Image Steganography: A Research Paper

VaibhaviSushil, M.Tech Scholar, Department of Computer Science & Engineering, Buddha Institute of Technology GIDA, Gorakhpur, (U.P.) India

AbhishekShahi, Assistant Professor, Department of Computer Science & Engineering, Buddha Institute of Technology GIDA, Gorakhpur, (U.P.) India

ABSTRACT

Computer based communications are nowadays are at threshold of making our lives easier in the world, from exchanging electronic documents, to communicating with each other , to sharing information, and to check the bank balances and pay bills. Information security is an essential part that must be taken into consideration to ensure communication .Steganography is the technique that restricts the unauthorized users to have access to an important data. By using Steganography, the information can be hidden in the carrier items such as images, sound files, text files, videos while performing the data transmission. Therefore, this is a research paper for the various studies done on steganography. It also analyzes the work that have been conducted on steganography and clarifies the strength and weakness points on each work separately.

KEYWORDS : *Steganography, Types of steganography, Techniques of steganography, Image steganography and types*

Introduction

People use many different techniques to hide the elements and information they deemed valuable from past to present. Steganography of one of them. Steganography is not a new work area[1]. The first use of steganography was originated in 440 BC. Internet is generally used for data transfer. It has significantly become an important part in today's life. In fact. It has become increasingly important to secure the information exchanged while making the use of cyber space [2], for many business industries, government organization and for individuals. The reason why questions of online privacy and cyber security have raised the issue of secret writing into limelight is fact that today our social, economic and professional lives are completely dependent on emailing, net posting, e-commerce, electronic banking etc. [3]. Steganography wonders hiding the presence of secret information. It is the alternative of cryptography, where alternations are made to messages. Hence steganography and cryptography are cousins in spy craft family. In cryptography, the messages are scrambled which makes it unintelligible, while on the other hand steganography makes a message invisible by hiding it [4]. In steganography, the secret messages are kept out of sight by the use of a cover medium (i.e., carrier) before it is passed on a public communication channel. It therefore, obstruct the unauthorized access to the messages and also it protects its confidentiality. The secret messages can be encrypted or compressed before the application of steganography increases the security level and reduction in the amount of data to be embedded [1]. This paper represents the various steganography techniques that have been proposed recently for hiding the secret information within the cover-images efficiently. It is a high security technique for long data transmission.

The goal of image steganography is to hide a secret message inside an image. In a typical scenario, the sender hides the secret message inside the cover image and transmits it to the receiver who recovers the message.

Traditional approaches to image steganography are only effective up to a relative consignment of around 0.4 bits per pixel. Beyond this, they look after to introduce artifacts that can be detected easily by automated steganography analysis tools and in most of the cases by human eye. With the arrival of deep learning, in past decades, a new class of image steganography approaches is emerging. These approaches use neural network as either component in a traditional algorithm, or as end-to-end solution and it takes in a cover image and a secret image and then combine them into a steganographic image. All these attempts have proved that deep learning can be used for practical and to end image steganography, and the embedding rates can be easily achieved competitively with those managed through traditional techniques.

Image steganography-

As stated before, images are the most popular cover objects used for steganography. Many different kinds of image file formats exist, in the domain of the digital image, and most of them are for specific application. For all these different types of image file formats, different steganographic algorithm exist.

Image definition-

To a computer, an image is a collection of number that add up to different light intensities in different areas of the image [1]. This numeric representation forms a grid and the individual points are known as pixels. Most of the images available on the internet consists of a rectangular map of the image pixels where each pixel is located and its color [2].

The number of bits in a color scheme is called the bit depth. It refers to the number of bits used for each pixel [3]. In the current color scheme, the smallest bit depth is 8, means that are 8 bits that are used to describe the color of each pixel [3]. Monochrome and greyscale image uses 8 bits for each pixel and they are able to display 256 different shades of grey. Typically, the digital colored images are in 24 bit files and they are RGB color model, known as true color [3]. All the color variations of a 24 bit images are derived from the three primary colors: Red, Green and Blue, where each primary color is represented by 8 bits [1]. Thus, in one given pixel, there are 256 different quantities of red, green, and blue and by adding up more than 16 million combinations, results in more than 16 million color [3].

Image compression –

When working the larger images of greater bit depth, then the images tend to become too large to pass on over a standard internet connection. In order to display an image in sensible amount of time, to reduce the images file size techniques must be incorporated. To analyze and condense image data, these techniques makes the use of mathematical formulas to result in smaller file sizes. This process is called compression [2].

There are two types of compression in image steganography:

1- Lossy Compression
2- Lossless Compression [4].

Both the methods are used to save storage space, but the procedures that they implement are different.

Lossy compression –

Lossy Compression are used for creating smaller files by throwing out the excess amount of image data from the original image. It detaches the details that are too small for the human eye to differentiate [2], resulting in close approximation of the original image, although not an exact duplicate. An example of an image format that uses this compression technique is JPEG (Joint Photographic Experts Group) [1].

Lossless Compression –

On the other hand, it never removes image information from the original image, but instead of this it is used to represent the data in mathematical formulas [2]. The rectitude of the original image's is maintained and the decompressed image output is bit by bit identical to the original image input [4]. The most popular image formulas that are used by lossless compression is GIF (Graphical Interchange Format) and 8 bit BMP (a Microsoft Windows Bitmap File) [1].

In choosing which steganographic algorithm to be used compression plays a very important role. Lossy compression techniques results in smaller image file sizes, but it is used to increase the possibility that the embedded message may be partly lost due to the fact that the excess amount of image data will be removed [5]. although it does not compress the image to such a small file size [1]. For both of these compression types different Lossless compression although, is used for keeping the original digital image complete without the chance of lost, steganographic algorithm have been developed.

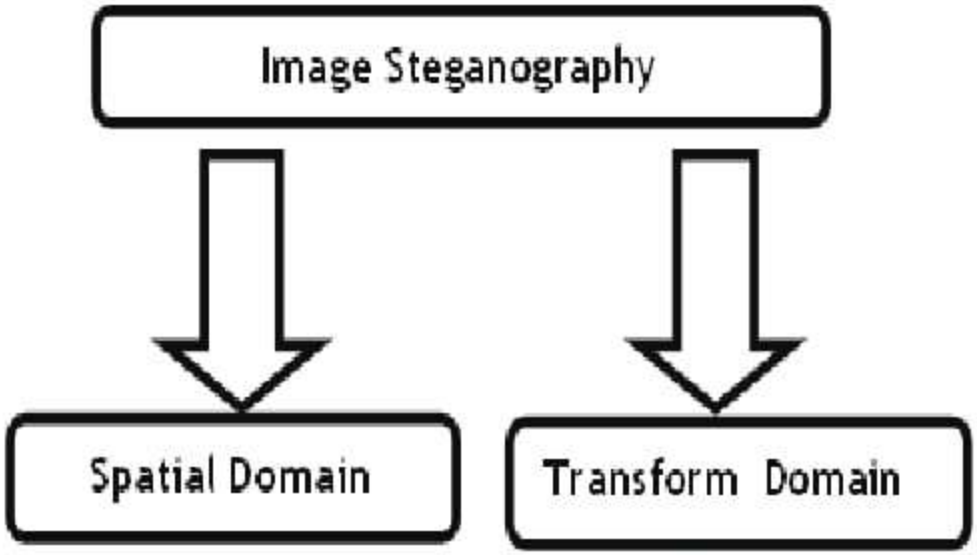

Figure 2: Image Steganography Classification

IMAGE AND TRANSFORM DOMAIN-

Image steganography techniques can be categorized into two groups- those in the image domain and those in the transform domain [6]. Image also knows spatial- domain techniques submerge the message in the intensity of the pixels directly, while the transform- also known as frequency- domain, images are first transformed and then the messages are submerged in the image [7].

Image domain techniques enclose bit wise methods that apply bit in section and noise manipulation and are something characterized as "simple systems" (IT). This image format technique which are most suitable for image domain steganography are lossless and the techniques used are typically dependent on image format [8].

In the transform domain, steganography involves the manipulation of algorithm and image transform [9]. These methods are used for hiding the message in more significant areas of the cover image making it moresturdy[10]. Many transform domain methods are independent of the image format and the embedded message may survive changes between the lossy and the lossless compression [8].

IMAGE DOMAIN-

i- LSB (Least Significant Bit)- Least Significant Bit insertion is a common, simple approach used for embedding information in a cover image [2]. The least significant bit is also known as 8th – bit of some or all bytes, inside an image is changed to a bit of the secret message. When we use a 24 bit image a bit of each of the red, blue and green color compression can be used, since each of them are represented by a byte. In other words 3 but can be stored by each pixel.

For example- a grid for 3 pixels of a 24 bit image can be represented as follows-

100101101 00011100 11011100
110100110 11000100 00001100
111010010 10101101 01100011

If the number 200 whose binary representation is 11001000, is submerged into least significant bit of this part of the image then the resulting grid is as follows-

100101101 00011101 11011100
110100110 11000101 00001100
111010010 10101100 01100011

Although the number was embedded into the first 8 bytes of the grid, only the underlined 3 bits are needed to be changed according to the embedded message. On average, only half of the bits in an image will be required that is to be modified to hide a secret message using the maximum cover size [11]. Since there are total 256 possible number of intensities of each primary color, by changing the LSB of a pixel results in small changes in the intensity of colors. These changes cannot be recognized by the human eye, therefore the message is successfully hidden.

ii- LSB and Palette Based Images- Palette based images for example, GIF are another popular image file format that are being commonly used on the internet. By definition, a GIF image cannot have a greater depth than 8. Thus a GIF can store 256 maximum number of colors. GIF images are indexed image where the colors that are being used in the images are stored in a palette, which is sometimes referred to as a color lookup table [2]. Each pixel is represented as a single byte and the pixel data are represented as an index to the color palette [1]. The colors of the palette are typically the colors of the palette are ordered from the most used to the least used colors to reduce the lookup time [8].

GIF image can also be used for LSB steganography, although it needs some extra care that should be taken. The problem with the palette approach used with the GIF image is that one should change the least significant bit of a pixel. Since the index to the color palette is changed it results in a completely different color (IT). One of the possible solution is to sort the palette so that the color differences between consecutive colors are minimized [12]. Another solution is to add up the new colors which are visually similar to the existing colors in the palette. It requires the original image to have less unique colors than the maximum number of colors [4]. Using the greyscale image is the final solution to the problem. There are 256 different shades of the grey in an 8-bit greyscale GIF image.

TRANSFORM DOMAIN-

To study how the steganography algorithm can be used when we are embedding the data in the transform domain. Firstly, one must explain the type of file format considered with the domain. The most popular image file format on the internet is the JPEG file format, because of the small size of the image.

JPEG compression-

To transform an image into JPEG format, the RGB color representation is firstly converted to a YUV representation. In this representation, the component Y seems same to the brightness and the U& V component stand for chrominance [4]. According to the research the human eye is more tactful to the changes in the luminance of a pixel than to change in its color [13]. This fact is utilized by the JPEG compression by downsampling the color data to turn down the size of the file. The color components (U & V) are halved in horizontal into vertical direction, thus decreasing the file `size by a factor of 2 [4]. Actual transformation of the image is the next step. For JPEG, the Discrete Cosine Transform (DCT) is used, but similar transforms are for example the Discrete Fourier Transform (DFT). These mathematical transforms are used for converting the pixels in such a way that they give the effectof "spreading" the location of the pixel values over the parts of the image [13]. The DCT is used for transforming a signal from an image representation to a frequency representation, by grouping the pixels into 8x8 pixel blocks and transforming the pixels blocks into 64 DCT coefficient each [11]. A change of a single DCT coefficient will affect all the 64 image pixels in that block.

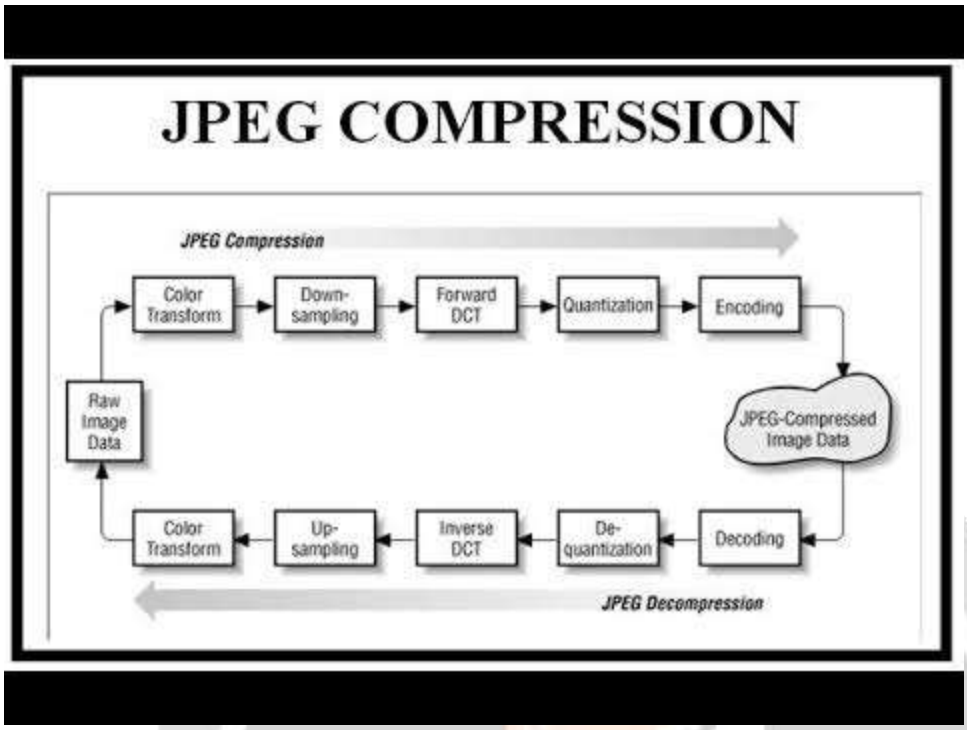

Fig - JPEG compression

Quantization is the next phase of the compression. Exploitation is the another biological property of the human eye. The human eye is equitably good at distinguishing the small differences in brightness over corresponding to a large area, but it is not so good as to differentiate between different strengths in a high frequency luminance [4]. This means that the strength of greater frequencies can be vanished, without changing the presence of the image. JPEG does this by dividing all the values in a block by a quantization coefficient. The results are rounded to integer values and for further reducing the size the coefficients arte encoded using Huffman coding [13].

JPEG Steganography- Originally, it was thought that steganography uses the lossy compression that results in parts of the image data being altered, it would not be possible to use steganography with JPEG image. One of the major characteristics of steganography is the fact that the information is concealed in an inessential bits of an object and since redundant bits are left out when JPEG is used. It was dread that the concealed message would be destroyed. However, the properties of compression algorithm have been used completely in order to develop a steganographic algorithm for JPEGs.

One of these properties of JPEG have been exploited to make the validation in the image invisible to the human eye. During the DCT transformation phase of the compression algorithm rounding error appears in co-efficient data that are not recognizable [1]. Although this property is that which classifies the algorithm as being lossy, this property can also be used for hiding the message.

It is neither impossible nor possible to embed information in an image that uses the lossy compression. Since the compression will destroy all the information present in the process. Thus it is important to recognize that the lossy and lossless stages are actually two stages of JPEG compression algorithm. The lossy stage is formed by the DCT and the quantization phase while in the lossless stage Huffman encoding is used for compressing of data.Steganography can take place between these two stages.

Image or Transform Domain-

Some stegnographic algorithm can either be categorized as being in the image domain or in the transform domain depending on the implementation.

Patchwork-It is the statistical technique that uses the redundant pattern encoding for embedding a message in an image [1]. The algorithm adds up the redundancy to the hidden information and then disperse it throughout the image [8].

A pseudo random generator is used for selecting the two areas of the patches patch A and patch B [14].All the pixels in the patch A is lightened up while in the patch B the pixels are darkened. In the other words, the pixel intensities are increased by the constant value, while the pixels of other patches are decreased with same constant value [3].The disparity changes in this patch subset encodes one bit and the changes are typically small and undetectable, while not changing the average glow [8].

A disadvantage of patchwork approach is that only one bit is embedded. One can embed more bits by first dividing the image into sub-image and then applying the embedding to each of them [15].The advantage of the patchwork approach is that we can distribute the secret message over the entire message, so that one patch should be destroyed, and the others may survive still [8].

The patchwork approach is used independently of the host image and proves to be quite vigorous as a hidden message can survive transformation between lossy and lossless compression.

Spread Spectrum-

In spread spectrum techniques, the hidden data are spread throughout the cover image making it harder to detect [9]. A Marvel et al. proposed a system that is used for combining the spread spectrum communication error control coding and the image processing for hiding the information in the image [16].

Spread spectrum communication can be defined as the process of spreading the bandwidth of a narrowband signal across a wide band of frequencies [16]. This process can be made by adjusting the narrowband waveform with a wideband waveform, such as white noise. After spreading, the energy in any one frequency band of the narrowband signal is low, and therefore it is difficult to detect [16]. In the case of spread spectrum image steganography the most is submerged in noise and then it is combined with the cover image to produce steganography image. The embedded image is not perceptible to the human eye or by computer analysis, because the power of the embedded signal is much lower than the power of the cover image without access to the original image [16].

Literature Review-

A research in [17] has proposed a updated image steganography method that is based on LSB technique in the area of the spatial domain. The methods introduced in it expresses a secret data in six bits of binary format by applying of LS Braille method rather than applying the American Standard Code for Information Interchange (ASCII) format. There methods covers the three bits of a secret data over one pixel of a true image that it collects three reasonable layers namely, red layer, blue layer, and green layer. In this method, the two binary bits are hidden in the green layer of a pixel. In addition to this, a secret data is hidden using second and the third LSB alongside with the least significant bit (LSB) of the blue layer. During the hiding procedure or process, only one bit of the blue layer is handled and changed as well. Their research problem was to improve the security of LSB steganographic technique. They worked on it and they dealt with the secret message and the cover image and the secret message as an input, and then converted each byte of the secret message into its binary format through using secret message is expressed in 6 bits only, as shown in figure 2.

Further, they converted the cover image into the three coherent layers: red layer, blue layer, and green layer. Then, the pixels of each green layer and the blue layer is represented by its binary format through the use of ASCII encoding format. In this method, they began with the blue layer and then they move on to the green layer of the pixel and so forth till the whole of the secret message is embedded. Two bits are utilized for embedding the blue layer. Besides, the message is hidden using the second and the third LSB is hidden alongside with the LSB. During each control of hiding merely 1 bit of the blue layer will be permitted to be changed by manipulating the last three binary bits of the blue layer in the pixel through the following equation-

b1 XOR b2 = r2----(1)

b2 XOR b3=r3-----(2)

where, b1 is the first least significant bit

b2 is the second least significant bit

b3 is the third significant bit

In this method, if the results r2 and r3 of a secret data are similar to the two bits, then it remains unchanged in the pixel. While they are different from those output, it alters merely single bit of cover image pixels. However, the method has some limitations. The limitation is that it provides a limited space for embedding a message, since it uses only 3 bits of message; 2 bits are hidden in blue layer and 1 bit is hidden in green layer.

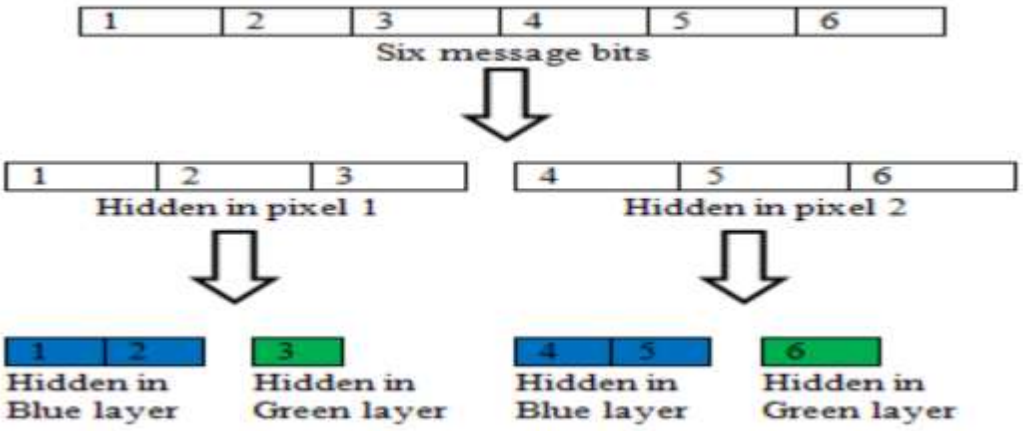

FIG 2-THE SECRET MESSAGE BIT

An additional research [18] presents a secured model for communication using image steganography. The main goal was to design and develop a tool named IMG Stego using Java Programming Language. It conceals a secret data in images based on LSB algorithm like 1- LSB and 2-LSB. IMG Stego tool offers to the end user two process operation:-embedding a secret data into image files as shown in figure 3 and extracting the embedded data from image files through using 1-LSB or 2-LSB algorithm. The IMG Stego tools is used for embedding the secret data on static color; based image that was BMP & PNG format. LSB replacement is used in their system to embed a secret message into an image file [19]. The IMG Stego Java based tools is used for altering the pixels of a cover image so that the least single binary bit of each byte is altered to a bit of a secret data, which is known as standard LSB or 1-LSB. Besides this, the 2-LSB which is different from standard LSB is utilized for allotting further data that is to be embedded over a cover image. Nonetheless, the IMGStego based tool is limited to small data size that it hides into image. the disadvantage of this proposed research is that it is restricted merely to BMP, PNG image format and it uses a sharable text key. Also, it does provides an encryption neither for secret message nor for a sharable key.

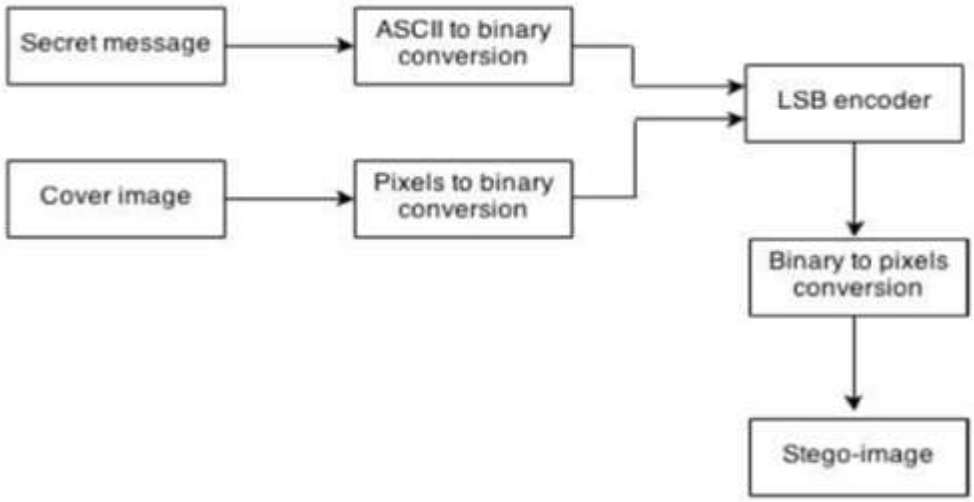

FIG 3-THE EMBEDDING PROCESS

As only a limited space is available for embedding a message. I have worked on increasing the space so that the message can be embedded in large number of space. The modification in the pixels of an image is done ; also the function is created enter data pixels in the image. GUI loop is also used. Tkinter and PIL mode are also learned.

EVALUATION OF DIFFERENT TECHNIQUES-

All the above mentioned algorithms with respect to image steganography are not void of weak and strong points. Consequently, it is important to decide the most suitable approach to be applied. As defined before, there are several parameters to measure the performance of the steganographic system. Fridrich in Fig. 2 shows the relationship between three parameters [21]. These parameters are as follows: •

Undetectability (imperceptibility): this parameter is the first and the primary requirement; it represents the ability to avoid detection, i.e., where the human eye fail to notice it. However, the techniques that do not alter the image in such a way to be perceptible to the human eye may still alter the image in a way that it is detectable by the statistical tests. Truly secure steganographic techniques should be undetectable neither by the human eye nor by the statistical attacks.

• Robustness: it is the second parameter that measures the ability of the steganographic technique to survive the attempts of removing the hidden information. Such attempts include, image manipulation (like cropping or rotating), data compression, and image filtering. Watermarks are an example of a robust steganographic technique

• Payload capacity: it is the third parameter that represents the maximum amount of information that can be hidden and retrieved successfully. When compared with watermarking, that requires embedding only a small amount of copyright information, steganography is seen to hide communication and consequently a sufficient embedding capacity is required. Accordingly and by using this parameter, small amounts of data could be hidden without being detected by the human eye. Larger amounts of information, on the other hand, may detect artifacts by the HVS or statistical tests.

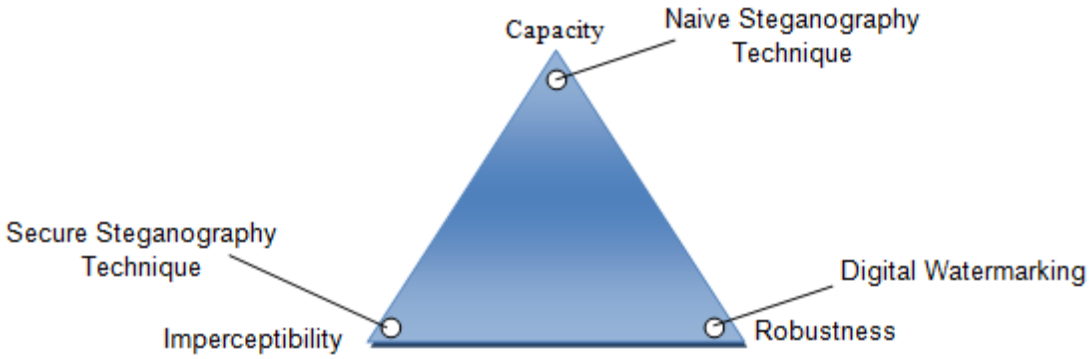

FIG -Competing factors in steganographic systems

Conclusion-

This paper presents a literature review on a variety of different methods algorithm and scheme in image steganography area in order to analyse and investigate them. The main steganographic techniques for both lossy and lossless image such as JPEG & BMP are reviewed in the paper.

All the major image file formats have different methods of hiding messages, with different strong and weak points respectively. Where one technique lacks in payload capacity, the other lacks in robustness. For example, the patchwork approach has a very high level of robustness against most type of attacks, but can hide only a very small amount of information. Least significant bit (LSB) in both BMP and GIF makes up for this, but both approaches result in suspicious filesthat increase the probability of detection when in the presence of a warden.

References –

1- Johnson, N.F. &Jajodia, S., "Exploring Steganography: Seeing the Unseen", Computer Journal February 1998.
2- "Reference guide: Graphics Technical Options and Decisions",Computer Journal February 1998.
3- Owens, M., "A Discussion of covert channels and steganography", SANS Institute, 2002.
4- Moerland, T., "Steganography and Steganalysis", Leiden Institute of Advanced Computing Science.
5- Dunbar, B., "Steganograpic techniques and their use in an open- systems environment", SANS Institute, January 2002.
6- Marvel, L.M., Boncelet Jr., C.G. &Retter, C., "spread Spectrum steganography", IEEE Transactionson image processing, 8:08, 1999.
7- Lee, Y.K. & Chen, L.H., "High capacity image steganographic model", visual Image signal processing, 147:03, June 2000.
8- Venkatraman, S., Abraham, A. &Paprzycki, M., "Significance of Steganography on data security", Proceedings of the International Conference on information technology: coding and computing, 2004.
9- Johnson, N.F. &Jajodia, S., "Steganalysis of images created using current steganography software", Proceedings of the 2nd Information Hiding Workshop, April 1998.
10- Wang, H &wang, S, "cyber warfare: steganography vs. steganalysis", communications of the ACM, 47:10, October 2004.
11- Krenn, R., "steganography and steganalysis", IBM Systems and journal, vol. 33, 1997.
12- Chandramouli, R., Kharrazi, M. &Memom, N., "Image steganography and steganalysis: Concepts and practice", Proceedings of the 2nd international workshop on digital watermaking, October 2003.
13- Currie, D.L. & Irvine, C.E., "Surmounting the effects of lossy compression on syeganography", 19th national information systems security conference, 1996.
14- Handel, T. &Sandford, M., "hiding data in the OSI network model", proceedings of the 1st international workshop on information hiding, June 1996.

15- Petitcolas, F.A,P., Anderson, R.J. & Kuhn, M.G., "Information hiding – a survey", proceedings of the IEEE, 87:07, July 1999.
16- Marvel, L.M., Boncelet Jr., C.G. &Retter, C., "spread spectrum steganography", IEEE transactions on image processing, 8:08, 1999.
17- P. Mahajan and A. Sachdeva, "A Study of Encryption Algorithms AES, DES and RSA for Security," vol. 13, no. 15, 2013.
18- R. Biswas, S. Bandyopadhyay, and A. Banerjee, "A FAST IMPLEMENTATION OF THE RSA ALGORITHM USING," pp. 1–15, 2014.
19- P. Gupta and S. Kumar, "A Comparative Analysis of SHA and MD5 Algorithm A Comparative Analysis of SHA and MD5 Algorithm," no.
20- Venkatraman, S., Abraham, A. &Paprzycki, M., "Significance of Steganography on data security", Proceedings of the IEEE.
21- Mohammed A. Saleh and Azizah Abdul Manaf. Optimal Specifications for a Protective Framework against HTTP-based DoS and DDoS
22- Attacks. International Symposium on Biometrics and Security Technologies (ISBAST 2014), May 2014 (IEEE).