Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 4, April -2017

Using Reversible Image Transformation an Overview on Document Leakage Detection with Steganography

[1]Miss. Mrunalinee S. Patil, [2]Prof. Z.I. Khan

[1,2]*Dept of Computer Science & Engg. Dr. R.G.I.T. &R , Amravati.*

Abstract:-*The world now depends more and more on the computer and the related systems directly or indirectly for living. Information security is one of the major concerns in this System. Steganography are the most popular or widely used information security scheme or techniques. It is the act of covert communications. The algorithms for hiding data or message in a cover image, and extracting a message from a stego-image were implemented using C# programming language. In this paper, LSB Replacement technique was adopted as the embedding method. In this paper we propose a reversible high capacity data hiding method applying on binary images. We also proposed leakage detection and prevention with the help of unique hash values of the document/images. The paper proposes a novel framework for reversible data hiding in encrypted image (RDH-EI) based on reversible image transformation (RIT). Different from previous frameworks which encrypt a plaintext image into a cipher text form; RIT-based RDH-EI shifts the semantic of original image to the semantic of another image and thus protects the privacy of the original image. In our proposed system, we proposed a new mosaic image technique in which we divide the image/pdf document into parts, and the parts will then transformed, encrypted and encoded into target images.*

Keywords: *- Steganography; Stego Image; BMP Image; Least Significant Bit (LSB); Reversible image data hiding (RIDH)*

I. INTRODUCTION

1.1. Steganography

Steganography is defined as a technique to hide data into images in such a manner, which is unperceivable. Steganography and Cryptography, both are used for security purposes but with different implementation and approaches. In cryptography, the text file get converted to other form which provide confidentiality to sensitive data but in steganography we hide the actual data file in image form so that if leakage get occurred the third party fails to recognize the actual data. This provides confidentiality as well as security to the sensitive data. The idea is to hide text in image with the conditions that the image quality is retained along with the size of the image instead we can encrypt the data. So the need is, in cryptography output of an unreadable data files are being send over an internet is easily detectable that some important information is being conveyed. While in steganography hiding message in an image, along with the conditions, it make seem of just an exchange of picture between two user ends.

The steps being followed in steganography are as under:-

1. Firstly the text message is being written, then encryption of the message is done.
2. Later, text is hidden in the selected media like image file and transmitted at the receiver side.
3. At receiver end, reverse method is done to implement and recover the original text message.

1.1.1. Classification of Steganography:

For decades people try to develop innovative methods for secret communication. Classification of information hiding can be depicted as follows:

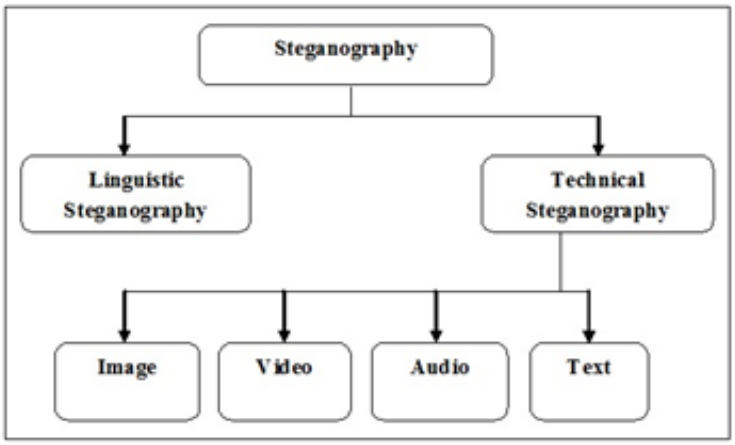

Fig.1.Classification of Steganography

Steganography is mainly of two types, linguistic steganography and technical steganography as shown in figure no.1. In linguistic steganography, machine readable data is encoded to innocuous natural language text, thereby providing security against any arbitrator tolerating natural language as a communication medium.[13] In this approach, linguistic properties of a text are customized to hide information. Language has the property that a small local change to a text, e.g. replacing a word by a word with same context, may result in text which is anomalous at the document level, or with respect to the state of the world. Hence finding linguistic transformations which can be applied reliably and often is a challenging problem for Linguistic steganography.[14]

Technical Steganography: It is a carrier rather than a text which can be presented, as any other substantial medium such as microdots and invisible inks. In this context, the cover_media is the file in which we will hide the secret_data, which may also be encrypted using the stegokey. The resultant file is the stego_media. There are four ways to implement steganography:

A. Using text.
B. Using images.
C. Using audio files.
D. Using video files

1.1.2. Process of Steganography:-

Various techniques are used in the field of steganography by arranging the different bits of the character of the text message in the image file and other media. In order to encrypt the data two files are needed: (i) image file and (ii) the text file containing the data. Our algorithm is simple and flexible using LSB (Least Significant Bit) technique. We have selected the formats that commonly use lossless compression that is BMP, PNG, TIFF and GIF. When data is streamed, it is captured after the header and chopped into 8 bits. In 24-bit BMP, Therefore comparing bit values byte by byte both of text and image. The technique we are using is LSB i.e. storing in LSB of a byte (pixel).

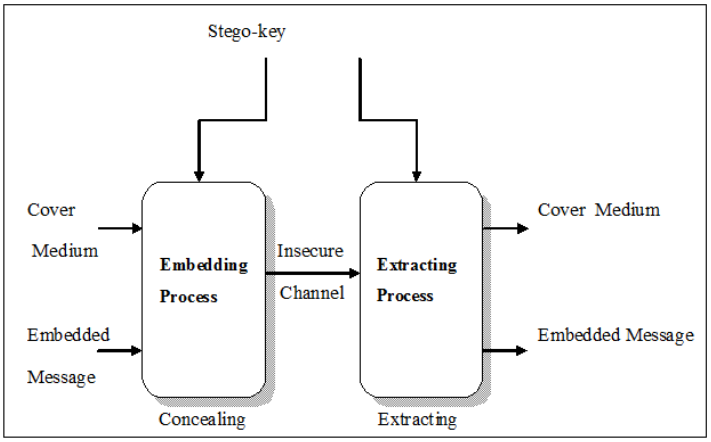

Fig.2. Steganographic Process

Some of the techniques used in steganography are domain tools or simple system such as least significant bit (LSB) insertion. In our security model, we divide the document / scanned images of document into parts. The parts will be reversibly transformed, encrypted and then encoded into target images which enhances the security of document and reduces possible attacks. At the time of file transfer we will check the access permission for the user as well as hash value of that document. If the document is an official document and particular employee is trying to leak that particular document, system will automatically prevent it and send notifications to higher authority for further actions to be taken. In order to protect the data files being open source the data leakage must be detected in the early stage.

Embedding data, which is to be hidden into an image, requires two files. The first is the image that will hold the hidden information, called the cover image. The second file is the message- the information to be hidden. When combined the cover image and the embedded
message make a stegoimage or stego-file as shown in figure 2.

Steganography system is designed for encoding and decoding a secret file embedded into an image file using random LSB insertion method in which the secret data is spread out among the image data in a seemingly random manner. This could be achieved using a secret key.

1.1.3. Steganographic Techniques

There have been many techniques for hiding information or messages in images

A. Least significant bit insertion (LSB)
B. Masking and filtering
C. Transform techniques

Least significant bits (LSB) insertion is a simple approach to embedding information in image file. The simplest steganographic techniques embed the bits of the message directly into least significant bit plane of the cover-image in a deterministic sequence.

Masking and filtering techniques, usually restricted to 24 bits and gray scale images, hide information by marking an image, in a manner similar to paper watermarks. The techniques performs analysis of the image,

Transform techniques embed the message by modulating coefficients in a transform domain, such as the Discrete Cosine Transform (DCT) used in JPEG compression, Discrete Fourier Transform, or Discrete Wavelet Transform. These methods hide messages in significant areas of the cover-image.

1.1.4. Applications of Steganography

A. Confidential communication and secret data storing
B. Access control system for digital content distribution
C. Digital watermarks
D. Modern printers

There are many applications for digital steganography of images, including copyright protection, feature tagging, and secret communications.

1.2. Data Leakage Detection (DLD) and prevention:-

Data leakage is defined as the accidental or unintentional distribution of private or secret data to an unauthorized entity. We proposed leakage detection and prevention with the help of unique hash values of the document/images. At the time of file transfer we will check the access permission for the user as well as hash value of that document. If the document is an official document and particular employee is trying to leak that particular document, system will automatically prevent it and send notifications to higher authority for further actions to be taken. In order to protect the data files being open source the data leakage must be detected in the early stage. This review paper deals with idea of preventing the data from being out sourcing by giving a special inscription to sensitive data from being reproduce using image steganography technique.

Data leakage prevention is a technique used to hide the confidentiality of data being accessed by unauthorized user

Most DLP solutions include a suite of technologies that facilitates three key objectives:

• Locate and catalog sensitive information stored throughout the enterprise

• Monitor and control the movement of sensitive information across enterprise networks

• Monitor and control the movement of sensitive information on end-user systems [19].

1.3. Reversible Image Transformation (RIT) with steganography

- At the time of document uploading, user have to choose security level of that document.
- Security level can be 1 to 4.
- System will divide the uploaded document/image into specified no of parts (1 to 4).
- The parts then transformed in reverse way i.e. The text/bytes will be arrange in reverse fashion.
- The transformed parts will be encrypted using any advanced encryption algorithm.
- The encrypted part of document/image will be then encoded into target image using LSB algorithm according to number of parts.
- Required encryption keys (as per the number of parts) will be stored into the database in encrypted format.

1.4. Reversible Image Data Hiding (RIDH)

It is a special category of data hiding technique, which ensures perfect reconstruction of the cover image upon the extraction of the embedded message. The reversibility makes such an image data hiding approach particularly attractive in the critical scenarios. The majority of the existing RIDH algorithms are designed over the plaintext domain, namely, the message bits are embedded into the original unencrypted images. The early works mainly utilized the lossless compression algorithm to compress certain image features, to vacate room for message embedding.

In this paper, we propose an encrypted-domain RIDH scheme by specifically taking the above-mentioned design preferences into consideration. The proposed technique embeds message through a public key modulation mechanism and performs data extraction by exploiting the statistical distinguish ability of encrypted and no encrypted image blocks.

II. LITERATURE REVIEW

The word steganography is originally derived from Greek words which mean “Covered Writing”. It is defined as "hiding information within a noise; a way to supplement encryption, to prevent the existence of encrypted data from being detected" [11]. It has been used in various forms for thousands of years. In the 5th century BC Histaiacus shaved a slave’s head, tattooed a message on his skull and the slave was dispatched with the message after his hair grew back [12, 13, 14, 15].

Data hiding technique [7] is a new kind of secret communication technology. It has been a hot research topic in recent years, and it is mainly used to convey messages secretly by concealing the presence of communication. There have been proposed many techniques about data hiding. A large number of popular data hiding tools, such as S-Tools 4, HideBSeek, Steganos and StegoDos [10] etc, that are based on LSB replacement. By using information hiding techniques, it is possible to fuse the digital content within the image signal regardless of the file format and the status of the image.

Photographic Experts Group (JPEG), and to a lesser extent - the Portable Network Graphics (PNG). One of the earliest methods to discuss digital steganography is credited to Kurak and McHugh [11], who proposed a method which resembles embedding into the 4 LSBs (least significant bits). They examined image downgrading and contamination which is known now as image-based steganography.

Neil et al [3] on the subject Exploring Steganography: he referred to steganography as a ‘covert writing’. In his research, he paid more attention to the selection of cover image as there is the tendency for some cover images ending in broadcasting the hidden message. He said that images are array of numbers representing light intensities at various points. He also noted that JPEG is lossy and most steganographers neither use them nor encourage its use but the 24 bit

image format such as BMP image file formats does the job well. In his experiment, 25 files and 2 message files were carefully selected.

Some recent attempts were made on embedding message bits into the encrypted images. Puech et al. [4] used a simple substitution method to insert additional bits into AES encrypted images. Opposite to the nonseparable schemes, there is another type called separable RIDH approaches, in which the data extraction and image decryption can be separately carried out, then exploited at the decoder side to reconstruct the original image. Zhang [5] designed a method to embed additional message bits into stream cipher encrypted images by flipping three LSBs of half of the pixels in a block. The data extraction can be performed by utilizing the local smoothness inherent to natural images. This method was later improved by Hong et al. [6] through a side match technique. As local smoothness does not always hold for natural images, data extraction errors can be observed in the high-activity regions. Furthermore, Zhang [5] proposed a separable RIDH method such that the protection scopes of data hiding key and encryption key are gracefully separated.

III. METHODOLOGY

3.1. Algorithm for Steganography-

LSB substitution is the process of adjusting the least significant bit pixels of the cover image. It is a simple approach for embedding message into the image. The Least Significant Bit insertion varies according to number of bits in an image. For an 8 bit image, the least significant bit i.e., the 8th bit of each byte of the image is changed to the bit of secret message. For 24 bit image, the colours of each component like RGB (red, green and blue) are changed. LSB is effective in using BMP images since the compression in BMP is lossless. But for hiding the secret message inside an image of BMP file using LSB algorithm it requires a large image which is used as a cover. The simple algorithm for OPA explains the procedure of hiding the sample text in an image. This is the reason we employ using it.

Step1: A (LSB) are substituted with the data to be hidden.
Step2: The pixels are arranged in a manner of placing the hidden bits before the pixel of each cover image to minimize errors.
Step3: Let n LSBs be substituted in each pixel.
Step4: Let d= decimal value of the pixel after the substitution.
d1 = decimal value of last n bits of the pixel.
d2 = decimal value of n bits hidden in that pixel.
Step5: If (d1~d2)<=(2^n)/2 then no adjustment is made in that pixel
Else
Step6: If(d1<d2)
d = d – 2^n
If(d1>d2)
d = d + 2^*n*.
This'd' is converted to binary and written back to pixel. This method of substitution is simple and easy to retrieve the data and the image quality better so that it provides good security[17].

IV. CONCLUSION:-

In this paper, we conclude that our proposed system provides a good and efficient way to conceal data and reached the destination in a safe manner. Steganography using LSB with more than one bit used for the hidden data gives us more space to store data. We have also presented an image steganographic system using LSB approach. In this paper we have presented an enhancement of the image steganographic system using LSB approach to provide a means of secure communication. A stego-key has been applied to the system during embedment of the message into the cover-image. In our proposed approach, the message bits are embedded randomly into the cover-image pixels instead of sequentially.

It is concluded that the original image and the final embedded image appear to be identical to the human eye. LSB makes use of BMP image, to be able to hide a secret message inside a BMP file; one would require a very large

cover image. For this reason, LSB Steganography has also been developed for use with other image file formats. This paper presents an experimental application of image steganography in a secure communication between two parties.

REFERENCES

[1] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-LSB data embedding," IEEE Trans. Image Process., vol. 14, no. 2, pp. 253–266, Feb. 2005.

[2] M. U. Celik, G. Sharma, and A. M. Tekalp, "Lossless watermarking for image authentication: A new framework and an implementation," IEEE Trans. Image Process., vol. 15, no. 4, pp. 1042–1049, Apr. 2006.

[3] X. Zhang, "Separable reversible data hiding in encrypted image," IEEE Trans. Inf Forensics Security, vol. 7, no. 2, pp. 826–832, Apr. 2012.

[4] W. Puech, M. Chaumont, and O. Strauss, "A reversible data hiding method for encrypted images," Proc. SPIE, vol. 6819, pp. 1–9, Feb. 2008.

[5] X. Zhang, "Reversible data hiding in encrypted image," IEEE Signal Process. Lett., vol. 18, no. 4, pp. 255–258, Apr. 2011.

[6] W. Hong, T.-S. Chen, and H.-Y. Wu, "An improved reversible data hiding in encrypted images using side match," IEEE Signal Process. Lett., vol. 19, no. 4, pp. 199–202, Apr. 2012.

[7] Provos, N., Honeyman, P, Hide and seek: An introduction to steganography, IEEE Security & Privacy Magazine 1 (2003) pp. 32-44.

[8] Silman, J., Steganography and Steganalysis: An Overview, SANS Institute 2001.

[9] N.F. Johnson, S. Jajodia, "Exploring Steganography: Seeing the Unseen", IEEE computer, Vol. 31, No. 2, pages 26-34, February 1998.

[10] P. Hayati, V. Potdar, E. Chang, A survey of steganographic and steganalytic tools for the Digital forensic investigator,available from: http://debii.curtin.edu.au/~pedram/images/docs/survey_of_steganograph y_and_steganalytic_tools.pdf

[11] A.Joseph Raphael, Dr.V Sundaram, "Cryptography and Steganography – A Survey", Int. J. Comp. Tech. Appl., Vol 2 (3), 626-630 ISSN:2229-6093.

[12] Gandharba Swain, Saroj Kumar Lenka,"A Hybrid Approach to Steganography Embedding at Darkest and Brightest Pixels", Proceedings of the International Conference on Communication and Computational Intelligence – 2010, Kongu Engineering College, Perundurai, Erode, T.N.,India.27– 29 December,2010.pp.529-534.

[13] Richard Bergmair ,"Towards Linguistic Steganography: A Systematic Investigation of Approaches, Systems, and Issues Oct-03 – Apr-04.

[14] Richard Bergmair. 2007 A comprehensive bibliography of linguistic steganography. In Proceedings of the SPIE Conference of security. Steganography and watermarking of multimedia contents, 6505.

[15] Chamkour Singh, Gauravdeep, "Cluster based Image Steganography using Pattern Matching", IJAIR, vol. 2, issue 5, 2013.

[16] Neil, F. and Jajodia, J. S. Exploring Steganography: Seeing the Unseen. George Mason University USA.

0018-9162/98/$10.00 © 1998 IEEE

[17] Neeta, D., Snehal, K. and Jacobs, D. Implementation of LSB Steganography and Its Evaluation forVarious Bits. I EEE Digital Information Management, 2006 1st International Conference on, Bangalore, Pp173 – 178, 6-6 Dec. 2006 .

[18] Mamta Jain and Saroj Kumar Lenka "A Review on Data Leakage Prevention using Image Steganography", International Journal of Computer Science Engineering (IJCSE), ISSN : 2319-7323 Vol. 5 No.02 Mar 2016 (56-59)

[19] Raman, Preeti, H. G. Kayacık and A. Somayaji. "Understanding Data Leak Prevention."6th Annual Symposium on Information Assurance (ASIA'11).2011.

