International Journal of scientific research and management
(IJSRM) ||Volume||2||Issue||12||Pages||1753-1757||2014||
Website: www.ijsrm.in ISSN (e): 2321-3418

Image Steganography and Steganalysis : A Survey

Asha Pathak[1], Vrushali Bhuyar[2]

[1] MCA student at G. S. Mandal's MIT college Aurangabad, India
[1]ashu.pathak1012@gmail.com
[2] Assistant professor at G. S. Mandal's MIT college Aurangabad, India

Abstract: Steganography is communication of secret data in an appropriate carrier, e.g. image, audio, video or TCP/IP header file. For hiding secret data in digital images, large varieties of steganographic techniques are available; some are more complex than others. And all of them have their respective pros and cons. This paper intends to give understanding and evolution of different existing digital image steganography techniques of data hiding in spatial, transform and compression domains. It integrates research work on steganography and steganalysis. Detection of steganography, estimation of message length, and its extraction are part of the field of steganalysis. The battle between steganography and steganalysis is unending. In this paper, a review report is presented for steganography and steganalysis

Keywords: *Steganalysis, Steganography, Data hiding*

1. INTRODUCTION

In internet communication one of the most important factors of information technology and communication has been the security of information. Everyday tons of data are transferred through the internet through e-mail, file sharing sites, social networking sites. As the number of Internet users are increasing, the concept of Internet security has also gain very much importance.

Steganography purpose is to hide the very presence of communication by embedding messages into innocuous looking cover objects. In today's digital world, invisible ink and paper have been replaced by much more versatile and practical covers for hiding messages such as digital documents, images, video, and audio files. As long as an electronic document contains perceptually irrelevant or redundant information, it can be used as a cover to hide secret messages. Each steganographic communication system consists of an embedding algorithm and an extraction algorithm. To accommodate a secret message, the original image, also called the cover-image, is modified by the embedding algorithm. As a result of embedding, the stego-image is obtained.

2. Steganography [4]

Steganography word is of Greek origin and essentially its meaning is concealed writing. Protection of the transmitted data from being intercepted or tampered has led to the development of various steganographic techniques.

The data hiding process in steganography starts by identifying a cover image's redundant bits. Redundant bits are the bits that can be modified without destroying its integrity. The embedding process then creates stego-image by replacing these redundant bits with the bits of the message to be hidden. In digital image steganography, the secret message is embedded within a digital image called cover-image. Cover-image carrying embedded secret data is referred as stego-image.

Applications of Steganography

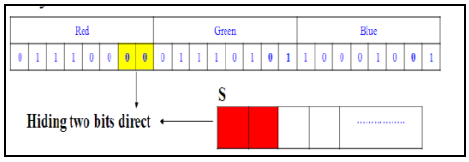
Steganography can be used for wide range of applications such as, in defence organisations for safe circulation of secret data, in military and intelligence agencies, in smart identity cards where personal details are embedded in the photograph itself for copyright control of materials. In medical imaging, patient's details are embedded within image providing protection of information and reducing transmission time and cost [1], in online voting system so as to make the online election secure and robust against a variety of fraudulent behaviours [2], for data hiding in countries where cryptography is prohibited, in improving mobile banking security [3], in tamper proofing so as to prevent or detect unauthorised modifications and other numerous applications.

Classification of steganographic techniques

1. SPATIAL DOMAIN-BASED STEGANOGRAPHIC TECHNIQUES

1.1 Least Significant Bit Embedding

Least significant bit embedding technique is the most popular steganography technique. Depending upon the binary coding of the secret message it hides the message in the binary coded image. The below figure depicts the knowledge of pixel values and secret messages. LSB makes changes in the image which are very easy to recognize and the stego images are easy to attack.

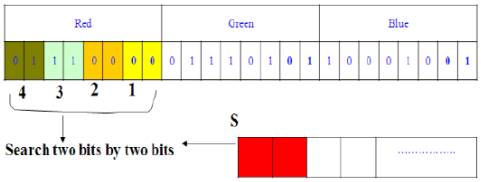


Algorithm for LSB Method

Steps to Hide the Message Using the LSB Method.

1.Choose the proper image for the cover medium.
2. Scan the image row by row and encode it in to binary form.
3. Encode the secret message into binary notation.
4. Calculate the sizes of the image and secret message. 5. Consider one pixel of the image.
6. Segment the image into three parts (Red, Green and Blue parts)
7. Hide two by two bits of the secret message in each position of the pixel at the last two significant positions.
8. Set the image with the newly considered values.
9. Set the image and save it.

1.2 Advanced LSB Embedding



The LSB method hides the secret text at the least two significant bits of the image pixels. Hence a change in the value of image pixels affects the image resolution, which ultimately leads to the reduction of image quality and make the image easy to attack. This LSB method is already attacked and broken. So in advanced LSB we are hiding the secret message based on searching about the identical values between the image pixels and secret messages. The image shown below gives a clear idea of arrangements of bits in the image pixel.
Algorithm for Advanced LSB Method. Steps to Hide the Message Using the LSB Method.

- Choose the proper image for the cover medium.
- Scan the image row by row and encode it in to binary form.
- Encode the secret message into binary notation.
- Calculate the sizes of the image and secret message.
- Consider one pixel of the image.
- Segment the image into three parts (Red, Green and Blue parts)
- Hide two by two bits of the secret message in each position of the pixel by searching the identical.
- If the search is unsuccessful then hide two by two bits of the secret message in the least significant position of pixel image.
- Store the positions of the hiding bits in a binary table.
- Set the image with the newly considered values.
- Set the image and save it.

1.3 Bit Plane Complexity Segmentation Steganography (BPCS)[5][14]

Bit plane complexity segmentation steganography was introduced by Kawaguchi, et al. The basic idea is that the higher bit planes can also be used for embedding information. Here, each block is decomposed into bit-plane. The LSB plane of the image would be a binary image consisting of the LSB of each pixel in the image and so on. In each segmented bit-plane its complexity is analysed. And based on a threshold value, block is divided into informative region and noise-like region and the secret data is hidden in noise regions without degrading image quality. This method provides high embedding capacity and least degradation of the cover image as compared to traditional LSB manipulation techniques.

1.4 A New Steganography Scheme based on an Index-color Image[6]

A new more efficient steganography Technique is proposed by Se-Min Kim et al. It can be applied in Internet environment. This technique can hide from 1 to 8 bits secret data per pixel, and has no distortion if the number of colors does not exceed 128. This scheme first divides the secret data into several parts based on the number of colors in the cover image, and then embeds secret data into the cover image, part by part by expanding palettes and modifying indices. Here, the cover image can be recovered easily without loss. Numerical experiments indicate that this technique introduces no distortion to the cover image, in contrast with other schemes that based on index-color, such as EZ Stego. This scheme can be applied to BMP, GIF and PNG image formats which use the index-color technique. Thus, this scheme suits the current Internet environment very well.

2. DATA HIDING TECHNIQUES IN FREQUENCY DOMAIN

Frequency domain methods hide messages in significant areas of the cover-image which makes them more robust to attacks such as compression, cropping or image processing methods than LSB approach and moreover they remain imperceptible to the human sensory system as well. Some of the frequency domain-based steganographic data hiding methods are:

2.1 JSteg and OutGuess[7]

JSteg developed by Derek Upham sequentially replaces the LSB of the DCT coefficients with the message's data. This technique does not require a shared secret; as a result, anyone who knows the steganographic system can retrieve the message easily, thus not so secure.
Outguess was proposed by Provos as a response to the statistical tests given by Andreas Westfeld. It improves embedding by selecting DCT coefficients randomly. Two versions are available: Outguess 0.13b which is vulnerable to extended version of $\chi 2$ -test and Outguess 0.2 which has the ability to preserve frequency counts statistics and hence remain undetected. Provos observed that while embedding not all the redundant bits were used and thus it is possible to use the remaining bits to correct statistical deviations that embedding created. Outguess 0.2 uses this phenomenon to avoid class of $\chi 2$ -tests.

2.2 Data Hiding Techniques: F3, F4 and F5[8]

F3 algorithm decrements the absolute value of non-zero coefficients only if the LSB does not match with the secret bit. And zero coefficients are skipped completely. Advantage of F3 is its resistance to statistical attack ($\chi 2$ -test). Drawbacks are its less capacity, surplus of even coefficients caused by shrinking and repetitive embedding required since receiver cannot differentiate between skipped 0 and the 0 generated due to shrinkage.

The F5 algorithm was introduced by German researchers Pfitzmann and Westfeld. F5 algorithm embeds

message bits into randomly-chosen DCT coefficients and does matrix embedding that minimizes the necessary number of changes to embed a message of certain length. The algorithm F5 comes after a series of F3 and F4. F5 is similar to F4 except that F4 does not use matrix encoding in embedding process. The major strengths and advantages of F5 are its high embedding capacity without sacrificing security and its resistance to statistical and visual attacks.

3. DATA HIDING TECHNIQUES IN COMPRESSION DOMAIN

In recent years, researchers have concentrated on embedding secret data into the compression domain. Various methods have been proposed for hiding data directly into the compressed codes of the image. Furthermore, the compressed codes transmitted attract less attention of the intruder.

3.1 Quantization based image steganography without data hiding position memorization [9]

This is a quantization-based steganography method of extracting hidden data without any memorization of data embedding positions. This method offers the user the flexibility of choosing data hiding positions. Hence it enables the user to select positions for embedding data on an individual image basis and/or the basis of the coding scheme being applied to the images.

This method requires no knowledge on data hiding positions when the data are extracted. This method offers the user the flexibility of choosing positions to hide data. It, thus, modifies the coefficients to embed data on the basis of the individual image and/or image compression technology.

Conventional steganography methods require memorizing positions for data hiding to extract the hidden data. Since they fix the positions within the whole image, it is difficult to choose data hiding positions on an individual image basis. On the other hand, methods that embed only one bit per one position are not able to embed multiple bits in any one position. Methods that are able to be extended to embed multiple bits in one position may degrade the image-quality considerably, because they have to embed L-level data using |_log2 L_| bit data, e.g., four bits are hidden for data that has nine levels.

It hides integer data in transformed coefficients after the quantization process of image compression. Hence, the hidden data are no longer distorted by the rest of the compression process.

Measurement of Steganography

There are several parameters to measure the performance of the steganographic system which are imperceptibility, robustness, and embedding capacity.

1. Imperceptibility is the primary parameter that is required for steganographic system to fulfil. It shows how difficult for third party to determine whether there is a hidden data in the stego-media or not. In other word it represents the ability to avoid attention of third party from detecting the stego-media. For example, people do not know there is a hidden message in image and when comparing the image with hidden message with original image, there is no difference between the two. Even though, no human eyes can detect the hidden message, there is possibility the stego-media being attacked by statistical attack. Thus, truly secure steganographic should not be undetectable either by human eyes or statistical attack.

1. Robustness here means how well the steganography systems resist the attempt of third party to extract hidden data. Some examples of distortion to hidden data that interceptors are attempting to do are image manipulation such as cropping and rotating the image, data compression and image filtering.

3. Embedding capacity is an important feature for steganography. Seganographic capacity shows the maximum information that can safely embedded in a stego-media without being statistically detectable. More the hidden message that are intended to embed in carrier, the more alteration should be made to the carrier which consequently turn out the stego-media to be detected by third party easily.

3. STEGANALYSIS

Steganalysis [4] is the process of identifying steganography by inspecting various parameter of a stego media. The primary step of this process is to identify a suspected stego media. After that steganalysis process determines whether that media contains hidden message or not and then try to recover the message from it. Steganalysis is the art of detecting steganography, which is identifying hidden message in suspected stego-media. Detection of steganography, estimation of message length, and its extraction is the part of the field of steganalysis. Steganalysis has recently received a great deal of attention both from law enforcement and the media.

Researches on this field tend to find statistical properties of images that the stego-system doesn't conserve or find methods that one can find out if the image was altered at all or not. Thus, steganalysis is considered successful if it can guess whether an image contains a hidden message or not with a probability higher than random guessing.

In practice, a steganalyst is frequently interested in more than whether or not a secret message is present. The ultimate goal is to extract and decipher the secret message. However, in the absence of the knowledge of the stego technique and the stego and cipher keys altogether, this task may be extremely time consuming or completely infeasible. Therefore, any additional information, such as the message length or its approximate placement in image features, could prove very valuable to the analyst. From the research done on steganalysis, the basis of the attacks is to just be able to distinguish between a cover and a stego image. As, it was mentioned earlier, finding a stego image may lead to even extracting the message

A. Categories of steganalysis attacks

These attacks can be categorized into 5 forms and it is assumed that the steganalyst always has the stego medium at least:

a. Chosen Stego Attack: In this scenario, the steganalyst is aware of the steganographic algorithm used to make the stego medium in order to match the intercepted stego medium. This description is based on the chosen cipher text attack but in the case of steganography it is more complicated to do. In theory, attempting to make new stego mediums to match the intercepted one sounds right, but in practice it is very hard to accomplish that since not only the cover medium is unknown but also the message embedded as well. So, accomplishing a match of the intercepted with the chosen stego is very hard to do in practice. A more realistic description of this attack could be that the steganalyst, since the steganographic algorithm and the stego medium are known, can make new stego images in order to derive a methodology to use for the specific algorithm and thus attack the intercepted medium. The above description is realistic since it is what some steganalyst do in the real world and not just in theory.

b. Stego Only Attack: In a stego-only attack the steganalyst does not have any other information available apart from the

stego medium. It is similar to the cipher text only attack and it is the hardest scenario for the cryptanalyst. Realistically, the only way a steganalyst would be able to attack it is by trying every possible known attacks on current steganographic algorithms.

c. Known Cover Attack: In a known cover attack apart from the stego medium, the original cover medium is also available. In this scenario, the steganalyst can find differences in the two mediums and hence attempt to find what kind of steganographic algorithm was used. This attack is similar to known plaintext attack.

d. Known Message Attack: A known message attack can be used when the hidden message is revealed. The steganalyst by knowing the hidden message can attempt to analyse the stego image for future attacks. Even by knowing the message, this may be very difficult and may even be considered equivalent to the stego-only attack. A targeted steganalysis technique works on a specific type of stego-system and sometimes limited on image format. By studying and analysing the embedding algorithm, one can find image statistics that change after embedding. The results from most targeted steganalysis techniques are very accurate but on the other hand, the techniques are inflexible since most of the time there is no path to extend them to other embedding algorithms. When a targeted steganalysis is successful, then it has a higher probability than random guessing, it helps the steganographic techniques to expand and become more secure.

e. A blind steganalysis technique is designed to work on all types of embedding techniques and image formats. In a few words, a blind steganalysis algorithm studies the difference in the statistical properties of pure and stego images and distinguishes between them. The studying process is done by training the machine on a large image database. Blind techniques are usually less accurate, but a lot more expandable.

f. Semi-blind steganalysis works on a specific range of different stego-systems. The range of the stego-systems can depend on the domain they embed on, i.e. spatial or transform.

B. Steganalytic Methods

Most steganographic programs embed message bits either sequentially or in some pseudo-random manner. In most programs, the message bits are chosen non-adaptively independently of the image content. If the image contains connected areas of uniform color or areas with the color saturated at either 0 or 255, we can look for suspicious patterns using simple visual inspection after pre-processing the stego-image. This attack is applicable to palette images for LSB embedding in indices to the palette. If, at the same time, the message is embedded sequentially, one can have a convincing argument for the presence of steganographic messages in an image. However, it may be impossible to distinguish noisy images or highly textured images from stego images using this technique. Although visual attacks are simple, they are hard to implement and they are not reliable.

a. Signature Steganalysis

Steganography alters the properties of image due to the insertion of message bits in the form of degradation or repeated patterns, which act as signatures that convey the existence of embedded message. Steganographic algorithm such as Hide & Seek produces stego-image that contain pixel values that are divisible by 4, which acts as a specific signature taking the insecure aspect for detection by steganalytic tools [60]. Similarly, steganographic tool Jpegx inserts secret message at the end of JPEG file marker, preceeding with hex code 5B 3B 31 53 00, which acts as a specific signature for detection of secret message in the stego-image

b. Blind Image Steganalysis Based on Statistical Analysis of Empirical Matrix [10]

Here the steganalytic technique for carrying out anti steganalysis test used is Blind Image Steganalysis Based on Statistical Analysis of Empirical Matrix plotted for that image.

A steganalysis method based on statistical analysis of empirical matrix (EM) is proposed to detect the presence of hidden message in an image. The projection histogram of EM is used to extract features composed of two parts: the moments of PH and the moments of the characteristic function of PH. A test database is constructed, based on which a detailed test for different categories of features and a comparison with previous methods are conducted.

c. Steganalysis using color model conversion [11]

The main threat in cyber crime for digital forensic examiner is to identify and interpret the concealed information inside digital medium such as image, audio and video. There are strong cases that hiding information inside digital medium has been used for planning criminal activities. It is very important to develop a steganalysis technique which detects the existence of hidden messages inside digital medium. Here focus is on universal image steganalysis method which uses RGB to HSI colour model conversion. Any Universal Steganalysis algorithm developed should be tested for various stego images to prove its efficiency. The developed Steganalysis algorithm is tested in stego-image database which is obtained by implementing various RGB LSB Steganographic algorithms. Though there are many stego-image sources available on the internet it lacks the information such as how many rows has been infected by the steganography algorithms, how many bits have been modified and which channel has been affected. These parameters are important for Steganalysis algorithms and it helps to rate its efficiency.

5. CONCLUSION

This paper presented the research work in the field of steganography deployed in spatial, transform, and compression domains of digital images. Transform domain techniques make changes in the frequency coefficients instead of manipulating the image pixels directly. Hence distortion is minimum and hence they are preferred over spatial domain techniques. But for good embedding capacity, spatial domain techniques give better results. There exists a trade-off between the image quality and the embedding capacity.

6. REFERENCES

[1] Nirinjan, U.C. & Anand, D. Watermarking medical images with patient information. In the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong, China, 1998, pp. 703-06.

[2] Katiyar, S.; Meka, K.R.; Barbhuiya, F.A. & Nandi, S. Online voting system powered by biometric security using steganography. In the 2nd International Conference on Emerging Applications of Information Technology (EAIT), Kolkata, India, 2011, pp. 288-291.

[3]Shirali-Shahreza, M. Improving mobile banking security using steganography. In the 4th International Conference on Information Technology, ITNG, Las Vegas, 2007, pp. 885-887.

Machado, R. EzStego, Stego Online. http://www.stego. com (Accessed on 15 April 2011).

[4]Johnson, N.F. & Jajodia, S. Exploring steganography: Seeing the unseen. IEEE Computer, 1998, 31(2), 26-34.

[5]Kawaguchi, E. & Eason, R.O. Principle and applications of BPCS-Steganography. In the SPIE Conference on Multimedia Systems and Applications, Boston, 1998, 3524, pp. 464-73.

[6]Cheng, Z.; Kim, Se-Min &Yoo, Kee-Young. A new steganography scheme based on an index-colour image. In the 6th International Conference on

Information Technology: New Generations, Las Vegas, Nevada, 2009, pp. 376-81.
[7] Quantitative Structural Steganalysis of Jsteg Jan Kodovský and Jessica Fridrich, Member, IEEE
[8]Westfeld, A. F5—A steganographic algorithm: High capacity despite better steganalysis. In the Proceedings of 4th International Workshop Information Hiding, Springer- Verlag, 2001, pp. 289-302.
[9]Quantization-Based Image Steganography without Data Hiding Position Memorization Yusuke SEKI∗ , Hiroyuki KOBAYASHI†, Masaaki FUJIYOSHI∗ and Hitoshi KIYA Department of Electrical Engineering, Tokyo Metropolitan University, Hachioji-shi, Tokyo 192–0397, Japan
[10]Blind Image Steganalysis Based on Statistical Analysis of Empirical Matrix Xiaochuan Chen1, Yunhong Wang2, Tieniu Tan1, Lei Guo11 National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[11]STEGANALYSIS USING COLOUR MODEL CONVERSION P.Thiyagarajan , G.Aghila and V. Prasanna Venkatesan CDBR-SSE Lab Department of Computer Science, Pondicherry University, Puducherry 05 014
[12]Maya, S.T.; Miyatake, M.N. & Medina, R.V. Robust steganography using bit plane complexity segmentation. In the 1st International Conference on Electrical and Electronics Engineering, 2004. Mexico, pp. 40-43.
[13]The JPEG Still Picture Compression Standard Gregory K. Wallace Multimedia Engineering Digital Equipment Corporation Maynard, Massachusetts Submitted in December 1991 for publication in IEEE Transactions on Consumer Electronics
[14]Digital Image Processing By Wilhelm Burger, Mark James Burge
[15]Digital Image Processing Third Edition Rafael C. Gonzalez, Richard E. Woods