Digital Image Steganography Concept and Evaluation

Seyyed Amin Seyyedi
Department of Electronic Computers, Belarusian State University of Informatics and Radioelectronics
Minsk 220013, Belarus

Rauf.Kh Sadykhov
Department of Electronic Computers, Belarusian State University of Informatics and Radioelectronics
Minsk 220013, Belarus

ABSTRACT

Steganography is the art of hiding the fact that communication is taking place, by hiding information in other information. There are many steganography techniques with different kind of container. Digital image are the most popular and frequency used on the internet. In this article we consider the main image steganography techniques and those advantages and disadvantages. Also attempt to identify the requirements of a good steganography algorithm and evaluation these algorithms.

Keywords

Steganography, image compression, LSB, Palette base LSB, Spread Spectrum, Patch work, Pseudorandom Permutations, DCT, DWT

1. INTRODUCTION

The objective of computer security includes protection of information and property from theft, corruption, or natural disaster, while allowing the information and property to remain accessible and productive to its intended users. Cryptography was created as a technique for securing the secrecy of communication and many different methods have been developed to encrypt and decrypt data in order to keep the message secret. Unfortunately it is sometimes not enough to keep the contents of a message secret, it may also be necessary to keep the existence of the message secret. The technique for solve the problem, is called steganography. [1, 3 , 4]

Steganography refers to the science of invisible communication. Unlike cryptography, where the goal on keeping the contents of a message secret, but the goal of steganography on keeping the existence of a message secret therefore we can use both of them for achieve more security . Table 1 shows comparison of different techniques for communicating in secret. Encryption allows secure communication requiring a key to read the information. An attacker cannot remove the encryption but it is relatively easy to modify the file, making it unreadable for the intended recipient. Digital signatures allow authorship of a document to be asserted. The signature can be removed easily but any changes made will invalidate the signature, therefore integrity is maintained. [4]

Steganography provides a means of secret communication which cannot be removed without significantly altering the data in which it is embedded. The embedded data will be confidential unless an attacker can find a way to detect it.

Table 1 Comparison of secret communication techniques [2]

Communication Techniques	Comparison Parameters		
	Confidentiality	***Integrity***	***Un removable***
Encryption	Yes	No	Yes
Digital signature	No	Yes	No
Steganography	Yes / No	Yes / No	Yes / No

2. An over view to steganography

2.1 Different type of steganography

Almost all digital file formats reusable for steganography, but the type of files have important role in steganography .When files are created there are usually some bytes in the file that aren't really needed (redundant) or at least aren't that important. This area of file can be replaced with the information that is to be hidden, without significantly altering the file. Image and audio files especially comply with this requirement, while research has also uncovered other file formats that can be used for information hiding. Figure 1 shows the different categories of file formats that can be used for steganography techniques. [1, 3]

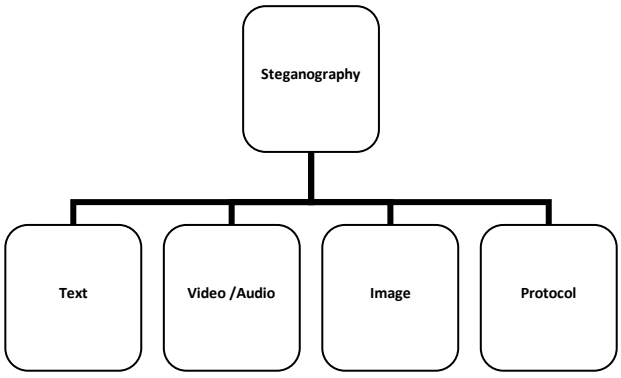

Figure 1 different types of steganography

2.2 Steganography systems

The total structure of steganography system involves the two process embedding and extracting that are showed in figure 2.

The embedding process involves: [4]

- The cover work(cover image) that contains the embedded message.
- Secret message (or image) which could be cipher text

- Stego gramme(Stego image) is the image, which contains the secret message.

An optional stego key or password or digital signature could be utilized to conceal and extract message.

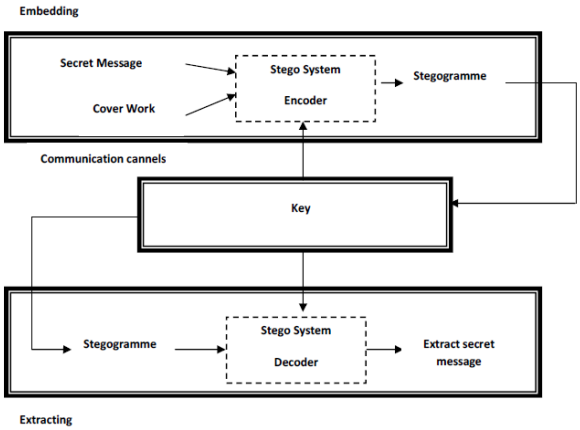

Figure 2 Structure of steganography

3. Image steganography

Secrets can be hidden inside all sorts of cover information. Most stenography utilities nowadays, hiding information inside images because images are popular used in internet. Almost steganography techniques depends on the image format and type of compression, it means that we can applied on various image format with varying of success. [5]

3.1 Classification of image steganography techniques

During the research into the different steganography techniques, there are many different ways keeping important message hiding. Image steganography methods can be divided in two groups, spatial domain methods [5][6] and frequency domain ones [5,6,7].In spatial domain secret message is directly embedded inside the least significant of image while for transform also known as frequency domain, images are first transformed and then the message is embedded in the image.

The spatial domain techniques involve [3, 6]:

- *Substitution system techniques*: Replace redundant or unneeded bits with secret message, Such as: Least Significant Bit (LSB) and palette base image techniques.
- *Statistical method techniques*: Embeds one bit of information only in a carrier and creates statistical change. Such as Pseudorandom Permutation (PP), patch work technique.
- *Spread spectrum techniques*: The stream of information to be transmitted is divided into small pieces .Such as Spread Spectrum (SS).

Frequency domain techniques, hide message data in the transform space of a signal such as Discrete Cosine transform(DCT) , Discrete wavelet transform(DWT) and some techniques is common in two categories involve patch work technique and spread spectrum.

4. Spatial domain techniques

4.1 LSB (Least Significant Bit)

The most frequently used steganography method is the technique of LSB substitution. In a gray-level image, every pixel consists of 8 bits. One pixel can hence display 2^8 =256 variations. The weighting configuration of an 8-bit number is illustrated in Figure 3.

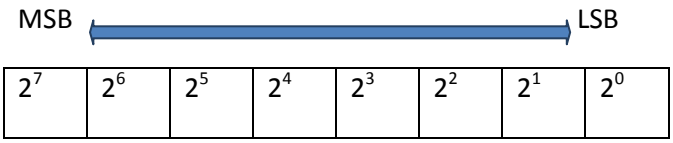

Figure 3 Weighting of an 8-bit pixel

The basic concept of LSB substitution is to embed the confidential data at the rightmost bits (bits with the smallest weighting) so that the embedding procedure does not affect the original pixel value greatly. The mathematical representation for LSB method is: [2]

$$x_i^{'} = x_i - x_i \mod 2^k + m_i \qquad (1)$$

In Equation (1), $x_i^{'}$ represents the i th pixel value of the stego-image, x_i represents the original cover-image, and m_i represents the decimal value of the i th block in confidential data. The number of LSBs to be substituted is denoted as k. The extraction process is to copy the k-rightmost bits directly. Mathematically the extracted message is represented as:

$$m_i = x_i^{'} \mod 2^k \qquad (2)$$

Hence, a simple permutation of the extracted m_i gives us the original confidential data. When using a 24-bit image, a bit of each of the Red, Green and Blue (RGB) color components can be used, since they are each represented by a byte and each byte can to represent 256 different colors. In other words, one can store 3 bits in each pixel. An 800 × 600 pixel image, can thus store a total amount of 1,440,000 bits or 180,000 bytes of embedded data [7]. For example a grid for 3 pixels of a 24-bit image can be as follows:

(00101101 00011100 11011100)

(10100110 11000100 00001100)

(11010010 10101101 01100011)

When the number 200, which binary representation is 11001000, is embedded into the least significant bits of this part of the image, the resulting grid is as follows:

(00101101 00011101 11011100)

(10100110 11000101 00001100)

(11010010 10101100 01100011)

Basically, the modification only happens in three of the underlined bits out of the eight bytes used. Any message that needs to be hidden in the least and second least significant bits and still be invisible to the human eye must occupy half or less than half of the bits of the changed image. (On average, only half of the bits in an image will need to be modified to hide a secret message using the maximum cover size).

In this method some parameters such as image file format, type of image compression (use lossless compression) and bit depth play the important roles. For example the capacity and invisibility is better in image file format that use 24-bit than 8-bit.There are many different approaches that LSB steganography to hide information inside an 8-bit image, but with varying degree of success. The cover images must be able to hide the existence of the message embedded inside it. Such as LSB on one bit, this approach used in GIF and BMP image file format. [6, 7]

Another method LSB on two bit, LSB on three bit, LSB on four bit, in these methods depended to application, we must to do tradeoff between image quality and capacity.

To make much harder detection, we can to use some method in color of RGB image. These methods call LSB on color that the color value in each pixel containing the data should be rotate, such as Red color only method ,middle quarter method.

4.2 Palette base and LSB

Palette-based image [8] includes an image index and the palette itself. The palette is usually consisting of a set of color vectors with an 8-bit unique index attached to each. For example GIF image, popular image format have been used in internet, by definition a GIF image cannot have a bit depth greater than 8, thus the maximum number of colors that a GIF can store is 256. The GIF images are indexed images where the colors used in the image are stored in a palette, sometimes referred to as a color lookup table.

The colors of the palette are typically ordered from the most used color to the least used colors to reduce lookup time. Palette approach used GIF image that should to change only one least significant bit. Its result depended to adjacent palette entries, if there are similar that means that there might be little or no noticeable change in human eyes. Otherwise the change would be evident. For solving this problem we can to use one of the two ways, sort the palette so that the color differences between consecutive colors are minimized (palette order) in this way we can store small amounts of secret information the palette using this method. The disadvantage of this method attacker can simply reorders the color vectors in the palette.

Second way adds new colors which are visually similar to the existing colors in the palette. (This value depends on the bit depth used) the palette base approach is better for gray scale image because In an 8-bit grey scale GIF image; there are 256 different shades of grey [6, 8]. The changes between the colors are very gradual, making it harder to detect.

The colors of the palette are typically ordered from the most used color to the least used colors to reduce lookup time. Palette approach used GIF image that should to change only one least significant bit. Its result depended to adjacent palette entries, if there are similar that means that there might be little or no noticeable change in human eyes. Otherwise the change would be evident. For solving this problem we can to use one

4.3 Pseudorandom Permutations

In this method the secret message bits can be distributed randomly over the whole cover image, these techniques further increase the complexity for an attacker, since it is guaranteed that sub sequent message bits are the same order. We have not restricted the output of the pseudorandom number generator in any way. We call a case collision ,if a collision occurs ,it will possible try to insert more than one message bit into one cover element, thereby computing some of them. To overcome to problem of collision, we could keep track of all cover bits which have already been used for communication in a set A. If during the embedding process one specific cover-element has not been used prior, we add its index to A and continue to use it. If however, the index of the cover element is already contained in A, we describe the element and choose another cover-element pseudo randomly. This technique ensures that subsequent message bits are not embedded in the same order thus making it even harder for an attack to succeed. As LSB, some of data which is stored randomly stored in LSB could disappear also it may generate a high noisy image if it stored large bits in MSBs. For instance figure 4. [9]

Pseudo random Number: 1, 7, 10, 13, 20, 26, 29, 35….

Cover Work *Secret message* *Stego gramme*

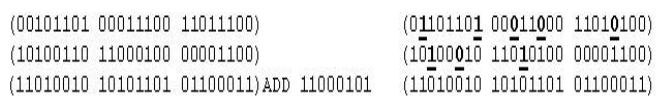

Figure 4 Pseudorandom Permutations method

4.4 Patch work techniques

Patch work is statistical techniques that encode information by changing several statistical properties of a cover image (adds redundancy to the hidden information and then scatters it with Gaussian distribution throughout the image) and use hypothesis testing in extracting process [7, 9].

A secret key is used to randomly select a subset of pixels from an image and then divide it into two distinct sets (patch A and patch B).The brightness of one set of pixels is shifted by a positive number while the brightness of those in the other set are shifted by the corresponding negative number, as show in figure 5.

The contents of the host image are independent of the Patchwork process, the contrast changes in this patch subset encodes one bit and the changes are typically small and imperceptible, while not changing the average luminosity.

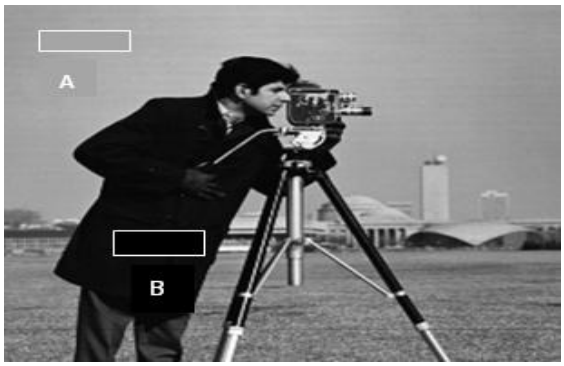

Figure 5 patch work method with two areas

In spread spectrum techniques, hidden data is spread throughout the cover-image, making it harder to detect .A system proposed by Marvel et al. Combines spread spectrum communication, error control coding and image processing to hide information in images .Spread spectrum communication describes the process of spreading the bandwidth of a narrowband signal across a wide band of frequencies .This can be accomplished by modulating the narrowband waveform with a wide bad waveform, such as with noise. After spreading, the energy of the narrowband signal in any one frequency band is low and therefore difficult for detection. When used for image steganography, the information will be embedded with noise before it is combined with the cover image. The signal of the embedded information is much lower than the cover image which makes it harder to recognize with the naked eye. Without the help of the original image, even computer analysis will have a hard time detecting the hidden information [9- 12].

5. Transform Domain Techniques

Some of the techniques are common in spatial and transform domain such as spread spectrum and patch work techniques.

For frequency domain methods [16-19], the first step is to transform the image data into frequency domain coefficients by some mathematical tools (e.g. FFT, DCT, or DWT). Then, according to the different data characteristics generated by these transforms, embed the secret message into the coefficients in frequency domain. After the embedded, coefficients are transformed back to spatial domain; the entire embedding procedure is completed. The advantage of this type of steganography is the high ability to face some signal processing or noises. However, methods of this type are computationally complex and hence slower.

To understand the steganography algorithms that can be used when embedding data in the transform domain, one must first explain the type of file format connected with this domain. One of the popular format that use in internet JPEG because the small size of images .This type of image use lossy compression (Compression attempts to reduce one or more of these redundancy types) [13].

5.1 Discrete Cosine Transform

One of the major characteristics of steganography is the fact that information is hidden in the redundant bits of an object and since redundant bits are left out when using JPEG it was feared that the hidden message would be destroyed (used lossy compression), Thus it is important to recognize that the JPEG compression algorithm is actually divided into lossy and lossless stages. The DCT and the quantization phase form part of the lossy stage, while the Huffman encoding used to further compress the data is lossless. Steganography can take place between these two stages. This method exploited human eyes system for compression.

Basically the idea behind JPEG steganography is described below in six steps while figure 6 shows the process: [15, 18]

1. Load a color image (bitmap format 24 bits), and part the colors into the red, green, and blue.
2. Convert the image formula from RGB to YCbCr
3. Every color plane is divided into 8x8 blocks(512x512 image we have 64 blocks)
4. Quantize these blocks with quantization coefficients. The DCT coefficients are divided by their corresponding quantization coefficients (quantization table) and rounded to the nearest integer
5. Entropy encoding
6. IDCT

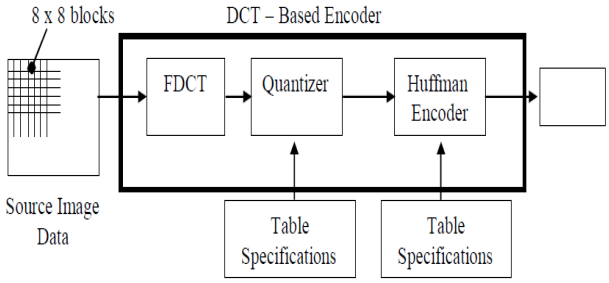

Figure 6 process of DCT in JPEG image format

5.2 Discrete Wavelet Transform

The Wavelet Transform (WT) has gained widespread acceptance in signal processing and image compression. Because of their inherent multi-resolution nature, wavelet-coding schemes are especially suitable for applications where scalability and tolerable degradation are important. Recently the JPEG committee has released its new image coding standard, JPEG-2000, which has been based upon DWT [19].

Despite all the advantages of JPEG compression schemes based on DCT namely simplicity, satisfactory performance, and availability of special purpose hardware for implementation; these are not without their shortcomings. Since the input image needs to be ``blocked" correlation across the block boundaries is not eliminated. This results in noticeable and annoying `blocking' artifacts" particularly at low bit rates as shown in Figure 7, 8.Lapped Orthogonal Transforms (LOT) [16, 17] attempt to solve this problem by using smoothly overlapping blocks. Although blocking effects are reduced in LOT compressed images, increased

computational complexity of such algorithms do not justify wide replacement of DCT by LOT.

Figure 7 Original Lena image

Figure 8 reconstructed Lena with DC component only to show Blocking artifacts

6. Evaluation the steganography techniques

All the steganography algorithms have varying strong and weak points. It is important to ensure that one uses the most suitable algorithm for an application. There is several criteria condition that we can to compare and select suitable the steganography algorithms. All steganography algorithms have to comply with a few basic requirements. The most important requirement is that a steganography algorithm has to be imperceptible. Now we consider these parameters [7, 9, and 20]:

Invisibility or Perception (Inv): First and mainly requirement of steganography algorithms which Human eyes cannot distinguish the difference between the original image and the stego-image (the image with confidential data embedded in). This requirement depended on the size of the secret message and format image and type of the cover image.

Capacity (Cap): This requirement expresses the amount of hide data that depended on image file format.

Robustness: The embedded data should endure any reprocessing operation that cover may be subjected to and still remain intact. There are two type of robustness namely robustness again statistical attack and robustness again manipulation attack.

Robustness Against Statistical Attack (RASA): Statistical stego analysis is the practice of detecting hidden information through applying statistical tests on image data (compare the frequencies). Many steganography algorithms leave a 'signature' when embedding information that can be easily detected through statistical analysis. To be able to pass by a warden without being detected, a steganography algorithm must not leave such a mark in the image as be statistically significant.

Robustness Against Manipulation Attack (RAMA): This type of attack work when the embedding algorithm causes noticeable artifacts in stego image. Such as cropping, rotating. Depending on the manner in which the message is embedded, these manipulations may destroy the hidden message.

Detect ability or Un-suspicious (Det): This is an important criterion that determines the success of a technique as a result of the complexity involved in detecting the hidden data in the carrier [2, 3]

*Domain type (Dom): T*his parameters indicate the type of steganography method (spatial or transform domain) .spatial domain are easy to implement and fast whereas transform domain methods are secure and slow.

*Independent of File Format (FF): T*here are several image file format that used in internet. Always the same image file format continuously communicated between two partners. It might seem suspicious and attackers try to change type of image file format.

The following table 2, we compare the steganography algorithms as discussed according the above criteria. In this table we have three levels defined as high, medium, low (H, L, and M). These levels show the degree of satisfied requirements. A high level indicates that the algorithm satisfied all requirements and the low level indicates that the algorithms have some weekends in this requirement and medium level means that the requirement depends to parameters. For example invisibility in LSB palette base depended on the type of file. This algorithm has better performance in gray scale image than another type of image.

The ideal, steganography algorithm would have a high level in every requirement. Unfortunately algorithms that are evaluated here, there is not one algorithm that satisfies all of the requirements. Thus a trade-off will exist in most cases, depending on which requirements are more important for the specific application.

6.1 LSB on BMP

The image file format BMP (raster image) are uncompressed, hence they are large. Unfortunately to be able to hide a secret

Table 2 Comparison Steganography algorithms

Al /Req	Inv	Cap	RASA	RAIM	Dec	FF	Dom
LSB(BMP)	H	H	L	L	L	L	S
LSB(JPEG)	H	H	L	L	H	L	S
LSB(palette)	M	M	M	L	M	L	S
PP	M	L	M	L	H	H	S
PW	H	L	H	H	H	H	S/T
SS	H	L	H	M	H	H	S/T
DCT	H	L	M	M	H	L	T
DWT	H	H	M	M	H	L	T

message inside a BMP file, one would require a very large cover image. Nowadays, BMP image are not often used on the internet and might arouse suspicion, therefore this technique easy detectable and don't satisfied robustness condition. Image pixel generally stored with a color depth of 1, 4, 8, 16, 24, 48 or 64 bits per pixel therefore a BMP is capable of hiding quite a large message, but the fact that more bits are altered results in a larger possibility that the altered bits can be seen with the human eye. (Trade-off between invisibility and amount of embedded information)

6.2 LSB on JPEG

The JPEG image format support 8 bits per color(RGB) for a 24 bit total ,therefore JPEG is capable for hiding a large message .Depended on the replacing method (color cycle) this technique make harder detection .The JPEG file format commonly used of lossy compression and during the compression process might to lost secret message hence not satisfied robustness condition.

6.3 LSB on palette

The GIF image limited an 8-bit palette or 256 colors, therefore LSB steganography to hiding information inside an-8bit image but with varying degree of success, because it's depended on the number of bit for used to hiding information. There are trade-off between security and invisibility. According to section 4-2, results after altered bit's depended to adjusted palette entries, in this moment type of image (color or gray scale) play very important role, as detect ability and RASA are depended to type of image and other disadvantage of this method attacker can simply reorder the colors in the palette.

6.4 Pseudorandom Permutation

This technique ensures that subsequent message bits are not embedded in the same order thus making it even harder for an attack to succeed. As LSB, some of data which is stored randomly stored in LSB could disappear also it may generate a high noisy image, if it stored large bits in MSBs. Type of image file format is not important in this techniques.

6.5 Patch work techniques

A disadvantage of the patchwork approach is that only one bit is embedded. One can embed more bits by first dividing the image into sub-images and applying the embedding to each of them [20]. The advantage of using this technique is that the secret message is distributed over the entire image, so should one patch be destroyed, the others may still survive. This however, depends on the message size, since the message can only be repeated throughout the image if it is small enough. If the message is too big, it can only be embedded once. The patchwork approach is used independent of the host image and proves to be quite robust as the hidden message can survive conversion between lossy and lossless compression.

6.6 Spread Spectrum

The Spread Spectrum technique spreads the all hidden data over the cover image. This will make it harder to detect. The signal of the embedded information is much lower than the cover image which makes it harder to recognize with the naked eye. This method robustness again attacks

6.7 Discrete Cosine Transform

Transform domain methods hide message in significant area of cover image which make than more robust than time domain (LSB), compression, cropping and some image processing. However, a trade-off exists between the amount of information and robustness. Embedding information in DCT domain is simply done by altering the DCT coefficient. Coefficients are equal to zero, and changing too many zero to non-zeroes value will have an effect to compression ratio, therefore capacity in DCT less than LSB

6.8 Discrete Wavelet Transform

Embedding in DWT domain shows promising result, and outperforms DCT embedding especially in term of compression survival. Its multi resolution capability decomposes a signal in narrow levels of details that may help to embed the same data in multi-level coefficients and consequently provides good resiliency against various image impairments.

7. Conclusion and future works

7.1 Conclusion

The main steganography techniques were discussed in this paper. There are the several selection approaches to hiding information. The major file formats have different methods of hiding information with different advantages and disadvantages. The spread spectrum method almost satisfied all the requirement and robustness against statistical attacks but this method has low capacity. Based on the application, we can to select the suitable algorithm for hiding information.

Hiding information in transform domain secure than spatial domain. Discrete wavelet transform is a new approach and for hiding information because allows good localization in time and spatial frequency domain, better identification of which data is relevant to human perception.

7.2 Future works

We propose a new steganography technique which embedded the secret message in frequency domain. Creation new algorithm to improve embedding capacity and image quality.

8. REFERENCES

[1] Sara Khosravi, Mashallah Abbasi Dezfoli, Mohammad Hossein Yektaie. 2011 A new steganography method based HIOP (Higher Intensity Of Pixel) algorithm and Strassen's matrix multiplication, Journal of Global Research in Computer Science.

[2] R.Popa. An analysis of stegonagraphic techniques.1998. The Politechnica university of Timisoara .

[3] S. Katzenbeisser, F.A.P. Petitcolas.2000. Information Hiding Techniques for Steganography and Digital Watermarking, Artech House.

[4] Chang, C. C. and Chuang, L. Z. 2004. Introduction to the Visual Cryptography, Communication of the Chinese Cryptology and Information Security Association (CCISA) 1-14.

[5] G. J. Simmons.2002. "The prisoners' problem and the subliminal channel" in Proc. Advances in Cryptology (CRYPTO '83), 51-67.

[6] Arash Habibi Lashkari.2011. A survey on image steganography algorithms and evalution. Communication in computer and information science.

[7] Michiharu Niimi, Hideki Noda, Eiji KaWaguchi, Richard Eason.2002. "High Capacity and Secure Digital Steganography to Palette-Based Images", IEEE.

[8] T. Morkel, J.H.P. Eloff, M.S. Olivier.2002. An Overview of Image Steganography, University of Pretoria, South Africa.

[9] Bender, W., Gruhl, D., Morimoto, N., and Lu, A. 1996. Techniques for data hiding. IBM Systems Journal, 313-336.

[10] W. Bender, D. Gruhl, N. Morimoto, A. Lu.1996. Techniques for data hiding. IBM Systems Journal.

[11] Frederick Brundick and Lisa Marvel.2001. Implementation of Spread Spectrum Image Steganography. Army Research Laboratory,ARL-TR-2433.

[12] Currie, D.L. & Irvine, C.E. 1996.Surmounting the effects of lossy compression on Steganography. 19th National Information Systems Security Conference.

[13] Johnson, N.F. & Jajodia, S.1998. Exploring Steganography: Seeing the Unseen. Computer Journal.

[14] Meenu Kumari, A. Khare, Pallavi Khare.2010.JPEG Compression Steganography & Crypography Using Image-Adaptation Technique, Journal of Advances in Information Technology.

[15] Ahmed A. Abdelwahab and Lobna A. Hassaan.2008.A Discrete Wavelet Transform based technique for image data hiding . 25th National Radio Science Conference (NRSC)

[16] V.Srinivasa rao, Dr P.Rajesh Kumar, G.V.H.Prasad, M.Prema Kumar, S.Ravichand.2010. Discrete Cosine Transform Vs Discrete Wavelet Transform: An Objective Comparison of Image Compression Techniques for JPEG Encoder. International Journal of Advanced Engineering & Applications.

[17] Rao, K. R. and Yip, P.1990. Discrete Cosine Transforms - Algorithms, Advantages, Applications, Academic Press.

[18] Arne Jense and Anders la Cour-Harbo. 2001.Ripples in Mathematics: the Discrete Wavelet Transform.Springer.

[19] Bushra Kassim Al-Abudi.2002. "Colour Image Data Compression Using Multilevel Block Trunsaction Coding Technique", Phd Thesis, College of Science, University of Baghdad.

[20] W. Bender, D. Gruhl, N. Morimoto, A. Lu.1996 Techniques for data hiding. IBM Systems Journal.