High Capacity Steganography with Improved Transformation and Security Technique:A survey

Bhumin Mandaliya,Vinit Gupta
Department of Computer Engineering,
Hasmukh Goswami College of Engineering, Vahelal
Gujarat Technological University

***Abstract*- Security is the primary requirement of in any area. We can apply security in two point of views. One is the cryptography and the other one is steganography. Cryptography is the way to alter the plaintext data to cipher text by some algorithm. In steganography scheme, the data is never modified and just it is hidden behind some image. Central idea of this paper is compare and study of several algorithms of applying steganography and their various ways. Finally, it concludes with comparison of all these papers and included strategies.**

***Index Terms*- Cryptography, Steganography, Trtansformation, Encryption.**

I. INTRODUCTION

The word data security has got a wide range of publicity and has become a prominent research topic in both industry and academics. Due to advent of internet, there are lots of advantages of it and equal number of disadvantages related to security present in it.

Technologies such as cryptography, watermarking and steganography are some of the tools being in use to control the problem of such hacking and to secure the confidential data being transmitted. Therefore, the protection of digital data, especially confidential data becomes more and more important [1].

The term cryptography corresponds to rearranging the data, so that it will be confidential and unreadable by user. Steganography is the term to hide our text message or data under a cover image. So that it will be totally confidential.

Watermarking is a process of hiding text behind a less transparent image applied to a piece of paper. It is an identifying an image pattern in paper that appears as various shades of lightness when viewed by light. Water marks are often used as security features to prevent counterfeiting. This technique is called as watermark, since the paper in which the image is hidden is still wet/watery [2].There are different types of steganographic algorithms broadly classified as spatial domain steganography and transform domain steganography. In spatial domain, manipulation is done directly on cover data, which hides the payload data [1]. In transform domain steganography, the cover data is transformed into frequency domain representation by using any standard transform technique. The advantage of transformed domain steganography is, first it increases the security of data, second it makes the cover media to store more data without transformation [1].

II. THEORETICAL REVIEW

Steganography is the word derived from two greek words ‘Steganos’ and ‘Graphia’ which gives meaning of covered writing. Steganography can be said as an improved version of cryptography developed by overcoming drawbacks of cryptography.

In cryptography, the secret data itself was scrambled using a secret key whereas the concept of steganography makes us to change something called as cover data based on the proposed algorithm and the secret data technically known as payload.

Different sizes of data are embedded inside the image and the PSNR (Peak Signal to Noise Ratio) has been taken for each of the images has been verified. The given scheme is considered a better scheme, if and only if PSNR value improves from the previous algorithm.PSNR will be improved only if the noise is reduced. Less the noise, better the scheme is.

The cover image is applied with the HAAR transformation technique, which gives result in the frequency domain [1].

The payload data has been applied with a unique cryptographic algorithm which makes the secret data secured even before the application of steganography, whereas the cover data is applied with the HAAR transform which is one among the wavelet transform techniques to enable the cover data to store more secret data [1].

III. DEFINITIONS AND TERMINOLOGIES RELATED TO STEGANOGRAPHY

3.1 **PSNR:-** Peak Signal to Noise Ratio is a parameter under which we can verify capacity of steganographic algorithm. High the value of PSNR parameter, high the capacity of steganographic scheme is. High capacity steganography with improved security provides high security, better PSNR and improved capacity [1].

3.2 **Cover data :-** This is the image/vedio/audio file which will hide the original message.

3.3 **Payload data :-** This is the original message/plain text which we want to hide behind an image.

3.4 **Transformation :-** It is the technique, which allows more data to be stored, since it separates different components of image like Horizontal, Vertical, Diagonal and Approximate. The advantage of representing the cover data in frequency domain using transformation technique is, first it increases the security of the data stored by the cover media without transformation, second it makes the cover media to store more data compared to data stored by cover media without transformation [1].

3.5 **Wet Rate :-** Wet Paper Codes method randomly chooses pixels as a wet pixels from a cover image and the pixels not selected become the dry pixels to embed secret message. In steganographic scheme based on WPC, wet pixel can be selected by applying a random number generator. It referes to the percentage of wet pixels in whole cover image.

IV. STEGANOGRAPHY RELATED ALGORITHMS

4.1. A new scheme for information hiding based on digital images [1]

In this scheme, author has pointed out some of the drawbacks of Least Significant Bits (LSB) steganography and resolved it using Huffman code and Wet Paper Codes (WPC)..

While working with Wet Paper Code scheme, one parameter is used i.e Wet Rate. The wet rate refers to the percentage of wet pixels in the whole cover image. The WR should not be more than 50%.

Advantage:

- The proposed scheme by author achieved high embedding efficiency and large embedding payload.
- Besides this, it also achieves some degree of security.
- Author’s scheme entirely works on pixel selection randomization, therefore it is able to achieve high security.
- Only randomly selected dry pixels can be modified, so receiver can easily extract the message without knowing the rate and distribution of wet pixel.
- From experimental results, we can say that if wet rate increases, it also increases the corresponding PSNR which signs a nice quality of proposed algorithm.

Disadvantage:

- High security cannot be provided by considering selectable operations.
- Steganalysis attack can be possible easily, because in order to detect, extract and recover the hidden message, the steganalysis uses statistical tool to analyse pixel value distribution on a suspicious image.

4.2 Two stage Color Image Steganography using Discrete Cosine Transform [2]

In frequency domain, use of color images for secret data hiding may prove to be a decisive innvation. Here author has proposed concept for two color images for data hiding for grey/color message/image.

He has also applied the Discrete Cosine Transform rather than HAAR transform.
Author has also compared value of Peak Signal to Noise Ratio(PSNR) with several transformation techniques and gained the highest value of PSNR with DCT.

Advantage:

- The proposed concept employs Discrete Cosine Transformation technique. This algorithm is compared with other existing watermarking methods like DCT-based, Z transform based, and DFT based. The comparison result involves factors like image quality assessment, Visual interpretation, Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR) and Image Fidelity (IF).
- As per the comparison, DCT results in high PSNR values and more capacity of embedding bytes compared to QFT-based, DFT-based, etc.
- Two stage DCT transformation gives highest Peak Signal to Noise Ratio (PSNR) of 45.59dB compared to other techniques.
- Two stage DCT is specifically proposed for utilization of color images in increasing the security of data hiding.
- The embedded image in this work is very difficult to detect due to dynamic insertion position of authenticating message bits in carrier message.
- The noise effect is also reduced, therefore PSNR is high.

Disadvantage:

- The extraction process is the inverse of the insertion process. To obtain the corresponding intensity value in spatial domain Inverse Discrete Cosine Transform is applied on current block.
- While using frequency values for IDCT, sometimes it generates erroneous results like:It generates negative pixel value, which may be removed by subsequent increment of the value at 1st position, which is totally invalid.
  Occurrence of fractional spatial value, because DCT is being multiplied by ½. The problem may be tested out by changing value of sum obtained after the DCT operation with an even number.

4.3 Universal image Staganalytic method based on binary similarity measures [3]

The aim of this paper was to examine the accuracy detection of universal steganalytic models for popular steganographic tools in static images and also compare time consumption in process of extraction parameters and training model .

Advantage:

- Author has proposed the steganalytic method is based on extraction of binary similarity measures using Discrete Cosine Transform. Moreover, in experiments, he has also used Support Vector Machine classifier which can deal with linear and non-linear kind of data.
- Its primary goal is to maximize the distance between two classes of parameters i.e. cover image and stego image.
- It can be used in intelligence system, since it is a cryptanalysis technique.
- It has lower computational and time cost compared to previous feature based analysis.
- The results of author’s comparison also shows that inspite of short statistic vector, it’s able to provide very similar detection accuracy compared to other steganalysis methods.

Disadvantage:

- It is vulnerable to detect secret message embedded using popular steganographic tool.

4.4 Data hiding in still images based on blind algorithm of steganography [4]

The blind steganographic methods do not require an original image in the process of extraction what helps to keep a secret communication undetected to third party user or steganalysis.

The secret message is compressed before insertion in order to enlarge the capacity of the proposed system.

Advantage:

- The blind steganography methods do not require an original image in the process of extraction which maintains secrecy of communication.
- It supports full reconstruction of secret data without having original image present on the recipient side.
- Using Haar transformation, the secret message is compressed before insertion in order to enlarge the capacity of proposed system.
- Huffman coding is a popular method for compressing data with variable length codes. Huffman coding serves as a basis for several applications implemented on various platforms. It can be used to encode secret message before it is being embedded to the cover image. Moreover, in order to hide one bit of secret message, only two bits of cover image pixels are needed, which is moderately small compared to others.

Disadvantage:

- Increasing transmission has a negative impact on the secret message perceptibility in stego image. The selection of cover image is highly important to maintain method capacity.
- Haar transform and Huffman code has excessive computation overhead.

Advantage:

- In this paper, author has proposed a method to structure this reference data in such a way so that both the stego-image and secret image suffers absolutely no distortion.
- Besides this, author has provided AES algorithm to reference data encryption for safe transmission of data and high security.
- Moreover, author has created Reference Hash Table which has two attributes. One is pixel position and the other is pixel difference.
- RHT maps corresponding secret image pixel with the cover image pixel. If the pixel value is same then difference attribute will have the difference zero. To make this process faster he has also used Memoization Lookup Table (MIT) [6] which acts like a virtual cache.
- The pixel position which is already covered is entered to the MLT, so that same position pixel can be easily obtained from MLT. Hence he has drastically reduced the processing time.
- The dual layer encryption increases the data security many folds.

Disadvantage:

- This algorithm does not work for text Steganography.
- Increases overhead, since Reference Hash Table and image needs to be transmitted separately.Has to keep track of recording of all the values of Memoization Lookup Table (MLT), since those values are further used for next step recording.
- Huffman coding is a popular method for compressing data with variable length codes. Huffman coding serves as a basis for several applications implemented on various platforms. It can be used to encode secret message before it is being embedded to the cover image. Moreover, in order to hide one bit of secret message, only two bits of cover image pixels are needed, which is moderately small compared to others.

4.5 Digital multimedia archiving based on optimization steganography system [7]

In proposed scheme, author has employed genetic algorithm scheme to boost values of PSNR.

The genetic algorithm begins with no information of exact solution and based totally on replies from its progress operators like reproduction, crossover, etc.

Advantage:

- Proposed algorithm by author works on the genetic algorithm. A Genetic algorithm is an exploration and optimization technique built on the knowledge of genetics and natural collection.

- The GA has been used in author's work to accomplish high performance ranges in multifunction domains without suffering the challenges related with high dimensionality.
- Moreover, here they have also used PSNR as a fitness function. It is done in order to obtain an optimal mapping function to reduce the difference error between the cover and stego image, therefore, improving hiding capacity with low distortions.
- It is noticeable that the stego image does not have the distortion.
- From experimental results of author, it is apparent that, the PSNR values are considerably improved in proposed method compared to previous PSNR LSB methods.
- The proposed method is very efficient to hide even large size of metadata and therefore, provide and excellent environment for creating digital media to merge with related description.

Disadvantage:

- Genetic algorithm may not always converge with a successful output. It begins with no information of exact solution and based totally on replies from its progress operators such as reproduction to get most possible solution.

4.6 Protection of Executables employing a Novel Dual Stage Digital Data hiding scheme. [8]

In this work, author has proposed a dual stage steganography technique to hide .EXE files by first embedding them behind a 2D image file followed by embedding the image in an audio cover.

Advantage:

- Steganography can be achieved using several variants like:
  Hiding image behind video
  Hiding data behind video
  Hiding data behind audio
  Hiding audio behind video
  One of the coherent requirements of steganography is that the size of the cover must be higher than that of payload. The higher the size ratio, the better hiding achieved. This is achieved by hybrid steganography.
- Author's work deals with hybrid steganography where text message is first hidden behind one image and then result is hidden behind another payload leading more complex embedding which is difficult to track.
- Proposed technique by author emphasizes on achieving high PSNR for high BPP such that the model does not require too many extra bits for stegno process.
- Two levels of security is provided, in which 1st payload is hidden behind the cover image and in next level, the result is hidden behind the audio file.
- If audio file is compromised, then also it's next to impossible to reveal the message hidden in cover image.

Disadvantage:

- It is too much difficult to apply transformation, since hybrid steps are carried out to achieve Steganography.
- BPP (Bits per Pixel) will vary according to the varying payload bits.
- Proposed algorithm works on .EXE files, but they are more sensitive to errors and even single bit error may cause application to crash.
- Besides this, .EXE file will have larger size than normal text data files. So it is difficult to manage all these parameters simultaneously.

4.7 An analysis of LSB based image steganography techniques [7]

This is the method for embedding data into cover image. The least significant bit of each pixel of an image is altered to a bit of a message that is to be hidden.

4.8 Image Steganography by closest pixel pair mapping [8]

In this paper, author has proposed the method for steganography, which results in absolutely no distortion of the cover image. In order to provide high security, author has also used AES algorithm for safe transmission of data. Besides this, he has also used Reference Hash table (RHT) and Memoization Lookup Table to store pixel position and pixel difference.

4.9 Image steganography combined with DES encryption pre-processing [10]

Here, author has proposed the scheme of combining image steganography with pre-processing of DES encryption. When transmitting the secret information, firstly, encrypt the information intended to hide by DES encryption is encrypted, then it is written in an image.

Advantage:

- Encryption algorithm improves the lowest matching performance between the image and secret information by changing statistical characteristics of secret information to enhance image steganoghraphy [10].
- The experimental results showed that the robustness of image steganography combined with pre-processing of DES encryption is found much better than the way using LSB steganography algorithm directly.

Disadvantage:

- Transformation is not applied, so encryption in frequency domain is not possible.

V. CONCLUSOIN

This paper is regarding to application of steganographic schemes in various area. I have also studied various values of PSNR (Peak Signal to Noise Ratio) in papers and decided to improve the scheme of reference paper 1 by applying other transformation technique and enhanced cryptographic algorithm. I have attempted integrate my understanding among all the surveyed papers and tried to reveal an enhanced scheme which employs new transformation technique and improved security.

	Paper	Technique Used	Algorithm	Merits	Demerits
3.1	**A new scheme for information hiding based on digital images [1]**	LSB and WPC steganography	Random number generator	High embedding efficiency and large payload	Steganalysis attack is possible
3.2	**Two stage Color Image Steganography using Discrete Cosine Transform [2]**	DCT	Discrete cosine transformation for transformation	PSNR value is moderately high	Inverse in extraction is tough
3.3	**Universal image Staganalytic method based on binary similarity measures [3]**	Support Vector Machine	SVM	Lower computational time	Vulnerable for attack using steganalytic tool
3.4	**Data hiding in still images based on blind algorithm of steganography [4]**	Blind algorithm	Haar and Huffman coding	Supports reconstruction of image	Excessive computational overhead
3.5	**Image Steganography by closest pixel pair mapping [5]**	Pixel mapping	AES for encryption	Dual layer of encryption	Not suitable for text steganography
3.6	**Multimedia archiving based on optimization steganography system [7]**	Optimization	Genetic Algorithm	Suitable to hide large size data	Convergence is not assured
3.7	**Protection of Executables**	Dual stage	Dual stage steganography	Supports	Sensitive

	employing a Novel Dual Stage Digital Data hiding scheme. [8]	steganography		hiding of .exe files	to errors
3.8	**An analysis of LSB based image steganography techniques [9]**	LSB and MSB	LSB and MSB steganography	Converts message in binary	Binary to stego image conversion is required
3.9	**Image steganography combined with DES encryption pre-processing [10]**	DES preprocessing	DES for encryption	Robustness is high	Transformation is not applied

REFERENCES

[1] Yan-Ping Zhang, Juan Jiang, Chao Xu, Bo Hua, Xiao-yan Chen. “A new scheme for information hiding based on digital images.” ICCIS-2011.

[2] Anirban Goswami, Dipankar Pal, Nabin Ghoshal. “Two stage Color Image Steganography using Discrete Cosine Transform”. Spriger-2013.

[3] Martin Broda, Dusan Levicky, Vladimir Banoci, Gabriel Bugar, “Universal Image Steganalytic method based on binary similarity measures”, IEEE-2013.

[4] Gabriel Bugar, Vladimir Banoci, Martin Broda, Dusan Levicky, Denis Dupak. “Data hiding in still images based on blind algorithm of steganography”, IEEE-2014.

[5] Raniyah Abdullah, Zenon Chaczko, Anup Kale. “Digital multimedia archiving based on optimization steganography system”, IEEE-2014.

[6]. H.S. Jayramu, Arvind K. Gautam, “Protection of Executables employing a Novel Dual Stage Digital Data hiding scheme”, International Journal of Emerging Trends of Technology in Computer Science (IJETTCS).

[7] K. Thangadurai, G. Sudha Devi, “An analysis of LSB based image steganography techniques”, IEEE-2014.

[8] Adnaan Ahmed, Nitesh Agarwal, Sabyasachee Banerjee, “Image Steganography by closest pixel pair mapping”, IEEE-2014.

[9]. G.A. Srinidhi and K. B. Shivakumar, “High capacity wet pixel based steganography with improved security”, by Springer India, 2015.

[10]. Yang Ren-er, Zheng Zhiwie, Tao Shun, Ding Shilei, “Image steganography combined with DES encryption pre-processing”, IEEE 2015.