International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)

ISSN: 2456-236X

Vol. 02 Special Issue 03 | 2017

Image Steganography

SushantGandhi[1] , Mayank Mangal[2]

Department of Information Technology, Alamuri Ratnmala Engineering College, Shahapur

ABSTRACT

Steganography is the art of hiding the fact that communication is taking place, by hiding information in other information. Many different carrier file formats can be used, but digital images are the most popular because of their frequency on the Internet. For hiding secret information in images, there exists a large variety of steganography techniques some are more complex than others and all of them have respective strong and weak points. Different applications have different requirements of the steganography technique used. For example, some applications may require absolute invisibility of the secret information, while others require a larger secret message to be hidden. This paper intends to give an overview of image steganography, its uses and techniques. It also attempts to identify the requirements of a good steganography algorithm and briefly reflects on which steganography techniques are more suitable for which applications.

Keyword:-digital images, Text, Images, Audio/ video, Protocol .

1. INTRODUCTION

Since the rise of the Internet one of the most important factors of information technology and communication has been the security of information. 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 used to implement this, is called steganography.

Steganography is the art and science of invisible communication. This is accomplished through hiding information in other information, thus hiding the existence of the communicated information. The word steganography is derived from the Greek words “stegos” meaning “cover” and “grafia” meaning “writing” [1] defining it as “covered writing”. In image steganography the information is hidden exclusively in images.

The idea and practice of hiding information has a long history. In Histories the Greek historian Herodotus writes of a nobleman, Histaeus, who needed to communicate with his son-in-law in Greece. He shaved the head of one of his most trusted slaves and tattooed the message onto the slave’s scalp. When the slave’s hair grew back the slave was dispatched with the hidden message [2]. In the Second World War the Microdot technique was developed by the Germans. Information, especially photographs, was reduced in size until it was the size of a typed period. Extremely difficult to detect, a normal cover message was sent over an insecure channel with one of the periods on the paper containing hidden information [3]. Today steganography is mostly used on computers with digital data being the carriers and networks being the high speed delivery channels.

Steganography differs from cryptography in the sense that where cryptography focuses on keeping the contents of a message secret, steganography focuses on keeping the existence of a message secret [4]. Steganography and cryptography are both ways to protect information from unwanted parties but neither technology alone is perfect and can be compromised. Once the presence of hidden information is revealed or even suspected, the purpose of

steganography is partly defeated [4]. The strength of steganography can thus be amplified by combining it with cryptography.

Two other technologies that are closely related to steganography are watermarking and fingerprinting [5]. These technologies are mainly concerned with the protection of intellectual property, thus the algorithms have different requirements than steganography. These requirements of a good steganography algorithm will be discussed below. In watermarking all of the instances of an object are "marked" in the same way. The kind of information hidden in objects when using watermarking is usually a signature to signify origin or ownership for the purpose of copyright protection [6]. With fingerprinting on the other hand, different, unique marks are embedded in distinct copies of the carrier object that are supplied to different customers. This enables the intellectual property owner to identify customers who break their licensing agreement by supplying the property to third parties [5]. In watermarking and fingerprinting the fact that information is hidden inside the files may be public knowledge – sometimes it may even be visible – while in steganography the imperceptibility of the information is crucial [4]. A successful attack on a steganography system consists of an adversary observing that there is information hidden inside a file, while a successful attack on a watermarking or fingerprinting system would not be to detect the mark, but to remove it [5].

2. OVERVIEW OF STEGANOGRAPHY

To provide an overview of steganography, terms and concepts should first be explained. An overview of the different kinds of steganography is given at a later stage.

2.1 Steganography concepts

Although steganography is an ancient subject, the modern formulation of it is often given in terms of the prisoner's problem proposed by Simmons [9], where two inmates wish to communicate in secret to hatch an escape plan. All of their communication passes through a warden who will throw them in solitary confinement should she suspect any covert communication [10].

The warden, who is free to examine all communication exchanged between the inmates, can either be passive or active. A passive warden simply examines the communication to try and determine if it potentially contains secret information. If she suspects a communication to contain hidden information, a passive warden takes note of the detected covert communication, reports this to some outside party and lets the message through without blocking it. An active warden, on the other hand, will try to alter the communication with the suspected hidden information deliberately, in order to remove the information [5].

2.2 Different kinds of steganography

Almost all digital file formats can be used for steganography, but the formats that are more suitable are those with a high degree of redundancy. Redundancy can be defined as the bits of an object that provide accuracy far greater than necessary for the object's use and display [11]. The redundant bits of an object are those bits that can be altered without the alteration being detected easily [5]. 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 four main categories of file formats that can be used for steganography.

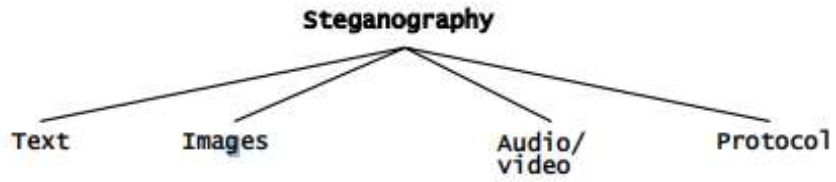

Fig1: Categories of steganography

Hiding information in text is historically the most important method of steganography. An obvious method was to hide a secret message in every n th letter of every word of a text message. It is only since the beginning of the Text Images Audio/ video Protocol Internet and all the different digital file formats that is has decreased in importance [1]. Text steganography using digital files is not used very often since text files have a very small amount of redundant data.

Given the proliferation of digital images, especially on the Internet, and given the large amount of redundant bits present in the digital representation of an image, images are the most popular cover objects for steganography. This paper will focus on hiding information in images in the next sections.

3. Image steganography

As stated earlier, images are the most popular cover objects used for steganography. In the domain of digital images many different image file formats exist, most of them for specific applications. For these different image file formats, different steganographic algorithms exist.

3.1 Image definition

When working with larger images of greater bit depth, the images tend to become too large to transmit over a standard Internet connection. In order to display an image in a reasonable amount of time, techniques must be incorporated to reduce the image's file size. These techniques make use of mathematical formulas to analyse and condense image data, resulting in smaller file sizes. This process is called compression [15].

In images there are two types of compression: lossy and lossless [1]. Both methods save storage space, but the procedures that they implement differ. Lossy compression creates smaller files by discarding excess image data from the original image. It removes details that are too small for the human eye to differentiate [15], resulting in close approximations 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) [14].

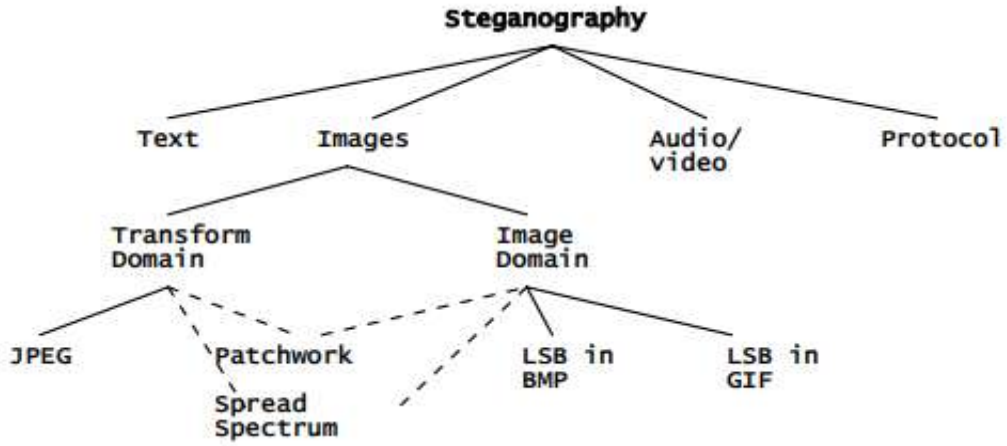

Fig 2: Categories of image steganography

3.1.1 Image Domain

- **Least Significant Bit**

Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover image [14]. The least significant bit (in other words, the 8th bit) of some or all of the bytes inside an image is changed to a bit of the secret message. When using a 24-bit image, a bit of each of the red, green and blue colour Text Images Audio/ video Protocol Transform Domain Image Domain JPEG LSB in BMP LSB in GIF Patchwork Spread Spectrum components can be used, since they are each represented by a byte. 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 [19]. 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)

Although the number was embedded into the first 8 bytes of the grid, only the 3 underlined bits needed to be changed according to the embedded message. 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 [19]. Since there are 256 possible intensities of each primary colour, changing the LSB of a pixel results in small changes in the intensity of the colours. These changes cannot be perceived by the human eye - thus the message is successfully hidden. With a well-chosen image, one can even hide the message in the least as well as second to least significant bit and still not see the difference [14]

In the above example, consecutive bytes of the image data – from the first byte to the end of the message – are used to embed the information. This approach is very easy to detect [4]. A slightly more secure system is for the sender and receiver to share a secret key that specifies only certain pixels to be changed. Should an adversary suspect that LSB steganography has been used, he has no way of knowing which pixels to target without the secret key [5]

3.2 Image or Transform domain

As seen in Figure 2, some steganographic algorithms can either be categorised as being in the image domain or in the transform domain depending on the implementation.

- **Patchwork**

Patchwork is a statistical technique that uses redundant pattern encoding to embed a message in an image [14]. The algorithm adds redundancy to the hidden information and then scatters it throughout the image

[17]. A pseudorandom generator is used to select two areas of the image (or patches), patch A and patch B [22]. All the pixels in patch A is lightened while the pixels in patch B is darkened [22]. In other words the intensities of the pixels in the one patch are increased by a constant value, while the pixels of the other patch are decreased with the same constant value [6]. The contrast changes in this patch subset encodes one bit and the changes are typically small and imperceptible, while not changing the average luminosity [17].

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 [23]. 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 [17]. 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 [14].

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 [23].

4. EVALUATION OF DIFFERENT TECHNIQUES

All the above mentioned algorithms for image steganography have different strong and weak points and it is important to ensure that one uses the most suitable algorithm for an application. All steganography algorithms have to comply with a few basic requirements. The most important requirement is that a steganography algorithm has to be imperceptible. The authors propose a set of criteria to further define the imperceptibility of an algorithm. These requirements are as follows:

- Invisibility – The invisibility of a steganography algorithm is the first and foremost requirement, since the strength of steganography lies in its ability to be unnoticed by the human eye. The moment that one can see that an image has been tampered with, the algorithm is compromised.

- Payload capacity – Unlike watermarking, which needs to embed only a small amount of copyright information, steganography aims at hidden communication and therefore requires sufficient embedding capacity.

The following table compares least significant bit (LSB) insertion in BMP and in GIF files, JPEG compression steganography, the patchwork approach and spread spectrum techniques as discussed in section 3, according to the above requirements:

	LSB in BMP	LSB in GIF	JPEG compression	Patchwork	Spread spectrum
Invisibility	High*	Medium*	High	High	High
Payload capacity	High	Medium	Medium	Low	Medium
Robustness against statistical attacks	Low	Low	Medium	High	High
Robustness against image manipulation	Low	Low	Medium	High	Medium
Independent of file format	Low	Low	Low	High	High
Unsuspicious files	Low	Low	High	High	High

* - Depends on cover image used

Table 1: Comparison of image steganography algorithms

The levels at which the algorithms satisfy the requirements are defined as high, medium and low. A high level means that the algorithm completely satisfies the requirement, while a low level indicates that the algorithm has a weakness in this requirement. A medium level indicates that the requirement depends on outside influences, for example the cover image used. LSB in GIF images has the potential of hiding a large message, but only when the most suitable cover image has been chosen.

The ideal, in other words a perfect,steganography algorithm would have a high level in every requirement. Unfortunately in the 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.

- LSB in BMP – When embedding a message in a "raw" image, that has not been changed with compression, such as a BMP, there exists a trade-off between the invisibility of the message and the amount of information that can be embedded. 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. The main disadvantage regarding LSB in BMP images is surely the suspicion that might arise from a very large BMP image being transmitted between parties, since BMP is not widely used anymore. Suggested applications: LSB in BMP is most suitable for applications where the focus is on the amount of information to be transmitted and not on the secrecy of that information.

- LSB in GIF – The strong and weak points regarding embedding information in GIF images using LSB are more or less the same as those of using LSB with BMP. The main difference is that since GIF images only have a bit depth of 8, the amount of information that can be hidden is less than with BMP. GIF images are especially vulnerable to statistical – or visual attacks – since the palette processing that has to be done leaves a very definite signature on the image. This approach is dependent on the file format as well as the image itself, since a wrong choice of image can result in the message being visible. Suggested applications: LSB in GIF is a very efficient algorithm to use when embedding a reasonable amount of data in a greyscale image.

- PEG compression – The process of embedding information during JPEG compression results in a stego image with a high level of invisibility, since the embedding takes place in the transform domain. JPEG is the most popular image file format on the Internet and the image sizes are small because of the compression, thus making it the least suspicious algorithm to use. However, the process of the compression is a very mathematical process, making it more difficult to implement. Suggested applications: The JPEG file format can be used for most applications of steganography, but is especially suitable for images that have to be communicated over an open systems environment like the Internet.

5. CONCLUSION

Although only some of the main image steganography techniques were discussed in this paper, one can see that there exists a large selection of approaches to hiding information in images. 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 files that increase the probability of detection when in the presence of a warden.

6. REFERENCES

[1] Moerland, T., "Steganography and Steganalysis", Leiden Institute of Advanced Computing Science, www.liacs.nl/home/ tmoerl/privtech.pdf
[2] Silman, J., "Steganography and Steganalysis: An Overview", SANS Institute, 2001
[3] Jamil, T., "Steganography: The art of hiding information is plain sight", IEEE Potentials, 18:01, 1999
[4] Wang, H & Wang, S, "Cyber warfare: Steganography vs. Steganalysis", Communications of the ACM, 47:10, October 2004
[5] Anderson, R.J. &Petitcolas, F.A.P., "On the limits of steganography", IEEE Journal of selected Areas in Communications, May 1998
[6] Marvel, L.M., Boncelet Jr., C.G. &Retter, C., "Spread Spectrum Steganography", IEEE Transactions on image processing, 8:08, 1999
[7] Dunbar, B., "Steganographic techniques and their use in an Open-Systems environment", SANS Institute, January 2002
[8] Artz, D., "Digital Steganography: Hiding Data within Data", IEEE Internet Computing Journal, June 2001
[9] Simmons, G., "The prisoners problem and the subliminal channel", CRYPTO, 1983
[10] Chandramouli, R., Kharrazi, M. &Memon, N., "Image steganography and steganalysis: Concepts and Practice", Proceedings of the 2nd International Workshop on Digital Watermarking, October 2003
[11] Currie, D.L. & Irvine, C.E., "Surmounting the effects of lossy compression on Steganography", 19th National Information Systems Security Conference, 1996
[12] Handel, T. &Sandford, M., "Hiding data in the OSI network model", Proceedings of the 1st International Workshop on Information Hiding, June 1996
[13] Ahsan, K. &Kundur, D., "Practical Data hiding in TCP/IP", Proceedings of the Workshop on Multimedia Security at ACM Multimedia, 2002