Steganographic scheme to avoid statistical Steganalysis

Jamal A. Othman
Iraq Commission for Computers & Informatics/ Baghdad
Email:Jamalothman2003@yahoo.com

Abstract

Most of the steganographic scheme embeds the hidden message through the least significant bits either in the spatial domain or frequency domain sequentially or pseudo randomly through the cover media (Based on this fact) statistical Steganalysis use different techniques to detect the hidden message, A proposed method is suggested of a stenographic scheme a hidden message is embedded through the second least significant bits in the frequency domain of the cover media to avoid detection of the hidden message through the known statistical Steganalysis techniques.

طريقةإخفاء لتجنب تقنيات الكشف الاحصائية

جمال أحمد عثمان
الهيئة العراقية للحاسبات والمعلوماتية
بريد إلكتروني : jamalothman2003@yahoo.com

الخلاصة

أكثر أساليب الاخفاء المعروفة تستخدم أسلوب تقنية الاغمار بالثنائي الاقل اهمية (LSB) لتضمين الرسالة السرية في المدى المكاني أو المدى الترددي بشكل متسلسل أو عشوائي مزيف لملف التغطية .ان أكثر طرق الكشف عن الرسائل المخفية بأستخدام التحليل الاحصائي المعروفة تستند الى هذه الفرضية في تصميم تقنيات إحصائية للكشف عن وجود رسائل سرية في ملفات التغطية ، نقترح في هذا البحث أستخدام الثنائي الأقل أهمية الثاني (Second LSB) بدلا من الثنائي الأقل أهمية في تضمين الرسالة السرية في ملف الغطاء وذلك لتجنب الكشف عن وجود الرسالة السرية في ملف الغطاء من خلال إستخدام تقنيات التحليل الاحصائي المعروفة .

Keywords: Steganography, Steganalysis, Discrete Cosine Transformation (DCT), LSB, Second Least Significant Bit (SLSB).

1- Introduction

Steganography is the art and science of hiding communication; a steganographic system thus embeds secret data in unremarkable cover media so as not to arouse an eavesdropper's suspicion, the information-hiding process in a steganographic system starts by identifying a cover medium's redundant bits (those that can be modified without destroying that medium's integrity). [1]

Steganalysis is the science of detecting hidden information. The main objective of Steganalysis is to break steganography and the detection of stego image is the goal of steganalysis. Almost all steganalysis algorithms rely on the Steganographic algorithms introducing statistical differences between cover and stego image. Steganalysis deals with three important categories: (a) Visual attacks: In these types of attacks with a assistance of a computer or through inspection with a naked eye it reveal the presence of hidden information, which helps to separate the image into bit planes for further more analysis. (b) Statistical attacks: These types of attacks are more powerful and successful, because they reveal the smallest alterations in an images statistical behavior. [2]

There are several techniques used to emblements Steganography some of them are simple and easy to detected other are very complicated and very hard or even impossible to be

detected through the known and published steganalysis. algorithms .Many well known stegnographical schemes works by embedding the secrete data in the least significant bits (LSB) in the spatial or frequencies domain sequentially or pseudo randomly by using a seed key , many of steganalysis schemes works to detect the hidden data relying on the hypothesis that the hidden data exist in the least significant bit of the cover image .A proposal in this paper of a stegnographical scheme by which the Second Least Significant Bit (SLSB) is used rather than the Least Significant Bit (LSB) to a void wide range of steganalysis scheme , in this paper a brief introduction to steganography , Discreet cosine transformation , least significant bit with more attention to steganalysis and finely a pseudo code for the proposed algorithm is given.

2-Steganography

The word steganography comes from the Greek steganos, which mean covered or secret and graphy mean writing or drawing. Therefore, steganography means, literally, covered writing [3]. Steganography is the art of hiding information and an effort to conceal the existence of the embedded information. It serves as a better way of securing message than cryptography which only conceals the content of the message not the existence of the message [4]. 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 [5].

The earliest recordings of Steganography were by the Greek historian Herodotus date back to around 440 BC. Herodotus recorded two stories of Steganographic techniques during this time in Greece. The first stated that King Darius of Susa shaved the head of one of his prisoners and wrote a secret message on his scalp. When the prisoner's hair grew back, he was sent to the Kings Aristogoras in Miletus undetected. The second story also came from Herodotus, which claims that writing medium was text written on wax covered tablets. Romans used invisible inks, which were based on natural substances such as fruit juices and milk. This was accomplished by heating the hidden text, thus revealing its contents. Invisible inks have become much more advanced and are still in limited use today. During the 15th and 16th centuries, many writers including Johannes Trithemius (author of Steganographia) and Gaspari Schotti (author of Steganographica) wrote on Steganagraphic techniques such as coding techniques for text, invisible inks, and incorporating hidden messages in music. [6]

Modern Steganography refers to hide information in digital picture, audio or text files …etc, each one of this digitals data has a many techniques can use with it [7]. The information hiding process in a steganographic system starts by identifying a cover medium's redundant bits (those that can be modified without destroying that medium's integrity). [8] The embedding process creates a stego medium by replacing these redundant bits with data from the hidden message. [9]

In modern steganography embedded message is detectable only if secret information is known namely, a secret key [8]. This is similar to Kerckhoffs'Principle in cryptography, which holds that a cryptographic system's security should rely solely on the key material. For steganography to remain undetected, the unmodified cover medium must be kept secret, because if it is exposed, a comparison between the cover and stego media immediately reveals the changes. [10] Christian Cachin proposed an information theoretic model for steganography that considers the security of steganographic systems against passive eavesdroppers. [11] In this model, we assume that the adversary has complete knowledge of the encoding system but does not know the secret key. His or her task is to devise a model for the probability distribution PC of all possible cover media and PS of all possible stego media. The adversary can then use detection theory to decide between hypothesis C (that a message

contains no hidden information) and hypothesis S (that a message carries hidden content). A system is perfectly secure if no decision rule exists that can perform better than random guessing. Essentially, steganographic communication senders and receivers agree on a steganographic system and a shared secret key that determines how a message is encoded in the cover medium. To send a hidden message, for example, Alice creates a new image with a digital camera. Alice supplies the steganographic system with her shared secret and her message. The steganographic system uses the shared secret to determine how the hidden message should be encoded in the redundant bits. The result is a stego image that Alice sends to Bob. When Bob receives the image, he uses the shared secret and the agreed on steganographic system to retrieve the hidden message. Figure (1) shows the model of the general steganographic methods. For example Derek Upham's JSteg is a publicly available steganographic system for JPEG images. Its embedding algorithm sequentially replaces the least-significant bit of DCT coefficients with the message's data. The algorithm does not require a shared secret; as a result, anyone who knows the steganographic system can retrieve the message hidden by Jsteg while OutGuess 0.1 is a steganographic system that improves the encoding step by using a pseudo-random number generator to select DCT coefficients at random. The least-significant bit of a selected DCT coefficient is replaced with encrypted message data. [12]

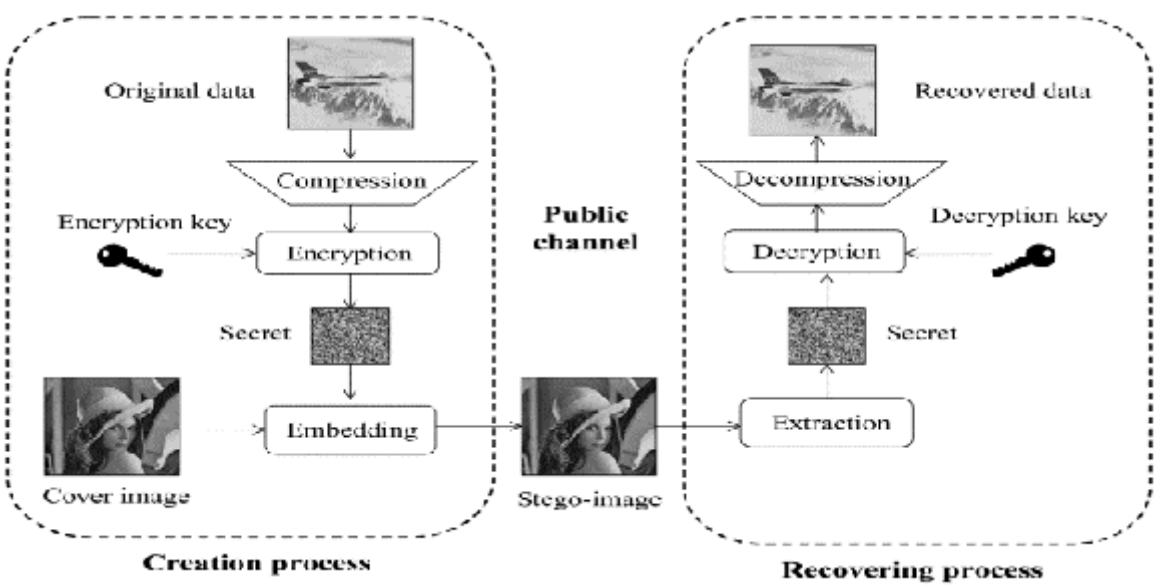

Fig. 1. The model of the general steganographic methods.

In general we can distinguish two main steganographic techniques: (i) Spatial domain technique: In this technique only the least significant bits of the cover object is replaced without modifying the complete cover object. It is a simplest method for data hiding but it is very weak in resisting even simple attacks such as compression, transforms, etc. (ii) Transform domain technique: The various transform domains techniques are Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) and Fast Fourier Transform (FFT) are used to hide information in transform coefficients of the cover images that makes much more robust to attacks such as compression, filtering, etc. [13]

3-LSB insertion method

The least significant bit insertion method is probably the most well known image Steganography technique. It is a common, simple approach to embed information in a graphical image file. [14]

The LSB technique either directly embed the secret data within the pixels of the cover image or used the image frequency domain to embed the secret data in LSB of the discrete cosine transform (DCT) coefficients. [15] In some cases LSB of pixels or the (DCT) coefficients visited in random or in certain areas of image and sometimes increment or decrement the pixel or coefficient value [16].

There are many algorithms used to embed secret data in a media file for example JSteg algorithm is one of them. The algorithm sequentially replaces the least-significant bit of discrete cosine transform (DCT) coefficients with message data. It does not require a shared secret and here is a simple pseudo-code algorithm to hide a message inside a JPEG image: [17]

Input: message, cover image
Output: steganographic image containing message
While data left to embed **do**
Get next DCT coefficient from cover image
If DCT ≠ 0 and DCT ≠ 1**then**
Get next **LSB** from message
Replace DCT **LSB** with message bit
End if
Insert DCT into steganographic image
End while

Although there are several disadvantages to this approach, the relative easiness to implement it, makes it a popular method. To hide a secret message inside a image, a proper cover image is needed. Because this method uses bits of each pixel in the image, it is necessary to use a lossless compression format, otherwise the hidden information will get lost in the transformations of a lossy compression algorithm. Disadvantages of using LSB alteration are mainly in the fact that it requires a fairly large cover image to create a usable amount of hiding space. Even nowadays, uncompressed images of 800 x 600 pixels are not often used on the Internet, so using these might raise suspicion. Another disadvantage will arise when compressing an image concealing a secret using a lossy compression algorithm. The hidden message will not survive this operation and is lost after the transformation. However, the LSB insertion method is easy to be attacked. [18]

4-Discrete Cosine Transformation (DCT)

The discrete cosine transforms (DCT) is a technique for converting a signal into elementary frequency components [19] .The DCT transformation works by separating images into different frequencies and then less important frequencies will discarded and only the most important frequencies are used to retrieve the images. The rapid growth of digital imaging applications, including desktop publishing, multimedia, teleconferencing, and high definition television (HDTV) has increased the need for effective and standardized image compression techniques. Among the emerging standards are JPEG, for compression of images [20]. Most of the compression standards employ the discrete cosine transform (DCT) Technique. Its application to image compression was pioneered by Chen and Pratt. [21]

Discrete Cosine Transformations (DCT)), are used by the JPEG compression algorithm to transform successive 8 x 8 pixel blocks of the image, into 64 DCT coefficients each. The general overview of the Jpeg process as follows: [22]

1. The image is broken into 8x8 blocks of pixels.
2. The DCT is applied to each block, working from left to write top to bottom.
3. Each block is compressed through quantization
4. Image is stored using the compressed blocks in reduced amount of space.
5. When needed images are reconstructed by decompressed the coefficients by using the Inverse Discrete Cosine Transformation (IDCT).

Each DCT coefficient $F(u,v)$ of an 8 x 8 block of image pixels $f(x,y)$ is given by:

$$F(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{\pi(2x+1)u}{2N} \cos \frac{\pi(2x+1)v}{2N} \quad \ldots. (1)$$

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } u = 0 \\ \sqrt{\frac{2}{N}} & \text{for } u \neq 0 \end{cases} \quad \ldots. (2)$$

After calculating the coefficients, the following quantizing operation is performed:

$$F_Q(u,v) = \frac{F(u,v)}{Q(u,v)} \quad \ldots.(3)$$

Where $Q(u,v)$ is a 64-element quantization table.

We can use the least-significant bits of the quantized DCT coefficients as redundant bits in which to embed the hidden message. The modification of a single DCT coefficient affects all 64 image pixels. Although a modification of a single DCT will affect all 64 image pixels, the LSB of the quantized DCT coefficient can be used to hide information. Lossless compressed images will be susceptible to visual alterations when the LSB are modified. This is not the case with the above described method, as it takes place in the frequency domain inside the image, instead of the spatial domain and therefore there will be no visible changes to the cover image.

5-Steganalysis

The development of techniques for steganography has led to an increased interest in steganalysis techniques. [23] The main idea behind steganalysis is that the existence of a secret data will modify the cover medium and change its statistical properties, so eavesdroppers can detect the distortions in the resulting stego medium's statistical properties. The process of finding these distortions is called statistical steganalysis. [10]

Through steganalysis we try to determine if an image (or other carrier) contains an embedded message Current steganalysis methods fall broadly into one of two categories: embedding specific or universal. While universal steganalysis attempts to detect the presence of an embedded message independent of the embedding algorithm and, ideally, the image format, embedding specific approaches to steganalysis take advantage of particular algorithmic details of the embedding algorithm. Given the ever growing number of steganography tools, universal approaches are clearly necessary in order to perform any type of generic large-scale steganalysis. [24] Another way to classify detection algorithms: one

based on inherent statistical properties and the other on class discrimination. Detection algorithms based on inherent statistical properties have the advantage that they do not need to find a representative training set; moreover, they often let us estimate an embedded message's length. However, each steganographic system requires its own detection algorithm. Class discrimination, on the other hand, is universal even though it doesn't provide an estimate of the hidden message's length, and creating a representative training set is often difficult. A feature vector can help detect several steganographic systems, once we get a good training set. [10]

We are going to present the detection statistical steganalysis for example cases of embedding through the least significant bits. For the case of sequential embedding in the least-significant bits of DCT coefficients (as seen in many stegnographic scheme for example JSteg) Andreas Westfeld and Andreas Pfitzmann noticed that steganographic systems that change least-significant bits sequentially cause distortions detectable by steganalysis.[25] In the simple case, the embedding step changes the least-significant bit of colors in an image. The colors are addressed by their indices *i* in the color table; we refer to their respective frequencies before and after embedding as n_i and n_i^* . Given uniformly distributed message bits, if $n_{2i} > n_{2i+1}$, then pixels with color 2*i* are changed more frequently to color 2*i* + 1 than pixels with color 2*i* + 1 are changed to color 2*i*. As a result, the following relation is likely to hold:

$$|n_{2i} - n_{2i+1}| \geq |n_{2i}^* - n_{2i+1}^*| \quad \ldots\ (4)$$

In other words, embedding uniformly distributed message bits reduces the frequency difference between adjacent colors. [10]

The same is true in the JPEG data format. Instead of measuring color frequencies, we observe differences in the DCT coefficients' frequency Westfeld and Pfitzmann used a χ2-test to determine whether the observed frequency distribution y_i in an image matches a distribution y_i^* that shows distortion from embedding hidden data. Although we do not know the cover image, we know that the sum of adjacent DCT coefficients remains invariant, which lets us compute the expected distribution y_i^* from the stego image. Letting n_i be the DCT histogram, we compute the arithmetic mean to determine the expected distribution and compare it against the observed distribution $y_i = n_{2i}$. The χ2 value for the difference between the distributions is given as where ν are the degrees of freedom. The probability of embedding is determined by calculating x for a sample from the DCT coefficients. [26]

$$P = 1 - \int_0^{\chi^2} \frac{t^{\frac{(v-1)}{2}} e^{\frac{-t}{2}}}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} \quad \ldots\ (5)$$

Where Γ is the Euler Gamma function. For the case of data that is randomly distributed across the redundant data The previous χ2-test does not detect the hidden data, However, it is possible to extend the χ2- test to be more sensitive to local distortions in an image and to detect the hidden data Using the extended test, we can detect pseudo-randomly distributed hidden data. Instead of increasing the sample size and applying the test at a constant position, we use a constant sample size but slide the position where the samples are taken over the image's entire range [10] Siwei Lyu and Hany Farid suggested a different approach based on discrimination of two classes: stego image and non-stego image Statistics collected from images in a training set determine a function that discriminates between the two classes. The

discrimination function determines the class of a new image that is not part of the training set. The set of statistics used by the discrimination function is called the feature vector. [27]

6-The Proposed Stegnographic Scheme

We propose to use of the Second Least Significant Bit (SLSB) rather than the Least Significant Bit (LSB) of the quantized Discrete Cosine Transformation coefficient of an image is implemented to avoid many known statistical steganalysis scheme and the pseudo code of the suggested algorithm is as follows:

Input: message, cover image
Output: steganographic image containing message
While data left to embed **do**
Get next DCT coefficient from cover image
If DCT ≠ 0 and DCT ≠ 1**then**
Get next LSB from message
Replace DCT **SLSB** with message bit
End if
Insert DCT into steganographic image
End while

An important note is to choose a proper cover image, such as the use of domestic hand paint, and to avoid well known hand painting as a cover image to avoid the comparison of the stego image with the cover image.

7-Conclusion

By embedding the secret data in the second least significant bit (SLSB) rather than the least significant bit (LSB) we will avoid many well known statistical steganalysis, this method can widely used for embedding an acceptable amount of data size. LSB embedding can easily be implemented and do not visually degrade the image to the point of being noticeable. However many statistical steganalysis had been developed to detect the hidden data in the cover media based on the hypothesis of embedding through the least significant bit.

Refrences

[1]. R.J. Anderson and F.A.P. Petitcolas, ***"On the Limits of Steganography,"*** J. Selected Areas in Comm., vol. 16, no. 4, 1998, pp. 474–481.

[2]. H S Manjunatha Reddy, K B Raja***," High Capacity and Security Stegnography Using Discrete Wavelet Transform"***, International Journal of Computer Science and Security (IJCSS), Volume (3): Issue (6).

[3] .Ali K.Hmood, B.B.Zaidan, A.A. Zaidanand Hamid A.Jalab, ***"An Overview on Hiding Technique in Images"***, Journal of Applied Sciences 10(18):2094-2100, 2010.

[4]. Arvind Kumar, Km. Pooja*," **Steganography- A Data Hiding Technique"**,* International Journal of Computer Applications (0975 – 8887) Volume 9– No.7, November 2010.

[5]. A.W.Naji, A.A.Zaidan, B.B.Zaidan, Shihab A, Othman O. Khalifa**, *"Novel Approach of Hidden Data in the (Unused Area 2 within EXE File) Using Computation Between Cryptography and Steganography"***, International Journal of Computer Science and Network Security (IJCSNS) , Vol.9, No.5 , ISSN : 1738-7906, pp. 294-300, May 30 (2009), Seoul, Korea.

[6]. Kefa rabah***, "Steganography the art of hiding data",*** Information Technology Journal 3(3): 245-269, 2004.

[7]. 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.

[8]. F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Information Hiding—A Survey," Proc. IEEE, vol. 87, no. 7, 1999, pp. 1062–1078.

[9]. M. I. Khalil, "Image Steganography : Hiding short Audio Messages within Digital Images", JCS&T vol 11 No.2, October 2011.

[10]. NIELS PROVOS, PETER HONEYMAN***," Hide and Seek: An Introduction to steganography",*** University of Michigan, 2003 IEEE.

[11]. C. Cachin***," An Information-Theoretic Model for Steganography, Cryptology",*** ePrint Archive, Report 2000/028, 2002, www.zurich.ibm.com/˜cca/papers/stego.pdf.

[12]. Khalil Challita and Hikmat Farhat, ***"Combining Steganography and Cryptography: New Directions",*** International Journal on New Computer Architectures and Their Applications (IJNCAA) 1(1): 199-208, 2011.

[13]. Nagham Hamid, Abid Yahya, R. Badlishah Ahmad & Osamah M. Al-Qershi, ***"Image Steganography Techniques: An Overview"***, International Journal of Computer Science and Security (IJCSS), Volume (6): Issue (3): 2012.

[14]. S .K. Moon, R.S. Kawitkar ***," Data Security using Data Hiding",*** International Conference on Computational Intelligence and Multimedia Applications 2007 ,247-251.

[15]. J. Fridrich and M. Goljan, ***"Digital image steganography using stochastic modulation",*** SPIE Symposium on Electronic Imaging, San Jose, CA, 2003.

[16]. A.Nag!, S. Biswas*, D. Sarkar*, P.P. Sar***," A novel technique for image steganography based on Block-DCT and Huffman Encoding",*** International Journal of Computer Science and Information Technology, Volume 2, Number 3, June 2010,103-112.

[17]. Dr. Ekta Walia, Payal Jain, and Navdeep, "***An Analysis of LSB & DCT based Steganographic",*** Global Journal of Computer Science and Technology, Vol. 10 Issue 1 (Ver 1.0), April 2010.

[18]. B.Swathi, K.Shalini, K. Naga Prasanthi, ***"A REVIEW ON STEGANOGRAPHY USING IMAGES",*** Asian Journal of Computer Science and Information Technology 2: 8 (2012)

[19]. Andrew B. Watson**," *Image Compression Using the Discrete Transform ",*** Mathematica Journal, 4(1) , p. 81-88, 1994.

[20]. Wallace, G.**" *The JPEG still picture compression standard"***. Communications of the ACM 34(4): 1991**,** 30-44.

[21]. Chen, W. H., and W. K. Pratt. , ***"Scene adaptive coder",*** IEEE Transactions on Communications COM-32: 225-232. 1984.

[22]. Mr.S. V. Viraktamath, Dr. Girish V. Attimarad, "***Impact of Quantization Matrix on the Performance of JPEG"***, International Journal of Future Generation Communication and Networking Vol. 4, No. 3, September, 2011.

[23]. Mehdi Kharrazi1, Husrev T. Sencar2, and Nasir Memon2***," Image Steganography: Concepts and Practice"***, WSPC/ Lecture Notes Series, 2004.

[24]. Siwei Lyu, ***"Natural Image Statistics for Digital Image Forensics",*** A Thesis Submitted to the Facultyin partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science, DARTMOUTH COLLEGE Hanover, New Hampshire August, 2005.

[25]. A. Westfeld and A. Pfitzmann, ***"Attacks on Steganographic Systems,"*** Proc. Information Hiding—3rd Int'l Workshop, Springer Verlag, 1999, pp. 61–76.

[26]. Hamdy A. Morsy, Zaki B. Nossair, Alaa M. Hamdy, and Fathy Z. Amer, ***"Utilizing Image Block Properties to Embed Data in the DCT Coefficients with Minimum MSE",*** International Journal of Computer and Electrical Engineering, Vol. 3, No. 3, June 2011.

[27]. H. Farid, ***"Detecting Hidden Messages Using Higher- Order Statistical Models,"*** Proc. Int'l Conf. Image Processing, IEEE Press, 2002.