IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735. Volume 8, Issue 1 (Sep. - Oct. 2013), PP 56-60
www.iosrjournals.org

A Survey on Digital Image Steganography and Steganalysis

Gangadhar Tiwari [1], Arun Kumar Yadav [2], Madhusudhan Mishra[3]
[1] *(IT Department, NIT Durgapur, India)* [2]*(CSE Department, M.G. Institute of Management and Technology, Lucknow, India)* [3]*(ECE Department, NERIST, India)*

Abstract: *Steganography is an information security approach used to hide messages inside suitable covers in such a way that it is not known to attackers. The cover files may be any digital data including Image or Audio files. For steganography several methods exists where each of them has some advantages and disadvantages. Steganographic applications have varying requirements depending upon the steganography technique used. In this paper we present an overview of image steganography and steganalsysis, its uses and techniques. It also attempts to identify the requirements of a good steganographic algorithm and compares their performance with respect to requirements.*

Keywords: *Image Steganography, Data Hiding, Steganalysis, Spread Spectrum, Patchwork*

I. Introduction

Steganography means covered writing. It is used to conceal information inside cover files making it impossible to detect them. Various steganography terminologies are discussed below.

1. Embedded data Type- It is the data to be hidden.
2. Stego-data Type- Output of the hiding process.
3. Cover-data Type- An input which is original form of stego data type.
4. Stego key- Additional secret data needed in hiding process.
5. Embedding- It is process of hiding embedded message.
6. Extracting- Getting embedded message out of Stego-message again.
7. Stego-analyst- From whom message is hidden

In this paper, Section-2 presents a comparison between Steganography and Cryptography. Section-3 describes various Steganographic Techniques. Section-4 is devoted on Image Steganalysis. Section-5 presents the conclusion and future scope.

II. Steganography vs. Cryptography

The term steganography means covered writing whereas cryptography means secret writing. In this section we compare steganography and cryptography based on key criterion [1].

1. Carrier for Steganography is any digital media while for cryptography it is text based image files.
2. Secret data for steganography is payload while for cryptography it is plain text.
3. In steganography Key is optional while for cryptography it is necessary.
4. At least two input files are required for steganography but one is sufficient for cryptography.
5. In steganography as well as cryptography, detection scheme is blind.
6. For authentication both require full data retrieval.
7. The objective of steganography is covert communication while cryptography is deployed for data protection.
8. Output of steganography is a stego file while that of cryptography it is cipher text.
9. Steganography is concerned with detecting capacity while cryptography stresses on robustness.
10. Attacks on steganography is known as steganalsysis while for cryptography it is called cryptanalysis.
11. Steganography fails when it is detected while cryptography fails when deciphered.
12. Steganography is a very old information security approach except its digital version while cryptography is a present era technique.

III. Different kinds of Steganography

3.1. Based on type of Original signal

There are four different types of steganography on the basis of original signal viz. Text, Audio, Image and Protocol. Given the proliferation of digital images on Internet and large amount of redundant bits present in it, images are the most popular cover objects for Steganography. This review will focus on hiding information in images in the next sections. The term Protocol Steganography refers to the technique of embedding information within messages and network control protocols used in network transmission.

3.2 Based on Key Usage

The various types of Steganographic techniques based on usage of key are: Pure Steganography, Secret key Steganography and Public key Steganography.

3.2.1 Pure Steganography

It is the process of embedding the data into the object without using any private keys. It fully relies on the secrecy and a cover image is used for data embedding, personal information to be transmitted, and encryption decryption algorithms to embed the message into image. It cannot provide the better security**.**

3.2.2 Secret key Steganography

It uses the same method as discussed above while employing secure keys. Individual keys are deployed for embedding the data into the cover object**.**

3.2.3 Public key Steganography

Public key Steganography uses two types of keys- one for encryption and another for decryption. The encryption key is kept private while decryption key is made public and available in a public database**.**

3.3 Based on Embedding Domain

There are number of image steganography algorithm proposed on the basis of embedding domain that includes **Spatial or Transform**, depending on redundancies used from domain for the embedding process [2].

3.3.1 Spatial Domain Embedding- LSB Modification

In spatial domain techniques messages are embedded in intensity of pixels directly. It encompasses bit-wise methods that apply noise manipulation and bit insertion. **LSB technique** is the most commonly used steganographic algorithm. It makes use of fact that least significant bits in an image could be thought of random noise and changes to them would not affect the image. LSB based methods can be divided into two groups viz.- **LSB Replacement** and **LSB Matching**. In the LSB replacement technique, the LSB of the pixels is replaced by the message to be sent. While in LSB Matching Technique the pixels are randomly incremented or decremented by secret bits. Popular Steganographic algorithms change LSB of pixels randomly while others amend pixels in particular areas of images. It is most suitable for applications where the focus is on the amount of information to be transmitted and not on secrecy of that information. LSB technique performs better in terms of speed and quality of data hiding however security is a major concern.

Palette based images are popular image file format used on the Internet. GIF images can also be used for LSB Steganography. An obvious demerit of the palette approach when used with GIF images is that any single change in the least significant bit of a pixel, leads to a completely different color. If adjacent palette entries are dissimilar the change would be evident. One solution is to add new colors which are visually similar to the existing colors in the palette that results in original image to have less unique colors than the maximum number of colors. In this method, care should be taken to select the right cover image. But any tampering with the palette of an indexed image leaves a trail making detection easier.

Another solution to the problem is to use grayscale images. In an 8-bit grayscale GIF image, there are 256 different shades of grey. The changes between the colors are very gradual, making it harder to detect. Moreover, the amount of information that can be hidden is less than BMP. GIF images are especially vulnerable to statistical/ visual attacks – since the palette processing that has to be done leaves a very definite signature on image. This approach is relies on image and its file format, as because a wrong choice of image would lead to message visibility.

Another approach for embedding messages in spatial domain is introducing statistically similar noise to pixel randomly e.g., scanner noise. A pseudo random noise generator is employed to generate the noise using a shared key**.** Embedding and detection of the message modulated by the generated noise is achieved by a parity function.

Naseem et al proposed a technique based on using Distributed LSB for Data Hiding in Images which has better hiding capacity and causes less degradation in the resulted stego image by using the lower bits to hide secret bits as per the new rule which takes into account the intensity level of each pixel [3, 6]. The proposed algorithm is reversible as the secret data is recovered properly. In order to protect the stego image from channel induced interference, channel encoding and modulation techniques can be used. In the recovery process each pixel is inspected for its intensity value. The last three bits are extracted if a pixel belongs to the first range. This process continues till exhausting first range pixels after which the second range of pixels are found and the lower four bits are extracted. However, security is a major concern as it does not uses encryption key. Moreover, it assumes a communication channel to be perfect which is not possible due to noise.

3.3.2 Transform Domain Techniques- JSteg and F5 Alogorithm

Here images are first transformed and then the message is embedded in it. These methods hide messages in more significant parts of the cover image, making it secure against common signal processing distortions. Transform domain techniques are does not depend on the image format and thus an embedded message survives compression attacks. A majority of transform domain techniques focus on using redundancies in discrete cosine transform (DCT) domain.

JSteg algorithm is a steganographic technique for embedding data into JPEG images. The image content is transformed into frequency coefficients so as to achieve compressed storage. It replaces LSBs of quantized Discrete Cosine Transform [4]. In this process the hiding skips all coefficients with the values of 0 or 1. It has high capacity and had a compression ratio of 12%. This algorithm is secure against visual attacks and offers higher embedding capacity for Steganographic messages but it is vulnerable to statistical attacks. The **F5 algorithm** embeds the message into randomly chosen Discrete Courier Transform (DCT) coefficients. By employing matrix embedding which minimizes the changes to be made to the length of certain message. It provides high Steganographic capacity, faster speed and can prevent visual and statistical attacks. This algorithm supports almost all image file formats. The performance of the algorithms differs with the type of cover image or source on which the data is embedded. The comparison of algorithms is given in Table-III below From the above review it is clear that both the spatial domain and transform domain steganographic techniques had their own weakness and hence some hybrid methods were required. We describe two hybrid techniques viz. Patchwork and Spread Spectrum Techniques in our next section.

3.3.3 Hybrid Techniques-Patchwork and Spread Spectrum

Patchwork is a statistical technique that uses redundant pattern encoding to embed a message in an image. It adds redundancy to the secret information and scatters it throughout the image. Two areas of image, patch ***A*** and patch ***B***, are selected using a pseudorandom generator All patch ***A*** pixels are lightened simultaneously darkening patch ***B*** pixels. The contrast changes in this patch subset encodes one bit and the changes are usually small and invisible, while keeping average luminosity intact. The patchwork approach is used independent of the host image and proves to be quite robust and message can survive conversion between lossy and lossless compression. Also, since the secret message is distributed over the entire image, if one patch is destroyed, the others exist which makes it suitable for transmitting sensitive information. However, it works well with small size messages. If the message is too big, it can only be embedded once. Moreover, using patchwork approach only one bit is embedded. The biggest disadvantage of the patchwork approach is the small amount of information that can be hidden in one image and overcoming the menace hampers the secrecy.

In **Spread Spectrum** technique hidden data is spread throughout the cover-image making it harder to detect. A system proposed by Marvel et al combines error control coding and image processing with spread spectrum technique to hide information in images. It can be defined as the process of spreading the bandwidth of a narrowband signal across a wide band of frequencies. It is achieved by adjusting the narrowband signal with a wideband signal, such as white noise. After spreading, the energy of the narrowband signal in any one frequency band is low and therefore difficult to detect. Here the message is embedded in noise and then combined with the cover image to produce the stego image. Thus the embedded image is imperceptible without access to the original image. It satisfies most requirements and is secure against statistical attacks, since the hidden information is uniformly scattered keeping intact the statistical properties of image. Spread spectrum techniques can be used for most Steganography applications, although it is highly mathematical and intricate approach.

3.4 Evaluation of various techniques

TABLE-1 below compares the effectiveness of common Image Steganographic Algorithms in terms of six parameters namely Invisibility, Bit rate, Robustness against Statistical Attack, Robustness Against Image Processing(IP) Attacks, Independence of file format and Use of Unsuspicious files[5, 7]. The levels at which the algorithms satisfy the requirements are defined as high, medium and low. If the algorithm completely satisfies the requirement it is rated high, whereas an algorithm with a weakness in requirements is rated low. A medium level indicates that the requirement depends on other factors like the cover image used. For e.g. LSB in GIF images can hide large messages provided suitable cover image are selected. It is to be noted that an ideal Steganographic algorithm would have a high level in every requirement which is non-existent yet creating a trade-off, depending on which requirements are more important for the specific application.

Table-1

	LSB in GIF	JPEG	Patchwork	Spread Spectrum
Invisibility	Medium	High	High	High
Bit rate	Medium	Medium	Low	Medium
Robustness against Statistical Attack	Low	Medium	High	High
Robustness against IP Attack	Low	Medium	High	Medium

Independent of file extension	Low	Low	High	High
Unsuspicious file	Low	High	High	High

IV. Image Steganalysis

Aiming at detecting secret information hidden in a given image using steganographic tool, steganalsysis has been of interest since the end of 1990's. Steganographic attacks consist of extraction, detection and destruction of hidden messages of the stego media [8].

4.1. Types of Attacks

There exist various attacks based on information available for analysis. It is divided into six main categories.
Stego-only attack- Only the stego-image is available for analysis.
Known cover attack- The original cover-image and stego- image are both available.
Known message attack- At some point, the hidden message becomes known to the attacker. Even with the message, this may be very difficult and may even be considered equivalent to the stego-only attack.
Chosen stego attack- The steganography tool and stego-image are known.
Chosen message attack- Steganalyst generates a stego-image from steganography tool from a chosen message.
Known stego attack- algorithm is known and both the original and stego-image are available.
According to the targeted steganographic tools, steganalsysis can be also broadly classified into **specific** and **universal steganalsysis**. Specific steganalsysis is designed to detect some particular steganalsysis while Universal steganalsysis is also called blind steganalsysis or universal blind steganalsysis. It can detect the existence of secret message without steganography algorithms and practicable than the specific steganalsysis.

4.2 Steganalysis Technique

There have been a number of steganalsysis techniques proposed each operating with its own unique approach. In essence steganalsysis technique operates by obtaining a set of statistical feature from the input image. Avcibas I first proposed a universal steganalsysis algorithm to detect embedded message in images through a proper selection of image quality metrics (IQM) and a multi-variant regression analysis [9]. Then he improved the algorithm and increased the detection precision. Jiang N improved some IQM standards and combined them into a new feature vector. It adopted the supported vector machine (SVM) classifier to distinguish between the original and stego images. There are many works reporting that high-order statistics are very effective in differentiating stego-images from cover images. Farid proposed a general steganalsysis algorithm based on image higher-order statistics. He modeled the universal steganalsysis by supervised learning for the first time and indicated that the supervised learning was effective for detecting stego images without knowing the statistics property of images and steganography methods. Later Farid used a wavelet-like decomposition to build higher-order statistical models of natural images [10]. For each image, firstly three-scale quadrature mirror filters (QMF) decomposition was made, then higher-order probability dimensional function (PDF) moments were extracted, finally the classifier was used to detect the test images. Holotyak et al used higher-order moments of the PDF of the estimated stego-image in the finest wavelet level to construct the feature vectors [11]. Shi et al proposed the use of statistical moments of the characteristics functions of the wavelet sub-bands. Since the n-th statistical moment of wavelet characteristics function is related to the n-th derivative of the corresponding wavelet histogram, the constructed 3-Dimensional feature vector has proved to be sensitive to embedded data. Usually, the steganalsysis algorithm based on the higher-order statistics achieved satisfactory performance on image files, regardless of the underlying embedding algorithm. However, since the image-embedding method is typically unknown to steganalyst, many researches focused on the design of a blind steganalsysis algorithm to detect the presence of steganography independent of the steganography algorithm used. In Geetha's work they employed the higher order statistical features that were collected from a new transform domain, i.e., curvelet transform domain [12]. This work is typically a novel and first attempt of employing these features for Steganalysis. Experimental results also proved that their claim was justified.

V. Conclusion

Hybrid Steganography techniques like Patchwork and Spread Spectrum Techniques are efficient enough for enabling covert communication. Combining them with Cryptographic techniques provides for high security. In the present era of internet, steganography has undergone a lot of research and so the steganalsysis. In this paper we presented a summary on Image Steganography and Steganalysis. However, steganography and steganalsysis have different requirement for different applications and not one algorithm is sufficient to provide a solution for all of them. There are many challenges which need to be investigated in Steganography and corresponding steganalsysis. For e.g. How to choose good cover image for hiding information? Which image formats are most suitable? Which steganographic/cryptographic algorithm should be employed to provide better speed and security etc? Therefore steganography and steganalsysis requires further research.

References

[1]. Abbas Cheddad A, Joan Condell, Kevin Curran, Paul Mc Kevit, Digital image steganography: Survey and analysis of current methods, *Signal Processing, Elsevier, 2009,* 72-75

[2]. Mehdi Kharrazi, Husrev T. Sencar, and Nasir Memon, Image Steganography: Concepts and Practice. Lecture Notes, Polytechnic University, Brooklyn, NY 11201, USA

[3]. Mohammad Kamran Khan, Mohammad Naseem, Ibrahim Mohammad Hussain and Aisha Ajmal, "Distributed Least Significant Bit Technique for Data Hiding in Images", *Multitopic Conference (INMIC)", IEEE*, 2011, 149 – 154

[4]. Wai Wai Zin and Than Naing Soe, "Implementation and Analysis of Three Steganographic Approaches", *Computer Research and Development (ICCRD), 3rd International Conference IEEE, 2*, 2011, 60-64

[5]. A.W. Naji , Teddy S. Gunawan, Shihab A. Hameed, B.B Zaidan, A.A Zaidan, Stego-Analysis Chain, Session One-Investigations on Steganography Weakness Vs Stego-Analysis System for Multimedia File, *International Association of Computer Science and Information Technology - Spring Conference, IEEE,*2009, 405 – 409

[6]. B.B.Zaidan, A.A.zaidan, *Enhancement of the Size of Hidden Data and the Quality of Image Using LSB Algorithm,* Master research, Kuala Lumpur University, Malaysia, 2008.

[7]. Ge Huayong, Huang Mingshenga, Wang Qian, Steganography and Steganalysis Based on Digital Image, *4th International Congress on Image and Signal Processing, IEEE, 1*, 2011, 252 – 255

[8]. S.Z. Wang, X.P. Zhang, Recent Advances in Image Based Steganalysis Research, *Chinese Journal of Computers*, *32*, 2009, 1247-1263

[9]. I. Avcibas, N. D. Memon, Steganalysis using Image Quality Metrics, *IEEE Trans Image Process*ing, *12*, 2003, 221-229

[10]. H. Farid, Detecting hidden messages using higher order statistical models, *IEEE transactions on signal processing*, *2,* 2002, 905-908

[11]. T. Holotyak, J. Fridrich, Blind Statistical steganalysis of additive Steganography using wavelet higher order statistics, Lecture notes in Computer Science, 2005, 273-274

[12]. S. Geetha, K.Kamaraj, Passive steganalysis based on higher order image statistics of curvelet transform, *International Journal for Automation and Computing*, *7*, 2010, 531-542

.