A Review on Image Steganalysis techniques for attacking Steganography

Shamim Ahmed Laskar[1] and Kattamanchi Hemachandran[2]
Department of Computer Science
Assam University, Silchar, Assam, India

Abstract

Steganography is a data hiding technique that embeds the secret message inside a digital media for providing a method of invisible communication. All digital file formats can be used for steganography of which digital images are the most popular because of its massive presence in the internet. For hiding secret information in images, there exist a large variety of steganographic techniques. Several efforts are made to establish ways of detecting whether or not an image contains a steganographic element. Steganalysis is the technique of detecting the presence of steganography that can serve as an effective way to judge the security performance of steganographic techniques. In this paper, we provide an overview of digital image steganalysis technique for detecting steganographic method and identify where to look for hidden information. These techniques are analyzed and discussed in terms of their ability to detect secret message in an image file. We have also reviewed some attacks on steganography.

Keywords: *Image steganography, attacks, steganalysis, target steganalysis, blind steganalysis*

1. Introduction

With the growth in exchange of digital data through network, the security of data became a matter of concern across the globe. The distribution of digital data raised a concern as the data are attacked and manipulated by unauthorized person. The Internet provides a method of communication to distribute information and thus approach of hiding secret message in different multimedia is increased [1]. Data hiding techniques are increasing day by day with more effective approach. Steganography is a data hiding technique aiming to transmit a message on a channel, where some other information is already being transmitted. The goal of steganography is to hide messages inside digital media in order to avoid drawing suspicion of the hidden data from a third party. It includes a vast array of secret communications methods that conceal the message's very existence [2].

Steganography does not alter the structure of the secret message, but hides it inside a medium so it cannot be seen. The technique replaces unused or redundant bits of the digital media with the secret data. The redundant bits are those bits that can be altered without drawing suspicion. The concept is to embed the hidden object into a significantly larger object so that the change is undetectable by the human eye. The security of the steganography depends on the data encoding system. Once the encoding system is known, the steganography system is defeated. Steganography programs allow the user to select a carrier that they wish to use as the vector to carry the hidden data [3]. All digital file formats can be used for steganography, but those formats that have a high degree of redundancy are more suitable. The digital media which are used for secret communication includes text, images, audio and videos which provide excellent carriers for hidden information. The most popular cover objects used for steganography are digital images. Digital images often have a large amount of redundant data, and this is what steganography uses to hide the message. A steganographic system involves two parties, the sender who encodes the secret message in the cover medium and the receiver who extracts the message from the cover. Steganography are mainly applicable in the field of computer security, cyber-security, digital forensics, copyright protection, feature tagging, secret communication, watermarking, intelligence agencies, military agencies etc.

The most important requirements for steganographic system are the security (undetectable) and capacity [1]. As steganography is emerging increasingly in the field of data protection method, simultaneously techniques of detecting secret data inside digital file are also emerging. The security of a steganographic system is defined by its ability to overcome from attacks. The attacks are based on the statistical and structural properties of image file. The method to detect the presence of steganography is called steganalysis. Most steganographic techniques involve changing properties of the cover source and there are several ways of

detecting these changes. While steganography deals hiding secret data, the role of steganalysis is to detect and estimate hidden data inside digital file from observed data with little or no knowledge about the steganography method and/or its parameters.

2. Image Steganography Techniques

- **Transform Domain Based Steganographic Techniques**

Transform domain based steganography techniques take advantage of Discrete Cosine Transformation (DCT) used in the JPEG compression process [2]. In transform domain technique image is first transformed into the frequency domain and then the message is embedded in the discrete coefficients and finally the inverse transform is performed to restore the image. JPEG is a lossy digital image compression system which reduces image's file size to a great extent. Firstly, an image is converted to a YUV representation space, breaks up each color plane into 8 x 8 blocks of pixels blocks and then each of these 8 x 8 pixel blocks are transformed into an array of 64 DCT coefficients [4]. Next a quantizer rounds each of these coefficients. Then an array of streamlined coefficients is formed, which are further compressed via a Huffman encoding scheme or similar. The quantized DCT coefficients are then used to embed hidden message [5]. Decompression is done via an inverse DCT.

- **Spatial Domain Based Steganographic Techniques**

Spatial domain based method modifies the secret data and the cover medium in the spatial domain, where the secret message may be hidden by altering least significant bit of an image file. LSB steganography is the most classic and simplest steganographic techniques, which embeds secret messages in a subset of the LSB plane of the image [2]. The main advantage of such a technique is that the modification of the LSB plane does not affect the human perception of the overall image quality. Spatial-domain technique embeds messages in the intensity of pixels of images directly. Thus the advantages of spatial domain technique is that it is clean and quite simple mathematical derivation and the possibility to engage statistical image models for obtaining upper bounds on the performance of the message length estimator. Spatial domain method does not need any transformation or decomposition on the original images. The method is simple and fit for real-time processing.

- **Palette based Steganographic Methods**

Palette based or Indexed colors image that enables 8 bits per pixel determines the 256 most frequently used colors in the image and creates a color lookup table, also called a color map or color palette, which is stored with the image [6]. This avoids changes in the image leaving the visual perception of the image unscathed. This is possible because two identical images may have completely different color-maps. It is natural that the information can also be embedded using pre-selected pixels and arranging the palette in a structure where neighboring colors are close in given distance, including Chroma Difference [7]. Palette-based images are used as cover images to provide a secure and fast transmission/storage over a communication system. Furthermore, since there was some alteration introduced by the color quantization, the stego message is able to pass as noise. If palette based images have a smaller resolution, they can be transmitted faster than 24-bit resolution images [7]. Palette based steganography technique mainly deals with GIF or BMP images and performs better when data are embedded in a grayscale image.

- **Wavelet Transformation in Steganography**

Wavelet transform is used to convert a spatial domain into frequency domain. The use of wavelet in image stenographic model lies in the fact that the wavelet transform clearly separates the high frequency and low frequency information on a pixel by pixel basis [8]. A one dimensional DWT is a repeated filter bank algorithm, and the input is convolved with high pass filter and a low pass filter. The simplest form of discrete wavelet transform (DWT) is Haar-DWT in which the low frequency wavelet coefficient are generated by averaging the two pixel values and high frequency coefficients are generated by taking half of the difference of the same two pixels [9]. The operation for Haar DWT has been applied to image processing especially in multi-resolution representation.

3. Steganalysis

Steganalysis is the art and science to detect whether a given digital medium contains hidden data. The art of steganalysis plays a greater role in the selection of features or properties of digital data to test for hidden data while the science helps in designing of the technique to detect or extract tests the hidden data. A steganalysis method is considered as successful if it can detect and extract the hidden data embedded [10].The objective of steganalysis is to detect messages hidden in cover objects, such as digital images [25]. Steganalysis can be termed as a method of attacking the

digital media for estimating whether the media contains secret data embedded in it. Steganalysis can serve as an effective way to judge the security performance of steganographic techniques. A good steganographic technique should be imperceptible not only to human eye, but also to computer analysis. The steganalyst (one who performs steganalysis) is assumed to control the process of transmission channel and trace out for suspicious data. In practice, the steganalyst is frequently interested in more than whether or not a secret message is present [25]. However, an original cover medium and its stego-version (with hidden message inside) always differ from each other in some aspects since the cover medium is modified during the data embedding. Thus the general steganalysis methods can find way for detecting secret communications.

The objectives of steganalysis are:

- Its objective is to detect the existence of a secret message in a binary image. The suspect image may or may not have hidden data encoded into them.
- To evaluate techniques that can be used to distinguish the images hidden with secret messages from those without. Some of the suspect image may have noise or irrelevant data encoded into them.
- Its purpose is to identify the type of steganographic method used to create the stego image. It tries to understand the internal mechanism used during the embedding operation.
- The steganalysis technique is used not only to detect the stego-image but it tries to recover the hidden data.
- The steganalysis technique also tends to estimate the length of the hidden message, and to trace the locations of the pixel bearing the hidden message.
- Steganalysis is designed to estimate the relative numbers of embedding changes in the digital image.

Types of Attacks

While the purpose of Steganography is to hide messages, there exist several attacks that one may execute to test for Steganographic data. The strength of a steganographic algorithm depends on its ability to successfully withstand attacks. Attacks and analysis of hidden data may take several forms: detecting, extracting, and disabling or destroying hidden data. An attack approach is dependent on what information is available to the steganalyst. A large number of attacks are implemented in steganography to evaluate see if digital media contains hidden data. Attacking steganographic algorithm is very similar to attacking cryptographic algorithms and similar techniques apply [11]. As Fabien A.P. Petitcolas [10] points out that there are six general protocols used to attack the use of Steganography.

Classification of attacks based on information available to the attacker [10]

Stego-only attack: Only the steganography medium/object is available for analysis.

Known-carrier attack: The carrier, that is, the original cover, and steganography media/object are both available for analysis or are known.

Known-message attack: In this case, the hidden message is known and can be compared with the stego-object/medium.

Chosen-stego attack: The steganography medium/object and tool (algorithm) are both available for analysis.

Chosen-message attack: Here a chosen message and steganography tool (or algorithm) is used to create steganography media for future analysis and comparison. The goal in this attack is to determine corresponding patterns in the steganography medium.

Known-steganography attack: The secret message, steganography medium/object and the steganography tool (algorithm) are known and available for analysis.

Steganography elimination techniques is involved with steganalysis that try to eliminate or destroying the hidden information as the purpose is to break the cover communication, but not to render the legitimate channel unusable. The most common attacks based on this factor are [11]:

Destroy everything attack –this type of attack aims in destroying the message completely and the attacker might not even try to retrieve the message.

Random tweaking attacks – here small changes in the files are added so that the message will be unreadable.

Add new Information – in some cases the attackers might use the same technique of data hiding to embed a new message into the stego-file. The original message might be overwritten.

Reformat attack – a common way to destroy the information hidden in a file is by changing the file format. This type of attack can produce a lot of damages to the hidden message.

Compression attack – the attacker might compress the file which might result in the total loss of the secret message embedded in the file, because the compression algorithms tend to remove extra in-formation during compression.

From the above, it is clear that if an attacker wants only to destroy the hidden message, he can do that very

easily by combining some of the methods of attack presented above.

4. Image based Steganalysis Techniques

Steganalysis can be broadly divided into targeted method and blind method [22]. A targeted steganalysis uses the knowledge about the steganographic technique to detect stego-objects created with that specific technique, while blind steganalysis aims to distinguish if a file has some hidden information with no information about the used steganographic technique. Blind and targeted steganalysis techniques have been greatly studied on digital images [12].

4.1. Targeted steganalysis

Targeted steganalysis are designed to evaluate mechanisms of particular embedding operations and fully utilizes the knowledge applicable to detect steganography. It concentrates on detecting some specific steganographic tool and has good performance on this steganographic tool if well designed [14]. In the case of a known algorithm, an attack that works for that specific algorithm is called targeted steganalysis attack. A targeted steganalysis technique works on a specific type of stego-system and sometimes limited on image format. By studying and analysing the embedding algorithm, one can find image statistics that change after embedding. The results from most targeted steganalysis techniques are very accurate but on the other hand, the techniques are inflexible since most of the time there is no path to extend them to other embedding algorithms. Also, when a targeted steganalysis is successful, thus having a higher probability than random guessing, it helps the steganographic techniques to expand and become more secure. A targeted Steganalysis can be of three types- Visual, Statistical and Structural attacks [22].

4.1.1. Visual attacks

Visual Attacks are simplest form of steganalysis that involves examining the stego-image with the naked eye to identify any kind of degradation. The steganographic method normally does not leave any kind of visual distortion on the image file due to modification of bits. However, when the parts of the image those were not altered as a result of embedding are removed, it is usually possible to observe signs of manipulation [22]. The human sight is trained to recognise known things and capable of identifying changes, such ability is used for the visual attacks [21]. The visual attack enables humans to distinguish between noise and visual patterns.

Visual Attacks can be implemented in many ways by inspecting different properties of the image. A visual attack could be established to display the spatial domain of the image by verifying its LSB. An image typically contains as many 1's as there are 0's in its least plane. Whereas text often have more 0's than 1's, and this produces a visual inconsistency. A steganalyst searches for such inconsistency in order to classify an image either as a stego-image or as normal image. Although such inconsistencies depend on way the data is embedded in the cover-image. Similarly, a steganalyst could also attack in the transform domain to evaluate whether or not the image contains signs of transform embedding. In addition to this, it is also possible to attack various image formats, (such as JPEG, GIF, BMP, PNG etc.) that can be used as an investigative tool against embedding approaches [21]. When all possible approaches are used, the steganalyst can determine whether or not the image is subjected to steganographic embedding.

In case of sequence embedding when a visual attack is applied on a suspected image, the steganalyst looks for visual distortions for the first n pixels where n is the length of the message. However, the exact value for n is not made apparent until the visual attack successfully indicates signs of embedding. On the other hand, it is much harder to perform a visual attack on randomized embedding as the data are embedded in the random pixels of an image. So it becomes much difficult in identifying the regions that have been altered as a result of random embedding. Thus visual attack does not reflect the modified regions clearly when the data are embedded in random pixel locations. Visual attacks on random embedding differ from sequential embedding in way that the steganalyst must have access to the cover-image image in order to compare it with the stego-image and is referred to as known cover attack. Then the steganalyst can obtain the distorted points by comparing and then calculating the difference between the two by subtracting one from the other.

Visual attacks can be a useful tool for steganalysis - at least for detecting steganography from the most basic steganographic implementations, or for known cover attacks. When the cover image is not available to the steganalyst, visual attack depends on three factors holding true for successful attack. The message must be embedded in a sequential order, its length must be less than the maximum size of the bit plane and it should not be encrypted. In order to detect a visual distortions within a suspect image visual attack rely on the fact that the message length is shorter than the full embedding capacity of that bit plane. It is no longer possible to see a change in form in the bit plane. Also, when a message is encrypted it can reduce the chance

of success for a visual attack by considerable proportions when the cover-image is not available.

A successful visual attack not only allows to detect inconsistencies within an image, it also investigates how the stego-system operated (i.e. sequential/ randomised embedding), and allows for an estimation of the message length. However, with research by Wayner [11], and Westfeld [21] suggesting that every bit plane is non-random, perhaps it could be possible to broaden these findings, and develop a more structured and logical approach to the decision making process. It is essential for a visual attack to determine appropriately the features of the image that can be ignored and those features that can be taken into consideration implementing a valued attack. Because a wrong guess can cause false-negatives [22]. The success of visual attack varies significantly depending on the steganographic method applied and the format of the image. As the attack can be applied in different embedding technique so examining properties of several image formats is not sufficient. The cover-image or the steganographic method is required in order to detect the distorted regions. Thus it is proves time-consuming in testing images for various methods of embedding. This is obviously an inefficient methodology and so other steganalysis methods are preferred. The main drawback with visual attack is the fact that it cannot be automated.

4.1.2. Statistical Attacks

In this type of attacks the statistical analysis of the images by some mathematical formula detects the presence of hidden data. Statistical attack is partially similar to visual attack. Generally the hidden message is more random than the original data of the image thus finding the formula to know the randomness reveals the existence of data [11]. A theory would be constructed that seemingly explains why the phenomenon occurs, and statistical methods can then be used to prove this theory to be either true or false. If the data structure is considered for a stego-image, then the statistics can be useful for steganalysis when proving whether or not the image contains a hidden message. Statistical tests can reveal that an image has been modified by steganography by determining that an image's statistical properties deviate from a norm. Some tests are independent of the data format and just measure the entropy of the redundant data [18]. The simplest test measures the correlation towards one and is not able to decide automatically if an image contains a hidden message [18]. There are several methods that are known to prove the existence of a hidden message via statistical approaches; each aimed at identifying signs of embedding for specific stego-systems. Here we discuss one of the most effective statistical attacks, as well as providing details of its effectiveness. The technique that belongs to Statistical attack is Chi-square Analysis.

- **Chi-square Analysis**

Westfeld and Pfitzmann outline an interesting statistical attack [21]. They observed that for a given image, the embedding of encrypted data changes the histogram of color frequencies in a particular way. In their case, the embedding process changes the least significant bits of the colors in an image. The colors are addressed by their indices in the color table. If n_i and n_i* are the frequencies of the color indices before and after the embedding respectively, then the following relation is likely to hold

$$|n_{2i} - n_{2i+1}| \geq |n_{2i}^* - n_{2i+1}^*| \quad (1)$$

In other words, the frequency difference between adjacent colors is reduced by the embedding process. In an encrypted message, zeros and ones are equally distributed. For $n_{2i} > n_{2i+1}$, the bits of the hidden message change n_{2i} to n_{2i+1} more frequently than the other way around.

Westfeld and Pfitzmann use a χ^2-test to determine whether the color frequency distribution in an image matches a distribution that shows distortion from embedding hidden data. In the following, Niels Provos and Peter Honeyman discussed that it applies to the DCT coefficients JPEG format [1].Westfeld and Pfitzmann test uses only the stego medium, the expected distribution y_i* for the χ^2-test has to be computed from the image. The assumption for a modified image is that adjacent DCT frequencies are similar. Let n_i be the DCT histogram, then take the arithmetic mean is taken as,

$$yi^* = (n_{2i} + n_{2i+1})/2 \quad (2)$$

to determine the expected distribution and compare against the observed distribution $y_i = n_{2i}$.

The χ^2 value for the difference between the distributions is given as

$$\chi^2 = \sum_{i=1}^{v+1} \frac{(y_i - y_i^*)^2}{y_i^*}, \quad (3)$$

where v are the degrees of freedom, that is, the number of different categories in the histogram minus one. It may be necessary to sum adjacent values from the expected distribution and from the observed distribution to ensure that there are enough counts in each category. Westfeld and Pfitzmann require that

each count is greater than four. If two adjacent categories are summed together, the degrees of freedoms need to be reduced by one.
The probability p that the two distributions are equal is given by the complement of the cumulative distribution function,

$$p = 1 - \int_0^{\chi^2} \frac{t^{(v-2)/2}\, e^{-t/2}}{2^{v/2}\Gamma(v/2)}\, dt \qquad (4)$$

where Γ is the Euler Gamma function.
The probability of embedding is determined by calculating p for a sample from the DCT coefficients. The samples start at the beginning of the image and for each measurement the sample size is increased. Because the test uses an increasing sample size and always starts at the beginning of the image, it detects changes only if the frequency histogram is distorted continuously from the beginning of the image.

According to Niels Provos and Peter Honeyman it is possible to extend Westfeld and Pfitzmann's χ^2-test to be more sensitive to partial distortions in an image [17, 18]. Observe that two identical distributions produce about the same χ^2 values in any part of the distribution. Instead of increasing the sample size and applying the test at a constant position, they used a constant sample size but slide the position where the samples are taken over the entire range of the image.

The test starts at the beginning of the image, and the position is incremented by one percent for every χ^2 application. This extended test does not react to an unmodified image, but detects the embedding in some areas of the stego image. To find an appropriate sample size, they choose an expected distribution for the extended χ^2-test that should cause a negative test result. Instead of calculating the arithmetic mean of coefficients and their adjacent ones, they take the arithmetic mean of two unrelated coefficients,

$$yi^* = (n_{2i-1} + n_{2i})/2 \qquad (5)$$

A binary search on the sample size is used to find a value for which the extended χ^2-test does not show a correlation to the expected distribution derived from unrelated coefficients.

4.1.3. Structural Attacks

Structural attacks are designed to take advantage of the high-level properties that are known to exist for a particular steganographic algorithm [22]. Steganographic methods leave behind a characteristic structure to the data. The attacker may detect the existence of secret message by examining the statistical profile of the bits or by identifying these characteristic structure changes. Structural attacks rarely analyse each image on its own merits. Instead, the images are scanned to see if they contain any of the known side-effects for various steganographic algorithms. Images that contain these properties are often subjected to further investigation. There are some cases where the image may possess signs of steganography when it may be perfectly innocent. This is why structural attack usually follows a more thorough investigation. Thus a steganalyst will typically need to obtain a better range of features that represent suspicious images. A common element of structural detectors is to estimate features so that macroscopic cover properties can be approximated from the stego object by inverting the effects of embedding as a function of features so that it matches cover assumptions best [25]. A successful structural attack relies on being able to identify a distinct difference between the cover-image and a stego-image, which means that there is a heavy reliance on either knowing the cover-image or knowing the embedding details of the steganographic algorithm and evaluating the consequences of the embedding strategy. It is rarely the case that a steganalyst will have access to one of these, and even rarer for them to have access to both, which only hampers the success of the attack.

Structural attacks are arguably more important to steganalysts than visual attacks because they can be applied against a wider range of embedding techniques. The attacks work best when the steganalyst has access to a known stego-image. Structural attacks are not used as a means of proving that an image contains steganography, rather it investigates whether an image contains signs of embedding which allows steganalyst to conclude with a higher confidence whether or not the image has been tampered. RS analysis and Pair analysis represents structural attack.

- **RS Analysis**

RS steganalysis is able to estimate the length of the embedded message on a digital image, for LSB steganographic methods [12]. RS steganalysis was introduced by Fridrich et al. for exploiting the correlation of images in the spatial domain [12, 25].The method is based on the fact that the content of each bit plane of an image is correlated with the remaining bit planes. The technique is based on analyzing how the number of R and S groups of pixels changes with the increase in message length embedded in LSB plane. The lossless capacity reflects the fact that the LSB plane – even though it looks random – is related to the other bit planes. Randomizing the LSBs decreases this capacity. To examine an image, they define Regular groups (R) and Singular groups (S) of pixels depending upon some properties. Then with the help of relative

frequencies of these groups in the given image, in the image obtained from the original image with LSBs flipped and an image obtained by randomizing LSBs of the original image, they try to predict the levels of embedding.

If I be the image to be analyzed with width W and height H pixels. Each pixel has values in P. For an 8 bits per pixel image, P = {0 . . . 255}. Then I is divided into G disjoint groups of n adjacent pixels. For instance, if n = 4 adjacent pixels can be chosen. Then a discriminant function f responsible to give a real number f(x1, . . . , xn) ∈ ℜ for each group of pixels G = (x1, . . . , xn) is defined. The objective of using f is to capture the smoothness of G. The discrimination function is given by

$$f(x_1, \ldots\ldots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|, \qquad (6)$$

The noisier the group G, the larger the value of the discrimination function becomes. The LSB embedding increases the noisiness in the image, and thus they expect the value of f to increase after LSB embedding.

They defined an invertible operation F on P called "flipping" which is a permutation of gray levels. The permutation F1: 0 ↔1, 2 ↔3, …, 254 ↔255 corresponds to flipping (negating) the LSB of each gray level. They further define shifted LSB flipping F−1 as −1 ↔ 0, 1 ↔ 2, 3 ↔ 4, …, 253↔254, 255 ↔256, or

$$F_{-1}(x) = F1(x + 1) - 1 \text{ for all } x \qquad (7)$$

For completeness, they also defined *F*0 as the identity permutation $F(x) = x$ for all $x \in P$.

A mask M is defined that represents which function to apply to each element of a group G. The mask M is an n-tuple with values in {−1, 0, 1}. The value -1 stands for the application of the function F−1; 1 stands for the function F1; and 0 stands for the identity function F0. Similarly, –M is defined as M's complement.

The number of regular groups for mask *M is denoted* as R_M (percent of all groups). Similarly, S_M and U_M denotes the relative number of singular and Unusable groups respectively. Then the discriminant function f is applied with the functions F {−1,0,1} defined through a Mask M over all G groups to classify them into three types of pixel groups—*R*, *S*, and *U*:

• Regular. $G \in R_M \Leftrightarrow f(F_M(G)) > f(G)$
• Singular. $G \in S_M \Leftrightarrow f(F_M(G)) < f(G)$
• Unusable. $G \in U_M \Leftrightarrow f(F_M(G)) = f(G)$

In these expressions, *F* (*G*) means that apply the flipping function *F is applied* to the components of the vector $G = (x1, \ldots, xn)$.

Then the received image is divided into small blocks of the same size. Define R_M as the ratio of blocks in which *f* increases when F_1 is applied to a part of each block, and S_M as the ratio of blocks with decreasing *f*, thus, $R_M + S_M \leq 1$. Similarly, another two parameters R_{-M} and S_{-M} can be defined when F_{-1} is applied to a part of each block and thus $R_{-M} + S_{-M} \leq 1$.

In typical images, applying the LSB flipping mask to the pixels in the group will more frequently lead in an increase in the discrimination function, rather than a decrease, and thus the total number of regular groups in an image will be larger than singular groups. If the received image does not contain secret data, F_1 and F_{-1} should equally increase the *f* value of blocks in a statistical manner. So,

$$R_M \approx R_{-M} > S_M \approx S_{-M} \qquad (8)$$

When a message with a relative length p is embedded in the cover image ($p = 1$ for full length embedding), the fraction of image pixels with the LSB flipped is, on average, $p/2$. Flipping the LSB of all image pixels will result in an image with a fraction of flipped pixels $1 - p/2$. In the process of steganalysis of an image, the actual value of p is unknown. The relative number of R and S groups is counted for the original image, and for the flipped version of that image that are used to estimate the value of p.

Thus when secret bits are embedded, the difference between R_M and S_M decreases whereas the difference between R_{-M} and S_{-M} increases. Thus,

$$R_{-M} - S_{-M} > R_M - S_M \qquad (9)$$

Therefore, an attacker can use the relation among the four parameters to detect the presence of secret information [12].

- **Pairs analysis**

Fridrich et al. proposed another steganalysis technique that detects the data hidden in palette images [25, 27]. Pairs Analysis is a steganalysis process that detects messages embedded in palette images as LSBs of indices to palette colors along a random walk. The principle of Pairs Analysis is based on the color pair. Let (*c*1, *c*2) be a color pair from the set of color pairs, *E where*

$$E = \{(c\pi(0), c\pi(1)), (c\pi(2), c\pi(3)), \ldots, (c\pi(P-2), c\pi(P-1))\} \qquad (10)$$

Here the colors $c1$, $c2$ are extracted from the whole image for example by scanning it by rows and/or columns. This sequence of colors is then converted to a binary vector by associating $c1$ with a "0" and $c2$ with a "1". This vector is denoted by say, $Z(c1, c2)$ and call it the *color cut* for the pair $(c1, c2)$. As palette images have a small number of colors, Z will show considerable structure for most pairs $(c1, c2)$. The embedding process will disturb this structure and increase the entropy of Z. The entropy of Z will be maximal corresponding to a random binary sequence. During embedding to color cuts for "shifted" color pairs, another set of color pairs say E' are formed by shifting the pairs. Thus

$$E' = \{(c\pi(1), c\pi(2)), (c\pi(3), c\pi(4)), \ldots, (c\pi(P-1), c\pi(0))\} \quad (11)$$

In the detection method using pair analysis, the color cuts are concatenated for all pairs in E into one vector $Z = Z(c\pi(0), c\pi(1)) \,\&\, Z(c\pi(2), c\pi(3)) \,\&\, \ldots \,\&\, Z(c\pi(P-2), c\pi(P-1))$ and all color cuts for shifted pairs E' into the vector $Z' = Z(c\pi(1), c\pi(2)) \,\&\, Z(c\pi(3), c\pi(4)) \,\&\, \ldots \,\&\, Z(c\pi(P-1), c\pi(0))$. Next, the quantity that will be used to measure the structure in the bit-streams Z and Z' is defined and the number of "homogenous" bit-pairs ('00', '11') in Z is calculated. Let $R(p)$ denote the expected relative number of homogeneous bit-pairs in Z after flipping of $p\times100\%$ of chosen pixels. Similarly, let $R'(p)$ be the relative number of homogeneous bit-pairs in Z'.

As the authors claimed that $R(p)$ is a parabola with its vertex at $p = ½$ and $R(1/2) = ½$. Because with an unknown message length q, $R(q)$ can be calculated from the stego image and the points $(q, R(q))$ and $(1/2, R(1/2))$ uniquely determine $R(p)$. Here $R(q) = R(1-q)$. They stated that $R'(p)$ is well modeled using a parabola, as well. The value of $R'(1/2)$ can be derived from Z', while the values $R'(q)$ and $R'(1-q)$ can be calculated from the stego image and the stego image with all colors flipped, respectively. Thus, a second-degree polynomial can be fitted through the points $(q, R'(q))$, $(1/2, R'(1/2))$, and $(q, R'(1-q))$ to obtain $R'(p)$.

Finally, another assumption $R(0) = R'(0)$ is stated, which says that the number of homogenous pairs in Z and Z' must be the same if no message has been embedded. Denoting $D(p) = R(p) - R'(p) = ap^2 + bp + c$, with a, b, and c yet undetermined constants, then $D(0) = c = 0$ and $D(1/2) = R(1/2) - R'(1/2) = a/4 + b/2$ can be written. Also, $D(q) = aq^2 + bq$ and $D(1-q) = a(1-q)^2 + b(1-q)$. Eliminating the unknown parameters a, b leads after simple algebra to the following quadratic equation for q:

$$4D(1/2)q^2 + [D(1-q) - D(q) - 4D(1/2)]q + D(q) = 0 \quad (12)$$

Because the coefficients of this quadratic equation are known, it can be solved for the unknown q. The root that is smaller is our approximation to the unknown message length q [25].

4.2. Blind Steganalysis

Blind steganalysis is an approach is needed that is capable of identifying the probability of embedding, even when it is not sure how the information might have been embedded. Blind steganalysis do not require prior knowledge about details of the embedding operations [13, 14]. Without a specific embedding knowledge, blind steganalysis extract from suspected stego objects a broad set of general statistical measures, which likely change after embedding [14].It tries to detect any steganographic tool, known or unknown in advance. Both sets of statistical moments are used as features for steganalysis. Blind steganalysis therefore works differently to targeted steganalysis because it assumes that nothing is known about either the algorithm or the cover image that was used to produce a suspect image. The attacks attempt to evaluate the probability of embedding based solely on the data of the suspect image. Such approaches are more common in real-world steganalysis because a steganalyst will rarely know much more about an image than what they can extract from the image itself.

Blind identification methods pose the steganalysis problem as a system identification problem and the embedding algorithm is represented as a channel and the goal is to invert this channel to identify the hidden message [19]. In this steganalysis method, each image is analyzed individually based on the computed statistics. Thus it is possible to derive analytical results that suggest the feasibility of successful steganalysis for certain types of statistical models for the original image and the secret message [20]. Digital images are known to be statistically non-stationary and this causes practical issues in implementing algorithms based on the blind identification model, as it assumes stationary of data. When the stationary condition is violated additional effort is needed to make steganalysis work. If the message embedding algorithm is nonlinear then the blind identification problem becomes more difficult [20]. Perhaps the most important aspect of blind steganalysis is ensuring that an estimate of the cover image can be derived that is as accurate as possible. The attacks that follow this procedure often compare the data in the estimated cover image to that of the suspect image. A blind steganalysis technique is expected to work on all types of embedding techniques and image formats.

Some of the methods of blind steganalysis schemes are:

1. Self-calibration mechanism [26]: Calibration process is used by the blind steganalysis schemes to estimate the statistics of the cover image from the stego image. For JPEG steganography, this is typically achieved by decompressing the stego image to the spatial domain followed by cropping the image by a few pixels on each side and compressing the image again using the same compression parameters.
2. Features capturing cover memory [15, 16]: Most steganographic schemes hide data on a per-symbol basis, and typically do not explicitly compensate or preserve statistical dependencies. Hence, features that capture higher dimensional dependencies in the cover symbols are crucial in detecting the embedding changes. Cover memory has been shown to be very important to steganalysis and is incorporated into the feature vector in several ways.
3. Supervised learning based steganalysis [23, 24]: Supervised learning based steganalysis techniques employ two phase strategies: (a) training phase and (b) testing phase. In the training phase it constructs a classifier to differentiate between stego and non-stego images using training examples. The learning classifier iteratively updates its classification rule based on its prediction and the ground truth. In the testing phase unknown images are given as input to the trained classifier to decide whether a secret message is present or not. However the choice of proper features to train the classifier is a critical step. If the selected features are not appropriate for the specific embedding algorithm then the detector may completely fail. There is no systematic rule for feature and parameter selection. It is extremely difficult or even impossible to identify portions of the image where a message is hidden.

The calibration process is perhaps the most important of the above that allows the steganalyst to get an accurate underlying model of the cover image without access to it. With this approach, statistical steganalysis is successful even though good universal statistical models for images are not available.

4.2.1. JPEG Calibration

JPEG Calibration is a technique specific to steganalysis based on histograms for creating an estimate of the cover image proposed by Fridrich et. al [26]. In the following, we discuss the method of calibration proposed by them. The method takes advantage of the fact that most stego-systems encode the message data in the transform domain during the compression procedure to produce JPEG stego-image. The idea of *calibration* is to estimate marginal statistics (histograms, co-occurrence matrices) of the *cover*'s transformed domain coefficients from the *stego* object by desynchronising the block transform structure in the spatial domain. The calibration is done by taking the feature differences of the cropped image with the original image. A stego-image in transformed domain representation in transformed domain representation is first converted to spatial domain then it is cropped by a small number of pixels at two orthogonal margins (e.g. cropped by 4 rows and 4 columns) and then re-encoded in the JPEG format. This ensures that the original (8×8) block is desynchronised in a subsequent transformation to the transform domain. The idea of calibration is to estimate the cover image from the stego-image. By cropping and re-encoding the image, a smoother DCT matrix of the image can be obtained. This cropped image when compared to the stego-image will have large differences in their statistics if the image was modified to embed data. If the stego-image was not modified in any way, the difference between the image and its cropped version will not have significant differences in their statistics. Visually and technically, the calibrated image is compared with the stego image based on statistical measures such as PSNR, Histogram etc.

The advantage of calibration is that most stego-systems encode the secret message in the transform domain to produce stego-image of JPEG format and as the JPEG compression algorithm operates by transforming the image into 8x8 blocks, and it is within these blocks the secret data are encoded. Thus the presence of secret data can be estimated by introducing a new block structure and comparing it with that of the stego-image. When testing the calibration process, it became apparent that the best methodology was to crop the image by 4 pixels in every direction (top, bottom, left, and right). Cropping from all edges ensures that the entire block structure is removed, and thus a more accurate estimation is derived.

The calibration method is able to detect and recover hidden data, so evaluating the difference between stego-image and calibrated image forms a detector. Though the calibration method is effective and robust detector but it has limitations. Known limitations of calibration include double-compressed JPEG images and images that contain *spatial resonance*. This occurs when the content has a periodicity close to the block size of the transformation.

5. Conclusion

Steganalysis is an effective mechanism to analyse the embedding performance of steganographic techniques. This paper presents a review of the main steganalysis techniques for detecting secret data inside

an image file and the techniques are discussed in their ability to detect secret message in an image file. We presented the strengths and weaknesses of various stego-systems in terms of steganalysis. A steganalyst is frequently interested in more than whether or not a secret message is present. The ultimate goal is to extract the secret message. However, in the absence of the knowledge of the stego technique and the stego-image, the process of detection and analysis can be time consuming or completely infeasible. Also, the analysis may give significant amount of false positives. Therefore, any additional information, such as the message length or its embedding position could prove very valuable to the analyst. From the various work done on steganalysis it is found that the basis of the attacks is to just be able to distinguish between a cover and a stego image. Even if the exact stego system that was used to make the stego isn't known, there are basically a few known steganographic embedding techniques that could have been used. Each of these steganalysis technique has pros and cons. There is no universal steganalysis technique and thus, it is up to the user (steganalyst) to choose an appropriate methodology based on the information that is available.

6. References

1. N. Provos and P. Honeyman, "Hide and seek: An introduction to steganography", Security *& Privacy, IEEE* 1, no. 3 (2003): 32-44.
2. N. F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen", *IEEE computer* 31, no. 2 (1998): 26-34.
3. C. Hosmer, "Discovering hidden evidence", *Journal of Digital Forensic Practice* 1, no. 1 (2006): 47-56.
4. C.C. Chang, T.S. Chen, and L. Z. Chung, "A steganographic method based upon JPEG and quantization table modification", *Information Sciences* 141, no. 1 (2002): 123-138.
5. S. A. Laskar and K. Hemachandran, "An Analysis of Steganography and Steganalysis Techniques", *Assam University Journal of Science and Technology*, Vol.9, No.II, pp.83-103, (2012), ISSN: 0975-2773.
6. S. K. Bandyopadhyay and I. K. Maitra, "An Application of Palette Based Steganography", *International Journal of Computer Applications* (0975 – 8887) Volume 6, No.4, pp.24-27, 2010.
7. S. S. Agaian, and J. P. Perez, "New Pixel Sorting Method for Palette Based Steganography and Color Model Selection", *The University of Texas, San Antonio* (2004).
8. T. Liu, and Z. D. Qiu, "A DWT-based color image steganography scheme", In *Signal Processing, 2002 6th International Conference on*, vol. 2, pp. 1568-1571. IEEE, 2002.
9. S. Audithan and R. M. Chandrasekaran, "Document Text Extraction from Document Images Using Haar Discrete Wavelet Transform", *European Journal of Scientific Research* 36, no. 4 (2009): 502-512.
10. S. Katzenbeisser and F. A.P. Petitcolas, *Information hiding techniques for steganography and digital watermarking*. Vol. 316. Norwood: Artech house, 2000.
11. P. Wayner, *Disappearing cryptography: information hiding: steganography & watermarking*. Morgan Kaufmann, 2009.
12. J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color, and gray-scale images", *Multimedia, IEEE* 8, no. 4 (2001): 22-28.
13. X. Y. Luo, D.S. Wang, P. Wang, and F. L. Liu, "A review on blind detection for image steganography", *Signal Processing* 88, no. 9 (2008): 2138-2157.
14. Chandramouli, Rajarathnam, Mehdi Kharrazi, and Nasir Memon, "Image steganography and steganalysis: Concepts and practice", In *Digital Watermarking*, pp. 35-49. Springer Berlin Heidelberg, 2004.
15. K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, "Steganalysis for Markov cover data with applications to images", *Information Forensics and Security, IEEE Transactions on* 1, no. 2 (2006): 275-287.
16. D. Fu, Y. Q. Shi, D. Zou, and G. Xuan, "JPEG steganalysis using empirical transition matrix in block DCT domain", In *Multimedia Signal Processing, 2006 IEEE 8th Workshop on*, pp. 310-313. IEEE, 2006.
17. N. Provos, "Defending Against Statistical Steganalysis", In *Usenix Security Symposium*, vol. 10, pp. 323-336. 2001.
18. N. Provos and P. Honeyman, "Detecting steganographic content on the internet", *Ann Arbor* 1001 (2001): 48103-4943.
19. R. Chandramouli and K. P. Subbalakshmi, "Current trends in steganalysis: a critical survey", In *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*, vol. 2, pp. 964-967. IEEE, 2004.
20. R. Chandramouli, "A mathematical framework for active steganalysis", *Multimedia Systems* 9, no. 3 (2003): 303-311.
21. A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems", In *Information Hiding*, pp. 61-76. Springer Berlin Heidelberg, 2000.
22. K. Patil, R. Gupta and G. Singh, "Digital Image Steganalysis Schemes for Breaking

Steganography", *International Conference on Advances in Communication and Computing Technologies (ICACACT) 2012* (IJCA) 11.

23. I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics", *Image Processing, IEEE Transactions on* 12, no. 2 (2003): 221-229.
24. S. Lyu and H. Farid, "Steganalysis using color wavelet statistics and one-class support vector machines", In *Electronic Imaging 2004*, pp. 35-45. International Society for Optics and Photonics, 2004.
25. J. Fridrich, M. Goljan, D. Hogea, and D. Soukal, "Quantitative steganalysis of digital images: estimating the secret message length",*Multimedia systems* 9, no. 3 (2003): 288-302.
26. J. Fridrich, M. Goljan, and D. Hogea, "Attacking the outguess", In *Proceedings of the ACM Workshop on Multimedia and Security*, vol. 2002. Juan-les-Pins, France, 2002.
27. J. Fridrich, M. Goljan, and D. Soukal, "Higher-order statistical steganalysis of palette images", In *Electronic Imaging 2003*, pp. 178-190. International Society for Optics and Photonics, 2003.

Authors

Shamim Ahmed Laskar received his B.Sc. and M.Sc. degrees in Computer Science in 2006 and 2008 respectively from Assam University, Silchar, where he is currently doing his Ph.D. His research interest includes Image Processing, Steganography, Information Retrieval and Data Security.

Prof. Kattamanchi Hemachandran is currently serving as the Head of the Department in the Department of Computer Science, Assam University, Silchar and is associated with this department since 1998. He obtained his M.Sc. Degree from Sri Venkateswara University, Tirupati and M.Tech and Ph.D Degrees from Indian School of Mines, Dhanbad. He is supervising many research scholars. His areas of research interest are Image Processing, Software Engineering and Distributed Computing.