Stego-Image Vs Stego-Analysis System

B.B Zaidan, A.A Zaidan, Alaa Taqa, Fazida Othman

Abstract -Steganography is the idea of hiding private, sensitive data or information within something that appears to be nothing but normal, in this article the author invent comprehensive study on this stego-image; in fact, there are some factors discussed experimentally such as steganography classification, applied algorithms, Stego-image, the impact of data hidden on the image texture, in the other hand the author named the most commend methods used by the attackers to against the data hidden in image. As it shown below there are three illustrious technique used, sequentially, statistical technique, try and error technique and finally histogram technique; theses techniques has been discussed in details and evidenced by some experiment result

Keyword-Steganography, Hidden Data, HVS, Stego Image ,Attacker, High Rate Data Hidden

I. INTRODUCTION

Steganography is the art and science of hiding messages. Steganography and cryptology are similar in the way that they both are used to protect important information. The difference between the two is that Steganography involves hiding information so it appears that no information is hidden at all. If a person views the digital object that the information is hidden inside, he or she will have no idea that there is any hidden information, therefore the person will not attempt to decrypt the information, this is the main objective behind steganography. Steganography comes from the Greek words Steganós (Covered) and Graptos (Writing), these days the sense of the word "steganography" usually refers to information or a file that has been concealed inside a digital Picture, Video or Audio file[1],[2]. What Steganography technically does is to make use of human awareness; human senses are not trained to look for files that have information hidden inside of them, although there are programs available that can do what is called Steganalysis (Detecting use of Steganography.) The most common use of Steganography is to hide a file inside another file. When information or a file is hidden inside a carrier file, the data is usually encrypted with a password [3][8][4]

Manuscript received June 20, 2009

B. B. Zaidan - PhD candidate, Department of Computer Science & Information Technology, University Malaya, Kuala Lumpur, Malaysia, <u>bilal@perdana.um.edu.my</u>.

A. A. Zaidan - PhD candidate, Department of Computer Science & Information Technology, University Malaya, Kuala Lumpur, Malaysia, phone: +60172452457, Postcode: 50603 and Email: awsalaa@perdana.um.edu.my or aws.alaa@yahoo.com.

Alaa Taqa - Visitor Researcher , Department of Computer Science & Information Technology, University Malaya, Kuala Lumpur, Malaysia, , Postcode: 50603 Emailtalaa_taka@yahoo.com.

F. Othman- Mrs. Fazidah Othman - Lecturer, Department of Computer Science & Information Technology, University Malaya, Kuala Lumpur, Malaysia, <u>fazidah@um.edu.my</u>.

II. CLASSIFICATION OF STEGANOGRAPHY TECHNIQUES

There are several approaches in classifying Steganographic systems. One could categorize them according to the type of covers used for secret communication or according to the cover modifications applied in the embedding process. The second approach will be followed in this section, and the Steganographic methods are grouped in six categories, although in some cases an exact classification is not possible. Figure 1 presents the steganography classification.



A. Substitution Systems

Basic substitution systems try to encode secret information by substituting insignificant parts of the cover by secret message bits. The receiver can extract the information if he has knowledge of the positions where secret information has been embedded. Since only minor modifications are made in the embedding process, the sender assumes that they will not be noticed by an attacker. It consists of several techniques that will be discussed in more detail, in the following subsection:

(i) Least Significant Bit Substitution (LSB)

The embedding process consists of choosing a subset $\{ j1..., jl(m) \}$ of cover elements and performing the substitution operation $cji \leftrightarrow mi$ on them, which exchange the LSB of cji by mi (mi can be either 1 or 0).

In the extraction process, the LSB of the selected coverelement is extracted and lined up to reconstruct the secret message.

In the case of a 24-bit bitmap each pixel is represented by 4 bytes. Of those, 3 bytes, or 24 bits, are used to store the red, green and blue values for the pixel. The fourth byte is reserved and should be zero. To store each character in the low order bit plane of the raster data, it is necessary to obtain an 8 bit representation of the character. For example,



the character 'A' is represented by the number 65. The equivalent binary representation is '0100 0001'. Each of the 8 bits used to represent the letter A is then stored in the low order bit of one byte of raster data. Thus, to store a single letter, 8 bytes of raster data are consumed. This leads to a limit of embeddable information of size lengthOfRasterData/8. Consider hiding the letter A in the first 8 bits of raster data of an image. The first 8 bytes could possibly be (from left to right, top to bottom):

 '1001 1001'
 '1110 0011'
 '0110
 1001'

 '0001 1100'
 '0110 0100'
 '1011
 0000'

 '1010 1001'
 '0110 0100'
 '1011
 0000'

And the character 'A' is:

'0100 0001^{*}

Therefore, we need to set bits 7, 5, 4, 3, 2, and 1 to zero, this is accomplished by ANDing with the mask '1111 1110'. The result for the first byte is:

	1001	1001
AND	<u>'1111</u>	1110'
	[•] 1001	1000'

So the low order bit is set to '0'. This is repeated for all bits that will be set to carry a '0'.

We now need to set bits 6 and 0 to '1'. This is accomplished by ORing with the mask '0000 0001'. The result for the second byte is:

·1110 0011
<u>'0000 0001'</u>
ʻ1110 0011'

Although the resulting bit has not changed, we have ensured that the least significant bit has been set to '1'.

Because the byte values for the red, green and blue pixels will only change by at most 1, the change in the resulting image will be imperceptible to the human eye. The resulting image will not, however, be well protected against statistical attack.

(ii) Pseudorandom Permutation

If all cover bits are accessed in the embedding process, the cover is a random access cover, and the secret message bits can be distributed randomly over the whole cover. This technique further increases the complexity for the attacker, since it is not guaranteed that the subsequent message bits are embedded in the same order [9], [10].

(iii) Image Downgrading and Cover Channels

Image downgrading is a special case of a substitution system in which image acts both as a secret message and a cover. Given cover-image and secret image of equal dimensions, the sender exchanges the four least significant bits of the cover grayscale (or color) values with the four most significant bits of the secret image. The receiver extracts the four least significant bits out of the stego-image, thereby gaining access to the most significant bits of the stego-image. Whereas the degradation of the cover is not visually noticeable in many cases, four bits are sufficient to transmit a rough approximation of the secret image.

(iv) Cover Regions and Parity Bits

Any nonempty subset of $\{c1,\ldots,cI(c)\}$ is called a cover-region. By dividing the cover into several disjoint regions, it is possible to store one bit of information in a whole cover-region rather than in a single element. A parity

bit of a region I can be calculated by:

B (I) == Σ LSB (cj) mod2 [27]. $j \in I$

```
(v) Palette-Based Image
```

There are two ways to encode information in a palettebased image; either the palette or the image data can be manipulated. The LSB of the color vectors could be used for information transfer, just like the substitution methods presented. Alternatively, since the palette does not need to be sorted in any way, information can be encoded in the way the colors are stored in the palette. For N colors since there are N! Different ways to sort the palette, there is enough capacity to encode a small message. However, all methods which use the order of a palette to store information are not robust, since an attacker can simply sort the entries in a different way and destroy the secret message[9][10].

(vi) Quantization and Dithering

Dithering and quantization to digital image can be used for embedding secret information. Some Steganographic systems operate on quantized images. The difference ei between adjacent pixels xi and xi +1 is calculated and fed into a quantize φ which outputs a discrete approximation ΔI of the different signal (xi - xi + 1). Thus in each quantization step a quantization error is introduced [9],[10].In order to store the ith message bit in the cover-signal, the quantized difference signal ΔI is computed. If according to the secret table ΔI does not match the secret bit to be encoded, ΔI is replaced by the nearest ΔI where the associated bit equals the secret message bit. The resulting value ΔI is those fed into the entropy coder. At the receiver side, the message is decoded according to the difference signal ΔI and the stegokey

B. Transform Domain Techniques

It has been seen that the substitution and modification techniques are easy ways to embed information, but they are highly vulnerable to even small modification. An attacker can simply apply signal processing techniques in order to destroy the secret information. In many cases even the small changes resulting out of loose compression systems yield total information loss. It has been noted in the development of Steganographic systems that embedding information in the frequency domain of a signal can be much more robust than embedding rules operating in the time domain. Most robust Steganographic systems known today actually operate in some sort of transform domain[3][9]. Transformation domain methods hide message in a significant area of the cover image which makes them more robust to attack, such as adding noise, compression, cropping some image processing. However, whereas they are more robust to various kinds of signal processing, they remain imperceptible to the human sensory system. Many transform domain variations exist. One method is to use the Discrete Cosine Transformation (DCT) as a vehicle to embed information in image. Another method would be the use of wavelet transforms. Transforms embedding embeds a message by modification (selected) transform (e.g., frequency) coefficient of the cover message. Ideally, transform embedding has an effect on the spatial domain to apportion the hidden information through different order bits in a manner that is robust, but yet hard to detect. Since an attack, such as image processing, usually affects a certain band of transform coefficient, the remaining coefficient would remain largely intact. Hence, transform embedding is, in general, more robust than other embedding methods [10]. *C. Spread Spectrum (SS) Techniques*

Spread spectrum techniques are defined as "Means of transmission in which the signal occupies a bandwidth in excess of the minimum necessary to send the information ". The band spread is accomplished by means of a code which is independent of the data, and a synchronized reception with the code at the receiver is used for dispreading and subsequent data recovery. Although the power of the signal to be transmitted can be large, the signal-to-noise ratio in every frequency band will be small, even if parts of the signal could be removed in several frequency bands, enough information should be present in the other bands to recover the signal. This situation is very similar to a steganography system which tries to spread a secret message over a cover in order to make it impossible to perceive. Since spread signals tend to be difficult to remove, embedding methods based on SS should provide a considerable level of robustness[9].In information hiding, two special variants of spread spectrum techniques are generally used: direct sequence, and frequency-hopping scheme. In directsequence scheme, the secret signal is spread by a constant called ship rate, modulated with a pseudorandom signal and added to the cover. On the other hand, in the frequencyhopping schemes the frequency of the carrier signal is altered in a way that it hops rapidly from one frequency to another. SS are widely used in the context of watermarking[9][10].

D. Statistical Steganography

Statistical steganography techniques utilize the existence of "1-bits" steganography schemes, which embed one bit of information in a digital carrier. This is done by modifying the cover in such a way that some statistical characteristics change significantly if a "1" is transmitted. Otherwise, the cover is left unchanged. So the receiver must be able to distinguish unmodified covers from modified ones.

A cover is divided into l(m) disjoint blocks B1..... B l(m). A secret bit, mi is inserted into the ith block by placing "1" into Bi if mi=1.

Otherwise, the block is not changed in the embedding process [10].

E. Distortion Techniques

In contrast to substitution systems, distortion requires the knowledge of the original cover in the decoding process. The sender applies a sequence of modifications to the cover in order to get a stego-system. A sequence of modification is chosen in such a way that it corresponds to a specific secret message to be transmitted. The receiver measures the difference in the original cover in order to reconstruct the sequence of modification applied by the sender, which corresponds to the secret message [3]. In many applications, such systems are not useful, since the receiver must have access to the original cover. If the attacker also has access to them, he/she can easily detect the cover modification and has evidence for a secret communication. If the embedding and extraction functions are public and do not depend on a stego-key, it is also possible for the attacker to reconstruct secret message entirely [10].

F. Cover Generation Techniques

In contrast to all embedding methods presented above, when secret information is added to a specific cover by applying an embedding algorithm, some Steganographic applications generate a digital object only for the purpose of being a cover for secret communication[9].

III. DESIGN WEAKNESSES OF LSB

There are several design weaknesses of this software which make it susceptible to attack [3][7]. One is statistical attack. The embedded information always begins at the first byte of raster data. Because the data is not spread out randomly, it would be possible to use statistics to determine if any bytes in the image were altered. For example, within images, regions of the image often contain repetitions of the exact same color across multiple pixels [9]. For this reason, it would be possible to discover that variations exist in neighboring pixels that should not be present. Another weakness is that the program does not allow the user to encrypt the message using a strong encryption technique first. This is not so much a design weakness as it is an unsupported feature. While far from a robust Steganography software package, the software does support a primitive ability to embed a text message inside a standard, widelyused digital image format [10].

IV. The affict of data hidden in the Image texture $$\rm Using\,LSB$$

In this section we present a mathematical formulation for Analysis of LSB based steganographic techniques. LSB based Steganographic techniques either change the pixel value by ± 1 or leave them unchanged [3][7]. This is dependent both on the nature of the hidden bit and the LSB of the corresponding pixel value. Let $\mathbf{I} = \{X\mathbf{i}, \mathbf{i} \in \Omega\}$ where Ω is an index set denote the mean subtracted cover image. The set Ω can be partitioned into three subsets A1, A2, and A3, where, $\Omega = \bigcup_{i=1}^{3} A\mathbf{i} A\mathbf{i} and A\mathbf{i} \cap A\mathbf{j} = \emptyset$ for $\mathbf{i} \neq \mathbf{j}$ Then, the pixel values in a LSB based stego-image $\mathbf{I} \mathbf{s} = \{Y\mathbf{i}, \mathbf{i} \in \Omega\}$ can be represented as

$$Yi = \begin{cases} Xi + 1 & \text{if } i \in A1 \\ Xi - 1 & \text{if } i \in A2 \\ Xi & i \in A3 & \dots \dots \dots \dots (1) \end{cases}$$

Currently we will call the algorithms of the hidden data in the image by the number of bit's that used for embedding data

(i) 1st LSB

Good enough since the change was not that big regarding to the amount of color for each pixel, we already mentioned about LSB algorithm above.

(ii) 2nd LSB

This level of LSB also good and suitable for any type of image, the change was not that big since we play in the 2nd LSB



Let $I = {Xi, i \in \Omega}$ where Ω is an index set denote the mean subtracted cover image. The set Ω can be partitioned into three subsets A1, A2, and A3, where, $\Omega = \bigcup_{i=1}^{3} Ai Ai and Ai \cap Aj = \emptyset$ for $i \neq j$, Then, he pixel values in a LSB based stego-image $I s = {Yi, i \in \Omega}_{can be represented as}$



Figure 2 showing the 2nd LSB bit in 24-bit image

(iii) 3rd LSB and the spicieal approuch 3-3-2

In 3rd LSB, the change is clear and it can be suspected so we prefer to apply HVS and its rules. The next section is about the human vision system (HVS).



Figure 3 Image before and after hidden opration (LSB)

(iv) Smooth Texture Problem

A lot of ways to calculate the smooth texture and the noisy texture its Quite far from this study but the need of that to understand what is the meaning of smooth texture, from the trying on the image we define this formula for the smooth area in the image as the test of the image showing



Then, this equation become

$$ST = \sum_{e=1}^{j-7} \sum_{r=1}^{i-7} Ch$$

Ch = F(W) check function on the matrix W After extract the 8*8 pixel from the image then the pixels on this matrix should be not equal or at least not near to each amount of color.



Figure 4 simple textures before	Flat Area in the
Hidden data	image





A. Try and Error

The process of hidden data by using Least Significant Bit (LSB), which is a common method, used the image as a cover for data to be hidden, but it's mostly subjected to attacks from attackers. Once attackers suspect there is Steganography implemented in the data, they will reproduce LBS and examine it to check whether it has a meaning or not. In this case the attacker use try and error to exam the image in case he or she suspect there is a data hidden in this image, mostly the suspect come when the attacker cutch only the stego file that contains the hidden data, In this case, attacker try to analysis this stego file, analysis is done by trial and error. In this case the attacker is sure of the existence for the hidden data inside the carrier cover and this flaw often happens for two reasons, the first reason is the sender or receiver are Intentionally or accidentally disclose and the second reason is the attacker use some techniques or methods involve either stealing the system or one of its parts or use of analytical methods to detect as following: Signal to - Noise Ratio (SNR): And that's done by calculating the noise ratio to the total volume of a file (SNR: Signal - to -Noise Ratio), the higher the ratio, the more is the doubt of the existence hidden data, knowing that the process of calculating noise is done by examining the sudden changes in the value of sample and comparing it to the neighbor.

This type of analysis is not successful here, because the concealment is done evenly to all parts of the cover. The reason of observing the hidden data is caused by the weakness of the Steganography system or insufficient cover of concealment that makes it noticeable in some part.[5][6]

B. Statistical Techniques

Most of the statistical techniques come when the attacker knows one of factor of the hidden information. Important issues due consideration to prevent the discovery of concealment is choosing the cover file, to find this cover, the most important characteristics to be met cover are as follows:

- The data of the cover should not be that important or relevant to the embedded data.
- In any case the attacker should not be able to know the original size for cover.
- Should not use a common cover like a popular image or well known audio file.
- keep switching the cover constantly and do not keep it fix, especially in case of sending more than one message at the same time, to prevent attacker to notice the difference between both cases.
- Send blank cover with the covers that contains the data for the purpose of camouflage, to reduce the chance of discovery.

The cover should be selected very randomly and high frequency, as its models values changes a lot and it doesn't contain large areas of equal value of adjacent samples values. The statistical techniques more efficient with gray level images, where the image not more than a matrix of pixels present the RGB, and R=G=B, if the information hidden in the gray level images there will be some diverge on the pixel in the level of RGB



Next figure shows how the LSB effect the images of gray color in the pixel level.



Figure 7 in the gray level we may notes the bits that effected by add or remover one which make the $R\neq G\neq B$

C. Histogram Techniques

The earlier experimental result and the assessment test shows we can use the histogram to classify the images into normal image and stego image, where the embedding operation have been in to two type of image (gray level image, color image), the new habits of the histogram shows in the color image, color become a group and distant peaks in other word the distant between summits In direct proportion with the increasing data hidden in the image, and this bizarre conduct is our start point to identify the new approach of the stego-analysis systems next figures showing the result of embedding operation also the histogram before hidden data and after hidden data [6][5]



Figure 8 images before and after data hidden

The picture almost the same even there are more than 50% of its size as data hidden.



Figure 9 showing the summit and the bizarre conduct for the histogram

As we know the histogram is the number of repetitions of each level of color, in other word how many time the level of the color x repetition in the image, where x between 0-256 if the image 24-bit.





Figure 10 images before and after data hidden



Figure 11 Histogram before and after data hidden

In the gray level image there is more than this feature, as we know the gray level image is the image that have the same value for the three color in each pixel, for example if pixel x has three color and the value of blue is 33 its mean each of the red and the green should be 33, so that we may get one of the gray level color, this feature guide us to define a way to implement a system detect the gray level image, where the embedding operation should give some result that affect the histogram, the new histogram of each image will give same level of color depend on the change of the three color , we will consider that if there is an embedding data there should be a change otherwise the data match with the picture and that is impassable [2][3][4][5].Next figures showing the images before and after hidden data also the histogram before and after data hidden.



Figure 12 Image before and after hidden data



Figure 13 color in the histogram of gray level image



Figure 14 Histogram before and after hidden data

VI. CONCLUSION

In this article the author invent comprehensive study on stego-image, on which the author study the impact of data hidden on the image texture, also the secrecy of using data hidden in the image. To provide a complete study the author also describe the most illustrious technique used by the attackers to extract the data hidden which are sequentially, statistical technique, try and error technique and finally histogram technique, each of these techniques has been given with examples, also the perfect environment for choice one of these techniques. In fact, by this study the author reviews widely about the stego-image, classification, challenges, and risk of hidden data in the image.

ACKNOWLEDGEMENT

This work was supported in part by the University of Malaya, Kuala Lumpur Malaysia.

REFERENCES

- A.A.Zaidan, B.B.Zaidan, M.M.Abdulrazzaq, R.Z.Raji and S.M.Mohammed, "Implementation Stage for High Securing Cover-File of Hidden Data Using Computation between Cryptography and Steganography". International Association of Computer Science and Information Technology (IACSIT), Volume 20, 2009, Manila, Philippines.
- [2] B.B Zaidan, A.A Zaidan, Fazidah Othman, R.Z.Raji, S.M.Mohammed, M.M.Abdulrazzaq, "Quality of Image vs. Quantity of Data Hidden in the Image", International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'09), 2009, Las Vigas, USA.
- [3] B.B.Zaidan, A.A.Zaidan, Fazidah Othman "Enhancement of the Amount of Hidden Data and the Quality of Image", Malaysia Education Security (MyEduSec08), Grand Continental Hotel, 2008, Kuala Trengano, Malaysia.
- [4] A.W. Naji, Teddy S. Gunawan and Shihab A. Hameed, B.B Zaidan, A.A Zaidan " "Stego-Analysis Chain, Session One" Investigations on Steganography Weakness Vs Stego-Analysis System for Multimedia File ", International Conference on IACSIT Spring Conference (IACSIT-SC09), Session 9, P.P 393-397, 2009, Singapore.
- [5] A. W. Naji, Shihab A. Hameed, Md Rafiqul Islam, B. B. Zaidan, Teddy S. Gunawan, and A. A. Zaidan," "Stego-Analysis Chain, Session Two" Novel Approach of Stego-Analysis System for Image File ", International Conference on IACSIT Spring Conference (IACSIT-SC09), Session 9, P.P 398-401, 2009, Singapore.
- [6] Mohamed Elsadig Eltahir, Laiha Mat Kiah, B.B.Zaidan and A.A.Zaidan," High Rate Video Streaming Steganography", International Conference on Information Management and Engineering (ICIME09), Session 10, P.P 550-553, 2009, Kuala Lumpur, Malaysia.
- [7] Fazida.Othman, Miss.Laiha. Mat Kih, A.Y.Taqa, B.B.Zaidan, A.A.Zaidan, "An Extensive Empirical Study for the Impact of Increasing Data Hidden on the Images Texture", International Conference on Future Computer and Communication (ICFCC 09), Session 7, P.P 477-481, 2009, Kuala Lumpur, Malaysia.
- [8] Johnson, N. F. S. D, Z., "Information Hiding: Steganography and Watermarking-Attacks and Countermeasures", Center for Secure Information Systems (CSIS), Boston/Dordrecht/London, George Mason University, 2006.

- [9] Bilal Bahaa Zidan,2009 master thesis "Enhancement of the size of data hidden and the quality of image using LSB algorithm" University of malaya
- [10] Aos.A.Z.Ansaef, 2009 master thesis "Securing Cover-File of Hidden Data Using Statistical Technique and AES Encryption Algorithm" university of malaya



Bilal Bahaa – he obtained his bachelor degree in Mathematics and Computer Application from Saddam University/Baghdad followed by master from Department of Computer System & Technology Department Faculty of Computer Science and Information Technology/University of Malaya /Kuala Lumpur/Malaysia, He led or member for many funded research projects and He has published more than 40 papers at various international and national conferences and

journals. His research interest on Steganography & Cryptography with his group he has published many papers on data hidden through different multimedia carriers such as image, video, audio, text, and non multimedia careers such as unused area within exe.file, he has done projects on Stego-Analysis systems, currently he is working on Quantum Key Distribution QKD and multi module for Steganography, he is PhD candidate on the Department of Computer System & Technology / Faculty of Computer Science and Information Technology/University of Malaya /Kuala Lumpur/Malaysia.



Aos Alaa Zaidan - He obtained his 1st Class Bachelor degree in Computer Engineering from university of Technology / Baghdad followed by master in data communication and computer network from University of Malaya. He led or member for many funded research projects and He has published more than 40 papers at various international and national conferences and journals, he has done many projects on Steganography for data hidden through different multimedia carriers

image, video, audio, text, and non multimedia carrier unused area within exe.file, Quantum Cryptography and Stego-Analysis systems, currently he is working on the multi module for Steganography. He is PhD candidate on the Department of Computer System & Technology / Faculty of Computer Science and Information Technology/University of Malaya /Kuala Lumpur/Malaysia



Alaa Yasen Taga received her master degree on Computer Science, M.SC. on Applied Object-Oriented Software Engineering Methodology for Patterns Printing and Coloring on Textile, from NCC, Iraq. Her Ph.D. in Computer Science, on "CONSTRUCTING ANTI-SPAM FILTER BASED ON NAIVE BAYESIAN CLASSIFIER", Mosel University. Her interest research area on pattern recognition, network security, information protection, data encoding and decoding, AI and security applications, ante-spam system.

steganography and finally skin detector, Dr. Alaa job title is lecturer at university of Mosul, currently she has been appointed as a visiting researcher at the faculty of computer science and information technology department of AI, she has ongoing project titled "increase the reliability of the skin detector".



Fazidah Othman- has received her master from University Technology Malaysia, Faculty of Computer Science currently she is a lecturer, Department of Computer System & Technology/University of Malaya /Kuala Lumpur/Malaysia, her research interest on the network security, intelligent agent for Machine translation tools, she has many publication on Steganography, Agent Technology for Proxy Server, her PhD work on the multicast group

security management.

