INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

Designing and Performance Evaluation of Text Data Hiding Technique Using Sequential Encoding and Decoding Technique

Amrit Preet Kaur[1], Gagandeep Singh[2]

Chandigarh Engineering College,
Department of CSE, Landran, India
[1]kaur.amritpreet13@gmail.com

ABSTRACT *- Hidden data detection in cover images by using steganography is termed as Image steganalysis of image. Lots of new image steganographic algorithms become content-adaptive, in order to improve security and processing speed. Great challenges are possessed by advanced content-adaptive steganographic techniques to steganalyzers. These challenges are especially possessed to the feature-based blind steganalyzers. So, a technique of encoding and decoding for image is proposed which is very much generalized to the case of a source with redundancy. Also, we have introduced the Computational entropy of the source analogous to the computational cut-off rate of the channel. Also, a range of transmission rates is found. The average number of decoding computations is finite for this rate. Our proposed technique also explores the tree code in such a way to try to minimise the computational cost and memory requirements to store the tree. Actually, it gives the possibility of encoding the source output into the channel input and also of decoding the output of the channel into source symbols. This technique avoids the intermediate operations for encoding and decoding. Performance of proposed algorithm will be evaluated by calculating computational or processing speed of simulation. All simulation will be implemented on MATLAB R2008a using image processing and general tool box.*

Keywords: *- Steganography, RGB, image, encryption, extraction, embedding.*

1. INTRODUCTION

Most of the information they acquire from a system is in a form that they can read and comprehend. This is one of the reasons that intruders successful in revealing the information to others. They also modify it, to misrepresent an individual or organization. Also, they use it to launch an attack. Use of steganography is only solution to this problem. Technique of hiding information in digital media is known as Steganography. This method is not to keep others from knowing the hidden information but it is to keep others from thinking that the information even exists, in contrast to cryptography. Most of information is kept electronically, due to advances. Consequently, the information security has become a basic issue. Steganography can be employed to secure information and it is a technique of hiding information in digital media. The message or encrypted message is embedded in a digital host before passing it through the network, in contrast to cryptography, so the existence of the message is unknown to intruder. This approach can be extended to copyright protection for digital media i.e. audio, video, and images.

The possibilities of today’s communications need the special means of data security especially on computer network. Network security is very important, as the amount of data being exchanged on the Internet increases. So, there is a great need of confidentiality and data integrity, which is required to protect the data against unauthorized access and use. This has resulted in growth of the methods of information hiding. The digital copyright such as audio, video and other source, may lead to large-scale unauthorized copying, because the digital formats provides high image quality even under multi-copying. Therefore, the special part of invisible information is fixed in every image. This part could not be easily extracted without specialized technique saving image quality simultaneously [12]. All these things are of great concern to the film, music, book and software publishing companies. Hiding of information is an emerging research area, which contains applications such as protection of copyright for digital media, fingerprinting, watermarking, and steganography. In applications of watermarking, the message contains information such as identification of owner and a digital time stamp, which usually applied for copyright protection. In applications of fingerprint, the owner of the data set embeds a serial number that uniquely identifies the user of the data set. This adds to copyright information to makes it possible to trace any unauthorized use of the data set back to the user. Steganography hides the secret message within the host data set and presence imperceptible.

2. STEGANOGRAPHY

The term steganography comes from the Greek word “*Steganos*”, which mean covered or secret and *graphy* mean writing or drawing. So, steganography means covered writing. Steganography is the process of hiding information such that its occurrence can’t be detected [7] and a communication is happening [8]. Information is encoded in a manner such that the existence of the information is hided. Together with existing wireless communication techniques, steganography can be used to process hidden exchanges. The main objective of steganography is to communicate securely in a completely unpredictable manner [9] and to avoid drawing suspicion to the transmission of a hidden data [10]. It is not about keeping others from knowing the secret information, but it is about keeping others from thinking that the information even exists. If a steganography technique causes anyone to suspect the carrier medium, then the technique has failed [11]. Until, information hiding techniques received much lesser attention from the research community and from industry, in contrast to cryptography. This situation is changing rapidly. The first academic conference on this topic was organized in 1996. There has been a rapid growth of interest in steganography for two main reasons:

(i) The publishing and broadcasting industries have now become much interested in techniques for hiding encrypted copyright marks and serial numbers in digital films, audio recordings, books and multimedia products.

(ii) Moves by different governments to avoid the availability of encryption services have motivated people to study techniques by which private messages can be embedded in cover messages.

The typical model of steganography consists of Carrier, Message and Password. Carrier media is also known as cover-object, to which the message is embedded and serves to hide the presence of the message. The model for steganography is shown on Figure 1 [1]. Message is the secret data that the sender wants to remain confidential. It can be plain text, cipher-text, colored or black and white image, or anything that can be hided in a bit stream such as a copyright mark, a covert communication, or a serial number. Password is known as *stego-key*. This key ensures that only recipient who knows the corresponding decoding key will only be able to extract the message from a *cover-object*. The *cover-object* with the secretly embedded message is then called the *stego-object*.

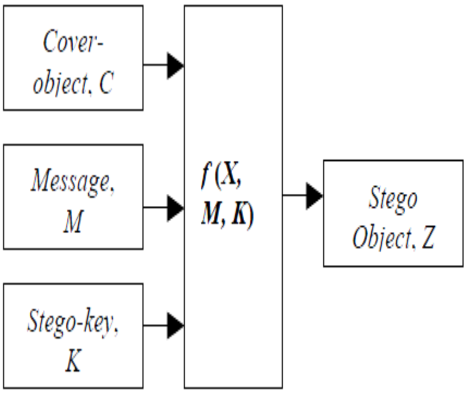


Fig. 1: Basic Steganography Model [1]

Recovery of hidden message from a stego-object requires the cover-object itself and a corresponding decoding key if a stego-key was used during the encoding process. The original cover image may or may not be required in most applications to extract the message. There are several suitable carriers below to be the *cover-object* [2]:

(i) Network Protocols such as TCP, UDP and IP.

(ii) Audio that using digital audio formats such as wav, midi, avi, mpeg and voc.

(iii) Disk and File that can hides and append files by using the slack space.

(iv) Text such as null characters, just alike morse code including html and java.

(v) Images file such as bmp, gif and jpg, where they can be both color and gray-scale.

In general, the information hiding process extracts redundant bits from *cover-object*. The process consists of two steps [4, 8].

(i) Redundant bits identification in a *cover-object*. These are those bits that can be changed without corrupting the quality of the *cover-object*.

(ii) The embedding process then selects the subset of the redundant bits to be replaced with data from a secret message. The *stego-object* is created by replacing the selected redundant bits with message bits.

3. DIFFERENT APPLICATIONS OF INFORMATION HIDING

There are number of applications of information hiding techniques, such as covert communications, Authentication, ownership proof, tracing of customer and data embedding. This application focuses on copyright protection. We broadly classify them and give examples as follows:

1) Covert Communication: In many situations, such as military and intelligence applications, officers used to send messages each other without being detected. In such cases, the adversary is usually the enemy.

2) Authentication: It is necessary to check the authenticity of input data, i.e. to determine whether the received data at hand is original, fake, or some altered version of the original. In medical applications, for instance, it is of great interest to have the original data for treatment of the patient. In that case, the data to be used may have been subjected to some unintentional changes (e.g., compression, etc.) and should not be used under such conditions. Also, in military applications, some documents may have been subjected to some changes by the enemy.

3) Identification and Proof of Ownership and Related Enforcement Issues: These applications are targeted by watermarking algorithms and intended for commercial purposes. In such a situation, a company that produces and sells digital audio clips (e.g. Sony) or a movie company that sells its products over the Internet is concerned with copyright issues. It is much profitable for hackers to crack these products and sell them at a cheaper price. Original producers would like to have legally valid proof that they are the real owners. Robust signature casting is a possible solution in such cases. A watermarking scheme would require the attackers to remove the watermark, while preserving the quality of the data, before selling the modified data. In the applications mentioned, there is nothing as such that prevents the intruders from copying the original copyrighted digital goods and using them. A precaution to get rid of such problem could be the following:

if a agreement is reached between the producers of such goods and the producers of digital media players, then the unauthorized users can possibly be discouraged from unauthorized copying through watermarking.

4) Tracing of customer: Such applications are mainly intended for the fingerprinting problem. In such a situation, for example, a movie company inserts user IDs in each product before selling it. Whenever an unauthorized user is caught playing the movie or selling it, that user and his accomplices (the parties that were involved in producing that unauthorized copy) would be identified.

5) Embedding of data in Communications: It is desirable to embed information into the host data before transmission. These are many non-adversarial applications for e.g. within an in-band captioning scenario, it is desired to send the captions of a digital video via embedding the captions in the video signal, yielding little or no distortion visually, thereby reducing the required bandwidth for data transmission.

4. EXISTING TECHNIQUES

A. A robust chaotic algorithm for digital image steganography

The steganographic scheme proposed in this article embeds a binary message in pseudo-randomly selected detail coefficients of a cover image; according to a discrete wavelet transform [6]. This helps imperceptibility since the more significant coefficients of the cover image are not altered. On the other hand, the use of a DWT results in relative robustness against steganalytic attacks as well as some image processing filters such as JPEG2000 compression. For improved imperceptibility and security, the proposed algorithm is designed to be edge adaptive. That is, a larger alteration of values is allowed at the edges of the image where it will be less visible to the human visual system. The general assumption of a binary message results in versatility of the proposed algorithm since the message can be a text file, an image, audio, video, or any other digital content, as long as it is represented as a stream of bits. Moreover, compression filters such as Huffman coding may be applied to the message for higher capacity. Similarly, although when its parameters are chosen properly the proposed algorithm is loss-less, error correcting codes may be used for improved robustness.

B. RGB Intensity Based Variable-Bits Image Steganography

The idea behind RGB image based steganography [5] algorithm is that, for ‘insignificant’ colors, significantly more bits can be changed per channel of an RGB image. Our idea is that, lower color-value of a channel has less effect on the overall colour of the pixel than the higher value. Therefore, more bits can be changed in a channel having ‘low’ value than a channel with a ‘high’ value. Therefore, authors propose the following algorithm.

1) Use one of the three channels as the indicator. The indicator sequence can be made random, based on a shared key between sender and receiver.

2) Data is stored in one of the two channels other than the indicator. The channel, whose colour value is lowest among the two channels other than the indicator, will store the data in its least significant bits.

3) Instead of storing a fixed no of data-bits per channel, no of bits to be stored will depend on the colour value of the channel. The lower the value, the higher the data-bits to be stored. Therefore a partition of the colour values is needed. Through experimentations, authors show that optimal partition may depend on the actual cover image used.

4) To retrieve the data, authors need to know which channel stores the data-bits. This is done by looking at the least significant bits of the two channels other than the indicator:

- If the bits are same, then the channel following the indicator in cyclic order stores the data.
- Otherwise, the channel which precedes the indicator in cyclic order stores the data.

Here, the cyclic order is assumed to be R-G-B-RG-B and so on. The appropriate bits can be set while the data is stored.

C. Triple-A: Secure RGB Image Steganography Based on Randomization

Figure 2 shows the Triple-A algorithm taking the message (M), the carrier image (C), and the password based generated key (K) depending on password (P), as inputs and produces the message (M) hidden inside the carrier image (C).

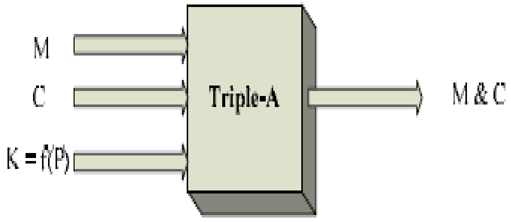


Fig. 2: Inputs and outputs of triple-A algorithm.

This algorithm can be divided into two major parts, In part one authors are interested in encrypting the message (M) using AES algorithm which will produce Enc (M, K). In our implementation the key K can be generated from a set of user passwords each with a specific key using simple XOR. This will add more security especially when it is necessarily to make the secret message available only if all the users present their passwords. In part two, the RGB Image is used as a cover media. It utilizes the advantage of the Bmp images, where every pixel is independent from the rest of the image file. Enc (M, K) is hidden according to our triple - A algorithm which needs to have a pseudorandom number generator (PRNG). The assumption for PRNG is to give two new random numbers in every iteration. The seeds of these PRNGs namely Seed1 (S1) and Seed2 (S2) are formed as a function of the Key (K). S1 is restricted to generate numbers in while S2 is restricted to the interval [3]. S1 random number is used to determine the component of the RGB image which is going to be used in hiding the encrypted data Enc (M, K).

5. PROPOSED TECHNIQUEOLOGY

To implement the proposed techniqueology MATLAB R2008a has been used as a software platform using image processing toolbox and some general commands. Proposed techniqueology is as follows:

Encryption and hiding of text message with in canvas image

Allow user to select of canvas image and text message file.

1. Allow user to select a proper encryption key.

Sequential Encoding

2. Conversion of text message into ASCII integer values including space.
3. Applying Header to Beginning of Message to be encoded.
4. Determine Message Image's Size for Encoding in Header.
5. Encrypting of message using XOR Key.
6. Conversion of integer values into binary.
7. Hide the data points using a RGBBGRRG Order behind RGB coloured image. Hiding this data along the columns moving from left to right through the target image.
8. Write Canvas Image to .BMP File. BMP, or bitmap format, is chosen because it does not use compression. JPEG compression destroys the message.

Decryption and Extraction of text message from message embedded image.

9. Import "Canvas Image" With Hidden Message.
10. Prompt User for Encryption Key to decrypt the message.

Sequential Decoding

11. Find and removal of header by taking modulus of message embedded image with 2, so as to get digital bits.
12. Next step is used to determine whether or not authors have reached the end of the image to extract the message. Authors then need to move to the next column and reset our pattern to the top row.
13. Decrypt and determine message by converting binary data into integer values.
14. Conversion of integer values into respective characters.
15. Writing of text and save text into user defined .txt file.

6. RESULTS

Simulation is carried on MATLAB R 2012a using image processing and generalized MATLAB toolbox. A colored RGB image and .txt file is taken as input to the program. Text file contains a text which is to be hiding behind the image after encoding. Then, image (containing text) is saved with a new name. This image is used for the extraction of text by sequential decoding process and extracted text will be saved in a new .txt file. The snapshot of the image used for hiding, text to be inputting, new image after hiding of text and extracted text is shown in figure 3, 4, 5 and 6. It is easily seen that both images (with text and without text) are almost analytically same and has almost no difference. Same is the case of text extracted and input text. Also, colored histogram of both the images (without text and with text) has been plotted as shown in figure 7 and 8. Both histograms are almost same, which implies that both images have same characteristics. Computational time of embedding as well as extraction has been calculated. Computational time for embedding the text is 0.1404 seconds and that for extraction is 0.1092 seconds, as shown in figure 9 and 10. At last, an output parameter i.e. capacity is also calculated for SCC, Triple- A and proposed algorithm, shown in Table 1.

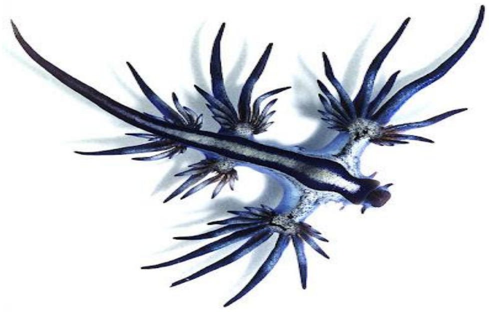
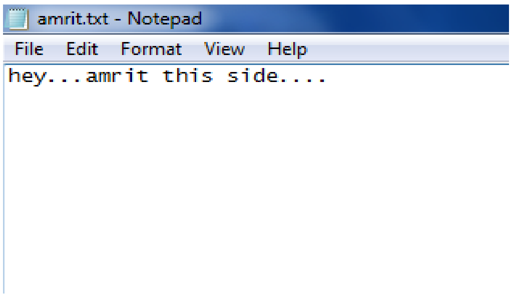


Fig. 3: Carrier image



hey...amrit this side....

Fig. 4: Input Texts

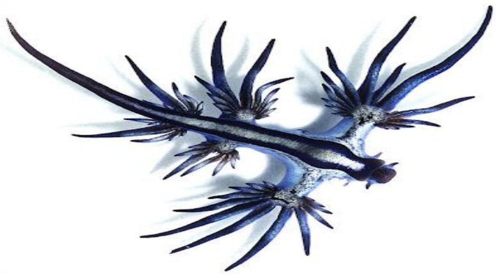
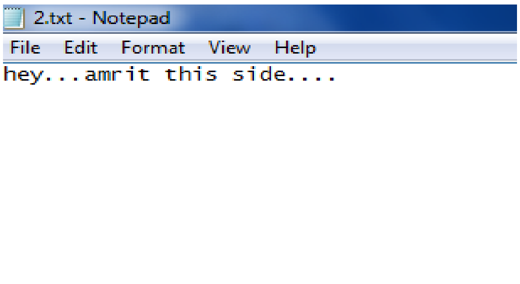


Fig. 5: Image with Text



hey...amrit this side....

Fig. 6: Extracted text

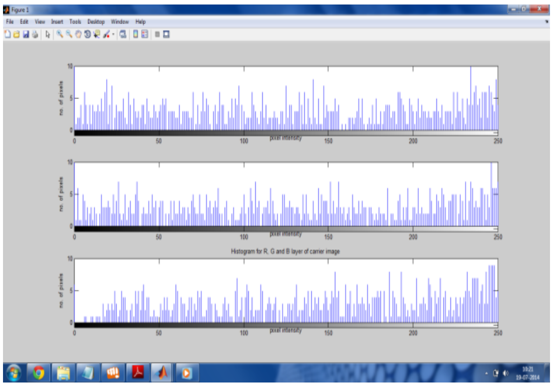


Fig. 7: RGB Histogram of Carrier image

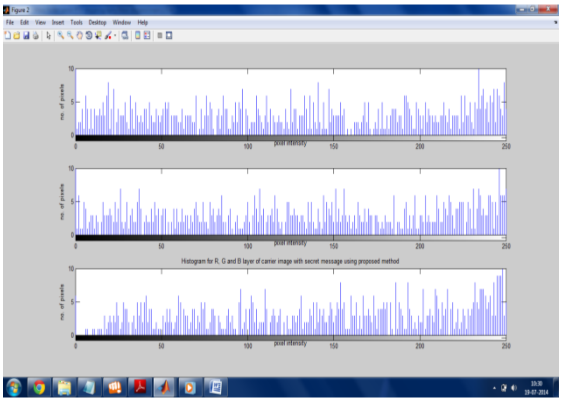
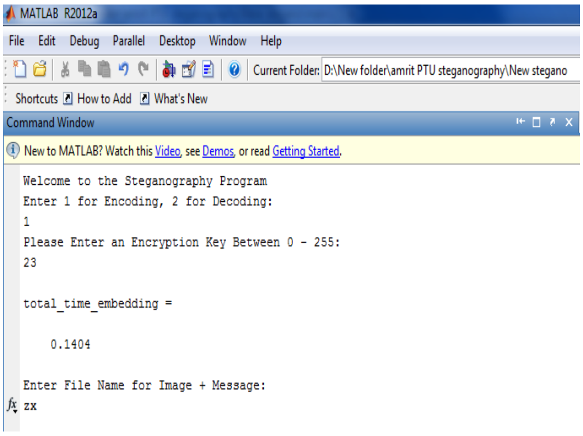


Fig. 8: RGB Histogram of image with Text

Table 1: Comparison Table of Capacity for existing algorithms as compared to proposed technique

Algorithm name	Number of pixels in carrier image	Size of text file in bytes	Capacity in %
SCC Algorithm	27984	28 KB	102.458
Triple-A Algorithm	7169	28 KB	399.94
Proposed Algorithm	2400	27 KB	112


MATLAB R2012a
File
Edit
Debug
Parallel
Desktop
Window
Help
Current Folder: D:\New folder\amrit PTU steganography\New stegano
Shortcuts
How to Add
What's New
Command Window
New to MATLAB? Watch this Video, see Demos, or read Getting Started.

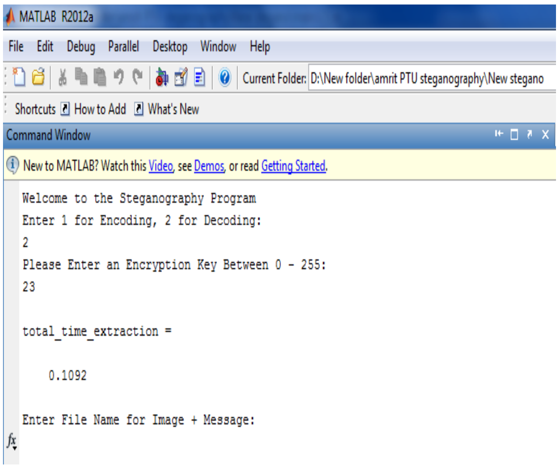
```
Welcome to the Steganography Program
Enter 1 for Encoding, 2 for Decoding:
1
Please Enter an Encryption Key Between 0 - 255:
23

total_time_embedding =

    0.1404

Enter File Name for Image + Message:
fx zx
```

Fig. 9: Computational time for Embedding


MATLAB R2012a
File
Edit
Debug
Parallel
Desktop
Window
Help
Current Folder: D:\New folder\amrit PTU steganography\New stegano
Shortcuts
How to Add
What's New
Command Window
New to MATLAB? Watch this Video, see Demos, or read Getting Started.

```
Welcome to the Steganography Program
Enter 1 for Encoding, 2 for Decoding:
2
Please Enter an Encryption Key Between 0 - 255:
23

total_time_extraction =

    0.1092

Enter File Name for Image + Message:
fx
```

Fig. 10: Computational time for Extraction

7. FUTURE SCOPE AND CONCLUSION

Steganography and steganalysis are important topics in information hiding. Steganography refers to the technology of hiding data into digital media without drawing any suspicion, while steganalysis is the art of detecting the presence of steganography. This work provides a survey on steganography and steganalysis for digital images, mainly covering the fundamental concepts, the progress of steganographic methods for images in spatial representation and in JPEG format, and the development of the corresponding steganalytic schemes. In this work, different techniques are discussed for embedding data in text, image, audio/video signals and IP datagram as cover media. We introduce a new idea in image based steganography, where variable no bits can be stored in each channel. Our algorithm uses actual colour of the channel to decide no of data bits to store. This approach leads to very high capacity with low visual distortions. Experimental results demonstrate that our algorithm performs better than other similar algorithms. All the proposed methods have some limitations. A technique is introduced as a new method to hide digital data inside image-based medium. The algorithm adds more randomization by using two different seeds generated from a user-chosen key in order to select the component(s) used to hide the secret bits as well as the number of the bits used inside the RGB image component. This randomization adds more security especially if an active encryption technique is used. The stego multimedia produced by mentioned methods for multimedia steganography are more or less vulnerable to attack like media formatting, compression etc. A secret key estimation algorithm for sequential message hiding is proposed. Extensive experiments show that the proposed algorithm is shown to perform well for stationary host signals. For non-stationary digital image data hiding, the secret key estimation accuracy is good when the embedding is done in mid and high frequency coefficients. Empirical results are presented to showcase the satisfactory performance of our proposed steganographic scheme. In particular, it is shown that the proposed algorithm has good imperceptibility according to output parameter i.e. PSNR. Furthermore, the algorithm is shown to be highly sensitive to its secret key. Finally, in comparison with some existing transform domain steganographic schemes, the algorithm proposed is shown to have superior performance. Further modifications to the algorithm are underway to improve its performance for low frequency embedding also. Also, proposed algorithm will be modified so as to employ other multimedia source as cover file. There are several ways to improve our variable bits algorithm: 1). Select the partition at run time, based on the cover media, rather than using the static (fixed) partition scheme for all cover images. 2). Use colour information of all three channels to determine the partition. This will lead to using different partition schemes for different parts of the image. Investigation into the applicability of the proposed algorithm to other low bit-rate speech codecs shall be the subject of future work. The steganalysis performance with different classifiers such as Fisher's linear classifier and logistic regression shall be part of future work.

References

[1] Chunfang Yang, Fenlin Liu, Xiangyang Luo, and Ying Zeng, Pixel Group Trace Model-Based Quantitative Steganalysis for Multiple Least-Significant Bits Steganography IEEE Transaction on information forensics and security, Vol. 8, No.1, January 2013.

[2] Yongfeng Huang, Chenghao Liu, Shanyu Tang , Steganography Integration Into a Low-Bit Rate Speech Codec IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 7, NO. 6, DECEMBER 2012.

[3] Adnan Gutub, Ayed Al-Qahtani, Abdulaziz Tabakh, Triple-A: Secure RGB Image Steganography Based on Randomization, 2009 IEEE.

[4] Shunquan Tan, Member, IEEE, and Bin Li, Member, IEEE, Targeted Steganalysis of Edge Adaptive Image Steganography Based on LSB Matchine Revisited Using B-Spline Fitting RGB Intensity Based Variable-Bits Image Steganography 2012 IEEE SIGNAL PROCESSING LETTERS, VOL. 19, NO. 6, JUNE

[5] Mohammad Tanvir Parvez and Adnan Abdul-Aziz Gutub, RGB Intensity Based Variable-Bits Image Steganography 2008 IEEE Asia-Pacific Services Computing Conference.

[6] M. Ghebleh, A. Kanso, A robust chaotic algorithm for digital image steganography, Commun Nonlinear Sci Numer Simulat 19 (2014) 1898–1907.

[7] Guo-Shiang Lin, Yi-Ting Chang, and Wen-Nung Lie, A Framework of Enhancing Image Steganography With Picture Quality Optimization and Anti-Steganalysis Based on Simulated Annealing Algorithm, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 12, NO. 5, AUGUST 2010.

[8] Graeme Bell and Yeuan-Kuen Lee, A Technique for Automatic Identification of Signatures of Steganography Software IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 5, NO. 2, JUNE 2010.

[9] Jun Zhang and Dan Zhan, Detection of LSB Matching Steganography in Decompressed Image, IEEE SIGNAL PROCESSING LETTERS, VOL. 17, NO. 2, FEBRUARY 2010.

[10] Weiming Zhang and Xin Wang, Generalization of the ZZW Embedding Construction for Steganography, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 4, NO. 3, SEPTEMBER 2009.

[11] Zhiyuan Zhang, Ce Zhu *a*nd Yao Zhao, Two-Description Image Coding With Steganography, IEEE SIGNAL PROCESSING LETTERS, VOL. 15, 2008 887.

[12] Jinwei Wanga,b,n, ShiguoLian c, On the hybrid multi-watermarking, Signal Processing 92 (2012) 893–904.