \Theta(xp(i), yi)$, where $xp(i)$ denotes cover pixels spatially close to pixel $i$. From practical experiments, it is possible to identify the quantity that should drive the costs.

The goal of this paper is threefold:

- To locally model the content of a cover image by describing the optical system which gives birth to a natural image?

- To exploit, as simply as possible, this model of natural image in the design of an almost optimal test.

- To numerically compare the proposed detection scheme with other steganalysis methods.

Steganography concerns the reliable transmission of a secret message buried in a host digital medium, such as digital image or audio file. This data hiding technique has been mainly used in information security applications and has received an increasing interest in the past decade. While a cyphered message can easily be detected the detection of data hidden within innocuous-looking digital media remains a difficult problem. More generally, the goal of steganalysis is to obtain any information about the potential steganographic system from an unknown medium. The “prisoners problem”, illustrates a typical scenario of Steganography and steganalysis. Alice and Bob, two prisoners, communicate by imperceptibly embedding a secret binary message $M$ into a cover-object $C$ to obtain an innocuous looking stego-object $S$.

Then, Alice send the stego-object $S$ to Bob through a public channel. Wendy, the warden, examines all their communications in order to detect whether the inspected object $Z$, contains a secret message $M$ or not.

1.1. State of the Art

Many steganographic tools are nowadays easily available on the Internet making Steganography within the reach of anyone, for legitimate or malicious usage. It is thus crucial for security forces to be able to reliably detect steganographic content among a (possibly very large) set of media file. In this operational context, the detection of a rather simple but most commonly found stego system is more important than the detection of a very complex but rarely encountered stegosystem. The vast majority of download able steganographic tools insert the secret information in the LSB plane. More precisely, two embedding functions have been widely studied: LSB replacement and LSB matching, also known as LSB±1. The LSB replacement method consists in substituting cover medium LSB by bits of secret message. The LSB matching scheme have been proposed as an improvement of LSB replacement; when the hidden bit to be inserted does not match cover medium sample LSB, it is proposed to randomly increment or decrement cover sample value. While substantial progress has recently been made in the detection of LSB replacement, the steganalysis of LSB matching remains a much harder problem. Therefore, the detection of steganographic algorithms based on LSB matching embedding remains a live research topic. It can be noted that many methods have been proposed to improve LSB replacement and/or LSB matching schemes. On the one hand, by using coding theory it has been proposed to improve embedding efficiency. Roughly speaking, the idea is to gather several samples and, using coding theory, to embed more than one bit of hidden data for each modification of cover medium samples. On the other hands, focusing on image Steganography, it has been proposed to choose the pixels in which it is more difficult to detect hidden bits. It has thus been proposed to locate edges within an image on the underlying assumption that textured areas are difficult to model which, consequently, makes the hidden data detection more difficult. Similarly, recently proposed algorithms select pixels location by minimizing a distortion function.

1.2. Contributions of the Paper

The recently proposed steganalyzers dedicated to LSB matching can be roughly divided into two categories. On the one hand, most of the latest detectors are based on supervised machine learning methods and use targeted or universal features set. As in all applications of machine learning, a difficult problem is to choose an appropriate feature set.
Moreover, the problem of measuring classification error probabilities remains open in the framework of statistical learning. On the other hand, it has been observed that LSB matching acts as a low-pass filter on the Histogram Characteristic Function (HCF) of a digital image. This pioneering work lead to an entire family of histogram based detectors. While histogram based detectors has been shown to be very efficient, they have been designed with a limited exploitation of cover medium model and hypothesis testing. Hence, their statistical properties are evaluated through numerical simulations but remain analytically unknown. In the operational context described above, the proposed steganalyzers must be immediately applicable without any training or tuning phase. For this reason, the use of a machine learning based detector is hardly possible. Moreover, the most important challenge for the steganalysts is to provide detection algorithms with an analytical expression for the false-alarm and missed detection probabilities without which the “uncertainty” of the result cannot be “measured.” The prior art LSB matching steganalyzers are certainly very interesting and efficient, buthead hoc design of these algorithms does not permit to calculate the detection errors probability. In addition, only a few theoretical results exist because steganalysis has seldom been thoroughly studied using hypothesis testing theory. Alternatively, the first step in the direction of hypothesis testing has been made in for LSB replacement to design a statistical test with known statistical properties. In the present paper, the statistical methodology proposed in the case of LSB replacement, based on a model of non identically distributed Gaussian samples, is extended for the LSB matching. It should be highlighted that for this extension is not immediate because by changing the embedding model the hypothesis testing problem is modified. Moreover, for the case of LSB matching, the design of an optimal test has never been studied even when all the signal parameters are known. Therefore, the goal of this paper is threefold:

1. Define the most powerful (MP) test in the theoretical case when the cover medium parameters are known, namely the expectation and noise variance of each sample.

2. Analytically calculate the statistical performance of the MP test in terms of the false-alarm and missed-detection probabilities. This result particularly allows us to calculate the decision threshold which warrants a prescribed false alarm probability, this is well known Constant False Alarm Rate (CFAR) detection. Moreover, the statistical performance of the MP test also provides an upper bound, which remains unknown, on the detection power of any detection scheme.

3. Design a practical efficient implementation of this test based on a simple local estimation of expectation and variance of each sample. Compared to the previous published works this paper proposes two main innovations:

1. It especially focuses on low signal-to-noise ratio for which the hidden data detection is the hardest.

2. This allows us to study the statistical properties of proposed test and provide an upper bound on the performance one can expect for any detector which aims at detecting LSB matching data hiding.

1.3. Organization of the Paper

The paper is organized as follows. The problem of LSB matching steganalysis is casted within the framework of hypothesis testing. Following the Neyman-Pearson approach, the MP Likelihood Ratio Test (LRT) is presented in Section 3 and its statistical performance is calculated in Section 4. Finally, the proposed practical implementation of the Generalized LRT (GLRT) is presented. To show the relevance of the proposed approach, numerical results on large natural databases of natural images and uncompressed sound.

RELATED WORKS

In [1] an adaptive least significant bit spatial domain embedding method. This method divides the image pixels ranges (0-255) and generates a stego-key. This private stego-key has 5 different gray level ranges of image and each range indicates to substitute fixed number of bits to embed in least significant bits of image. The strength of proposed method is its integrity of secret hidden information in stego-image and high hidden capacity. The limitation is to hide extra bits of signature with hidden message for its integrity purpose. It also proposed a method for color image just to modify the blue channel with this scheme for information hiding. This method is targeted to achieve high hidden capacity plus security of hidden message.

Yang et al., in [2] an adaptive LSB substitution based data hiding method for image. To achieve better visual quality of stego-image it takes care of noise sensitive area for embedding. Proposed method differentiates and takes advantage of normal texture and edges area for embedding. This method analyzes the edges, brightness and texture masking of the cover image to calculate the number of k-bit LSB for secret data embedding. The value of k is high at non-sensitive image region and over sensitive image area k value remain small to balance overall visual quality of image. The LSB’s (k) for embedding is computed by the high-order bits of the image. It also utilizes the pixel adjustment method for better stego-image visual quality through LSB substitution method. The overall result shows a good high hidden capacity, but dataset for experimental results are limited; there is not a single image which has many edges with noise region like ‘Baboon.tif’.

In [3] LSB based image hiding method. Common pattern bits (stego-key) are used to hide data. The LSB’s of the pixel are modified depending on the (stego-key) pattern bits and the secret message bits.
Pattern bits are combination of MxN size rows and columns (of a block) and with random key value. In embedding procedure, each pattern bit is matched with message bit, if satisfied it modifies the 2nd LSB bits of cover image otherwise remains the same. This technique targets to achieve security of hidden message in stego-image using a common pattern key. This proposed method has low hidden capacity because single secret bit requires a block of (MxN) pixels.

In [4] a Pixel value difference (PVD) and simple least significant bits scheme are used to achieve adaptive least significant bits data embedding. In pixel value differencing (PVD) where the size of the hidden data bits can be estimated by difference between the two consecutive pixels in cover image using simple relationship between two pixels. PVD method generally provides a good imperceptibility by calculating the difference of two consecutive pixels which determine the depth of the embedded bits. Proposed method hides large and adaptive k-LSB substitution at edge area of image and PVD for smooth region of image. So in this way the technique provide both larger capacity and high visual quality according to experimental results. This method is complex due to adaptive k generation for substitution of LSB.

In [5] a method of Multi-Pixel Differencing (MPD) which used more than two pixels to estimate smoothness of each pixel for data embedding and it calculate sum of difference value of four pixels block. For small difference value it uses the LSB otherwise for high difference value it uses MPD method for data embedding. Strength is its simplicity of algorithm but experimental dataset is too limited.

In [6] another pixel value differencing method, it used the three pixels for data embedding near the target pixel. It uses simple k-bit LSB method for secret data embedding where number of k-bit is estimated by near three pixels with high difference value. To retain better visual quality and high capacity it simply uses optimal pixel adjustment method on target pixels. Advantage of method is histogram of stego-image and cover-image is almost same, but dataset for experiments are too small.

In [7] have introduced a high capacity of hidden data utilizing the LSB and hybrid edge detection scheme. For edge computation two types of canny and fuzzy edges detection method applied and simple LSB substitution is used to embed the hidden data. This scheme is successful to embed data with higher peak signal to noise ratio (PSNR) with normal LSB based embedding. The proposed scheme is tested on limited images dataset. This method is not tested on extensive edges based image like ‘Baboob.tif’.

Madhu et al., in [8] proposed an image Steganography method, based on LSB substitution and selection of random pixel of required image area. This method is target to improve the security where password is added by LSB of pixels.

It generates the random numbers and selects the region of interest where secret message has to be hidden. The strength of method is its security of hidden message in stego-image, but has not considers any type of perceptual transparency.

In [9] an image steganographic method of mapping pixels to alphabetic letters. It maps the 32 letters (26 for English alphabetic and other for special characters) with the pixel values. Five (5) bits are required to represent these 32 letters and authors have generated a table where 4 cases design to represent these 32 letters. According to that table, each letter can be represented in all 4 cases. It utilizes the image 7 MSB (Most Significant Bits) (27 = 128) bits for mapping. Proposed method maps each 4-case from the 7 MSB’s of pixel to one of the 32-cases in that table. These 4-cases increase the probability of matching. This algorithm keeps the matching pattern of cover-image which is then used for extracting data from the stego-image. Proposed method does not required any edge or smoothness computations but secret data should be in the form of text or letter for embedding.

In [10], have introduced a data hiding technique where it finds out the dark area of the image to hide the data using LSB. It converts it to binary image and labels each object using 8 pixel connectivity schemes for hiding data bits. This method required high computation to find dark region its connectivity and has not tested on high texture type of image. Its hiding capacity totally depends on texture of image.

Babita et al., in [11] uses 4 LSB of each RGB channel to embed data bits, apply median filtering to enhance the quality of the stego image and then encode the difference of cover and stego image as key data. In decoding phase the stego-image is added with key data to extract the hidden data. It increases the complexity to applying filtering and also has to manage stego-key. Proposed scheme has high secret data hiding capacity.

In [12] a pixel indicator technique with variable bits; it chooses one channel among red, green and blue channels and embeds data into variable LSB of chosen channel. Intensity of the pixel decides the variable bits to embed into cover image. The channel selection criteria are sequential and the capacity depends on the cover image channel bits. Proposed method has almost same histogram of cover and stego-image.

Hamid et al., in [13] a texture based image Steganography. The texture analysis technique divides the texture areas into two groups, simple texture area and complex texture area. Simple texture is used to hide the 3-3-2 LSB (3 bits for Red, 3 bits for Green, 2 bits for Blue channels) method. On the other hand over complex texture area 4 LsB embedding technique is applied for information hiding. The above method used the both (2 to 4 LSB for each channel) methods depending on texture classification for better visual quality.
Proposed method has high hidden capacity with considering the perceptual transparency measures e.g. PSNR etc.

M. Chaumont et al., in [14] a DCT based data hiding method. It hides the color information in a compress gray-level image. It follows the color quantization, color ordering and the data hiding steps to achieve image Steganography. The purpose of method is to give free access to gray-level image to everyone but restricted access of same color images to those who have its stego-key. It has high PSNR plus with noticeable artifact of embedding data.

K. S. Babu et al., in [15] hiding secret information in image Steganography for authentication which is used to verify the integrity of the secret message from the Stegoimage. The original hidden message is first transformed from spatial domain to discrete wavelet transform (DWT); the coefficients of DWT are then permuted with the verification code and then embedded in the special domain of the cover image. The verification code is also computed by special coefficient of the DWT. So this method can verify each row of the image of modified or tampered by any attacker.

MODULES

Authentication Module:
In this project they are only two users one is sender and other one is receiver. Any user, who wants to the share and receive the data by using this project, must have to do registration in this project. After registration was done successfully they can login into this project by their username and password which they entered during registration process.

Send data request from client To sender:
Sender must share the data to receiver only after getting the request from receiver but receiver cannot send a request directly to the sender before that they have to get the connection path by entering the ip address of the sender. After getting the request, the sender can share the data to multiple receivers, that request contains the ip addresses of the receivers.

Image and data Steganography:
After getting request from the receiver only the sender can share the data, first the sender will select the data and then he will encrypt the data. While doing this process the user can see the original data and converted data. After converting the data the sender will hide the data into the image to make the content more secure. Finally the hidden data will to shared to the receivers who send request.

Receive data and splitting data from image:
After data was shared from the sender, the sender can view the path of the data how it was forwarding from the sender to the receiver. Once the path was completed the receiver can see the auto splitting of data and image, where the hidden by the sender to make the data more secure. Once the splitting of data and image was completed the data will be stored automatically.

Image Grey Scale and Encrypt:
This is the second one; in this the sender can share the image alone without doing any hiding process. In this before share the image, the sender will change the normal image into gray scale image and then that gray scale image will be converted in to some text format to make the transmission more secure. Then the converted text only share to the users who asked for the image.

Text and Voice Intimation:
This is the third one, in this the sender can share the text document alone without doing any hiding process. In this before share the data, the sender will change the normal data in to some other text format to make the transmission more secure. Then the converted text only shared to the users who asked for the data. And also there will be voice intimation for every sharing of datas.

Compress and View Transmission Path:
In this module, the sender will compress each and every data which he wants to share to other users. For example when the sender hide the data inside the image means the file sized will be increased, to reduce this file size the sender can compress the file so the size of the file will be reduced. In this same way, the user can compress the data and image alone which he wants to share. And also he can we the transmission path of the files between the sender and receiver.

EXPERIMENTS RESULTS

5.1. Cover model estimation and embedding
For a given relative payload _ and grayscale image x = {xi}, xi 2 {0, 255}, the sender first computes the costs I using (14). Even though pixel values are not realizations of independent zero-mean Gaussians, the pixels are locally strongly correlated. Assuming that Xi from a small (e.g., 3 x 3) neighborhood Ni have the same mean and variance, the variance vi of Xi can be estimated as vi = max{1,ENi[x2 i] − (ENi[xi])2}, (15) where ENi is the sample mean over Ni. The maximum with 1 was added for numerical stability. The actual embedding was simulated at the rate-distortion bound both for the MG algorithm and HUGO, which we included for comparison as the current state-of-the-art algorithm for images in raster format as of November 2012. In practice, the ternary version of STCs could be used to implement the actual embedding algorithm near its payload-distortion bound.

5.2. Steganalysis
To see how the detect ability increases with increased payload, steganalysis was carried out using supervised machine learning by building a binary classifier for the class of cover images and stego images embedded with a fixed relative payload. Images were represented using the state-of-the-art spatial rich model (SRM) [13] with q = 1 with total dimensionality 12,753. The machine learning was the ensemble [18] run with default settings with the Fisher linear discriminant as the base learner. The cover source was the BOSS base 1.01 [19] containing 10,000 grayscale
512 × 512 images originally obtained by seven cameras in raw format (DNG, CR2), demosaicked, and resized/cropped using a script also available online. The detection performance was evaluated in a standard manner using the minimal total error under equal prior probabilities of both hypotheses, \( PE = \min PFA \ (PFA+PMD) \), averaged over ten random splits of the database into two halves.

**Fig. 1.** Test image (left) and the embedding changes displayed in white when embedding relative payload 0.4 bpp (right).

Here selecting an image location we can select our own image file and uplink it and we have to assign the output Location so that the input file will be created there in the drive.

The input is selected from the image source as the ISO image file so that the encrypted data will be hidden in the ISO image. Here large amount of data will be compressed and encrypted as well as decrypted.

The content which is going to be encrypted will be processed in this step. For more security purpose the administrator will set a manual password for authentication.

**CONCLUSION**

The hurried development of multimedia and internet allows wide distribution of digital media data. It becomes much easier to edit, modify and duplicate digital information. In additional, digital document is also easy to copy and distribute, therefore it may face many threats. It became necessary to find an appropriate protection due to the significance, accuracy and sensitivity of the information.

Nowadays, protection system can be classified into more specific as hiding information (Steganography) or encryption information (Cryptography) or a combination between them. Cryptography is the practice of ‘scrambling’ messages so that even if detected, they are very difficult to decipher. The purpose of Steganography is to conceal the message such that the very existence of the hidden is ‘camouflaged’.
However, the two techniques are not mutually exclusive. Steganography and Cryptography are in fact complementary techniques. No matter how strong algorithm, if an encrypted message is discovered, it will be subject to cryptanalysis. Likewise, no matter how well concealed a message is, it is always possible that it will be discovered. By combining Steganography with Cryptography we can conceal the existence of an encrypted message. In doing this, we make it far less likely that an encrypted message will be found. Also, if a message concealed through Steganography is discovered, the discoverer is still faced with the formidable task of deciphering it. Also the strength of the combination between hiding and encryption science is due to the nonexistence of standard algorithms to be used in (hiding and encryption) secret messages. Also there is randomness in hiding methods such as combining several media (covers) with different methods to pass a secret message. Furthermore, there is no formal method to be followed to discover a hidden data.

REFERENCES


