ISSN No. 0976-5697

Volume 8, No. 5, May – June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Steganography Technique Based on the Huffman Codes and PM1 Technique

Sonal Dwivedi
Department of Computer Science and Engineering
Jamia Hamdard
New Delhi, India

Md. Tabrez Nafis
Department of Computer Science and Engineering
Jamia Hamdard
New Delhi, India

Abstract: In this age of technology, everyday huge amount of data is transferred across the network. The higher the amount of data transmission the larger the amount of security is needed. In this scenario security of the data stands as the basic need and it needs to be addressed and solved first. In this paper combination of a standard table and Plus Minus 1 (PM1) steganography using Genetic Algorithm (GA) is proposed which would not only allow transmission of a large amount of data but would also eliminate the need of embedding the look-up table along with the image.

Keywords: Steganography, PM1 technique, LSB matching, Genetic Algorithm, Huffman Code

I. INTRODUCTION

A. *Steganography*

The literal meaning of steganography is concealed writing. This technique is used to camouflage a text, image, audio, video or file within another document, audio, video or image. This technique puts the secret data in an incomprehensible form which does not catch attention for scrutiny.

B. *LSB Matching/Plus Minus 1 embedding*

LSB matching is an improved version of LSB replacement based steganography technique. This improved technique is much more difficult to detect in spatial domain images. This technique makes use of randomly increasing or decreasing the value of the Least Significant Bit (LSB) by one to change the original value with the bit to be embedded.

C. *Genetic Algorithm*

Genetic Algorithm is a search and optimization technique which is based on Darwinian principles of survival and reproduction (Goldberg 1989). This algorithm continuously modifies the population of individual solutions. The chromosome in GA is represented via binary encoding. Each chromosome represents a candidate solution in the searching space. A fitness function is usually needed in GA to assign a score (fitness) to each chromosome in current population. The initial population of individuals in GA is initialised by guess. The iterations through the individuals evolve is called generations. Genetic operators like selection, crossover and mutation are used for individuals to generate next generation of individuals. The process is carried out until some given criterion (e.g. fitness) is met. The genetic operators which are used to control the population of chromosomes in the simplest GA are described as follows:

- Selection

  This operator selects the chromosomes in the population for reproduction depending upon the fitness of the chromosome. The fitter the chromosomes the better are the chances of promoting the information they contain to the next generation.

- Crossover

  Crossover does its operation by randomly choosing a point to exchange the sequence of binary before and after that point between two chromosomes selected by the selection operator. This operation further reproduces the next generation of the offspring as shown below.

  Parent Chromosome1
  1111101|0101001110010000111101
  Parent Chromosome2
  0010101|0101111111100001111010
  Offspring
  1111101010111111111000011110 10
  0010101010100111001000011111 01

- Mutation

  This operation can be performed by randomly flipping the value of the single bit within the randomly selected chromosome from the population as shown in below given example.

  Parent Chromosome
  1111101010100111001000011111 01
  Offspring
  1011111110100111101000001111 01

II. LITERATURE REVIEW

The authors of "Stochastic approach to secret message length estimation in ±k embedding steganography" talk about a new method for the estimation of the number of embedding changes for non-adaptive ±k embedding in images [1]. Another method was introduced by the author of "LSB matching revisited" who uses the choice to set a binary function of two cover pixels to the desired value. The embedding is performed using a pair of pixels as a unit, where the LSB of the first pixel carries one bit of the information, and a function of the two pixel values carries another bit of information. Therefore, the modified method allows embedding the same payload as LSB matching but with fewer changes to the cover image [2]. The novelist of "A High-Capacity Steganography Scheme for

JPEG2000 Baseline System" talks about increasing the hiding capacity by their proposed JPEG2000 baseline system [3]. One of the most effective way to ensure the security and embedding capacity was implementing PM1 steganography using GA algorithm [4]. The authors for "Hiding data in images by simple LSB substitution" talks about hiding the data by using simple LSB substitution [5]. The method of LSB substitution using genetic algorithm was also proposed in an order to embed the data inside the image [6-8]. In paper [9], the authors proposed an alphanumeric technique for encryption using ciphers and number system. The proposed algorithm can be used in transmitting health care data during inter-organizational collaboration. Patient data can be encoded into alphabetical codes that cannot be cracked. Its advantage is that the codes look similar to medical codes (International Codes for Diseases, ICD) but actually are encrypted text. So it's a secure way of transmission of sensitive patient-data. The authors in [10] discussed about the quality and meaning of the sentence whenever the sentence was broken down at the second position VGNN which is the third case in most of the sentences, where the verb is allowed to complete the action. Also the errors were analyzed where the simplification parser was not able to return expected results. A list of those scenarios was presented. Working on these errors will significantly improve the accuracy of the shallow parser.

III. PROPOSED WORK

The embedding capacity along with security is an essential need of the steganography process. Currently existing algorithms focus either on capacity or on security but achieving both aspects at the same time still remains as a challenging task. PM1 steganography not only prevents the typical attacks against LSB based technique but also provides high embedding capacity. GA optimizes the performance by minimizing the factors like blockiness. The standard table is created using the Huffman code, which would be unique and its encryption along with the secret text would not be required inside the image. In case some un-authorized person manages to crack the Huffman code, without the standard Huffman table it would be impossible to crack the secret message. This is the central idea behind the proposed work. Creation of standard table:

Step 1: Assign the frequency to all alphabets (A-Z) as 1.

Step 2: Construct the Huffman tree of these alphabets (A-Z).

Step 3: Store the Huffman code of the corresponding alphabets in a reference table for future reference. This table will be our standard table.

While creating the table the ascending order of the alphabets must be taken into consideration in an order to obtain the same Huffman table every time [11].

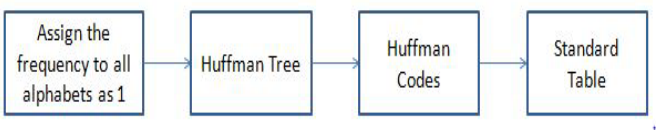


Figure 1. Steps to create standard look-up table

The embedding procedure of the proposed method can be divided into following steps:

Step 1: To attain the DCT coefficients 'X', applies entropy decoder and decode the JPEG image. Then using key based permutation (straddling mechanism(Westfeld 2001)) shuffle all the coefficients to obtain DCT coefficients X_p. Due to permutation the secret bits are scattered all over the cover medium, and the embedding density can be same everywhere (Westfeld 2001).

Step 2: Use the standard table to convert the secret text into binary mode. Here the length of the secret text will be represented by L(represented using 15 bits).

Step 3: Compare the combined message bit by bit with their corresponding non-zero AC coefficients to determine the number of coefficients that need modification, which is the length of chromosomes, Lc, used later in the GA. Then, the GA algorithm is used to select the optimal plus/minus solution for each coefficient to be modified (PM1 steganography).

Step 4: Using the plus/minus solution from the above step modify the corresponding non-zero AC coefficients. This would give the permutated version of stego quantised DCT coefficients X'_p.

Step 5: The permutated version of stego quantised DCT coefficients (X'_p) are inversely permuted to their original sequence version of X'. These are then delivered to Huffman encoder to get the stego JPEG image.

Step 6: The key and the stego JPEG image is then transferred to the receiver.

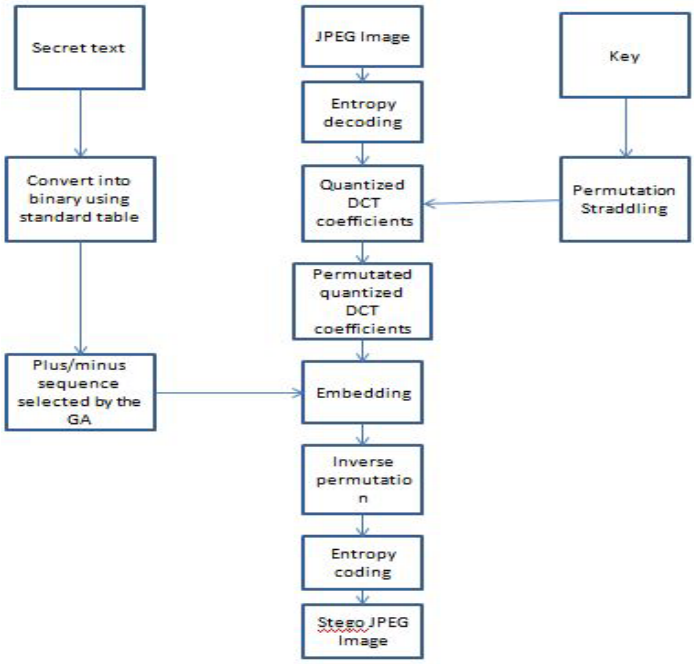


Figure 2. The embedding procedure.

The extraction procedure would be as follows:

Step 1: After receiving the key and the stego JPEG image, use the entropy decoder to recover the quantised stego coefficients X'. To attain the permuted DCT coefficients X'_p, the stego coefficients X' are shuffled based on the receiving key.

Step 2: From first 15 coefficients of X', the length of the secret message, L, is obtained. After this the secret message bits are extracted from the successive L non-zero AC coefficients.

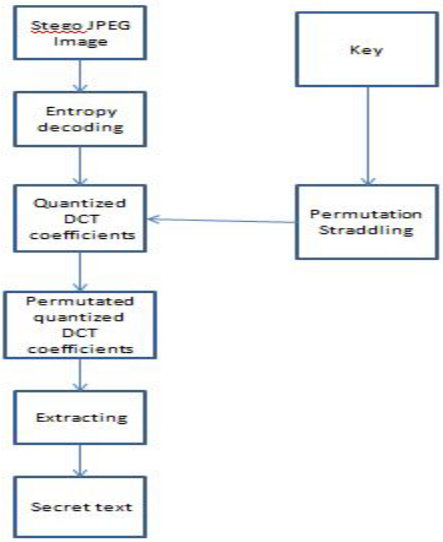


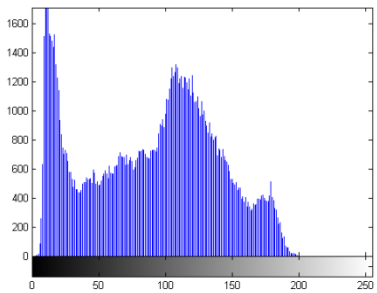
Figure 3. The extraction process.

IV. RESULTS

This section presents the experiments which were carried out to test the working of the proposed method. This method is stimulated using MATLAB R2013a on Windows 10 platform. A set of JPEG images were used to hide the secret text in each of them. The histograms show the difference between the original and the stego image.

Original Image

a) Flower

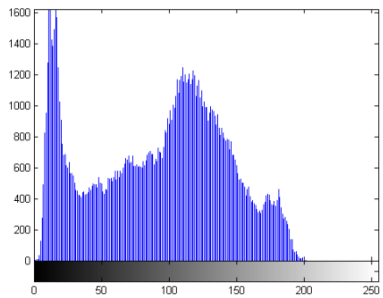


Histogram of Flower

a) Lion

Stego Image

b) FlowerSteg

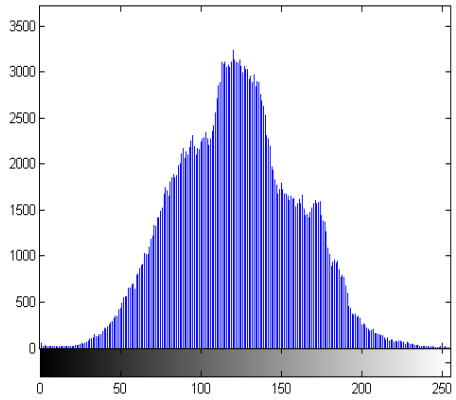


Histogram of FlowerSteg

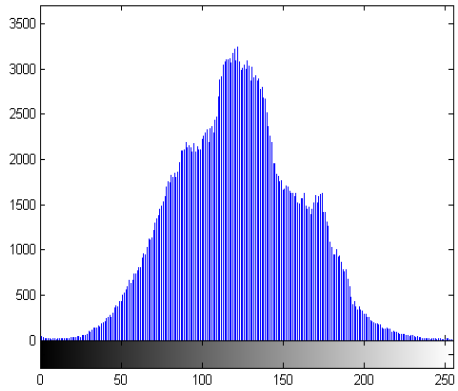
b) LionSteg

V. CONCLUSION

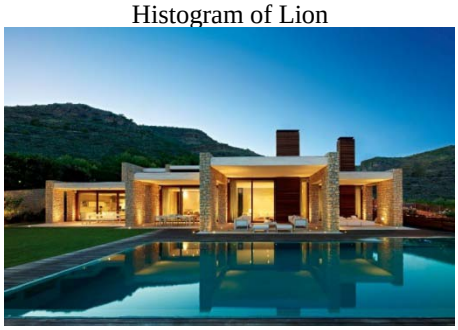
In this paper a combination of PM1 steganography using GA and standard table is introduced. The standard table is created using huffman coding technique. This standard table eliminated the need of encrypting the table inside the image for decryption purpose hence enhancing the security protocol of the image. The PM1 steganography is proved to be secure when it comes to the various attacks like $\chi 2$, along with that high embedding capacity is also achieved using this method. The optimisation of PM1 embedding is achieved by using GA. Hence, it can be concluded that the proposed combination of PM1 steganography using GA and standard table is not just effective in terms of capacity but also in terms of security.



Histogram of Lion

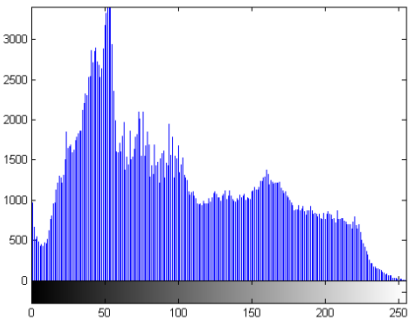


Histogram of LionSteg

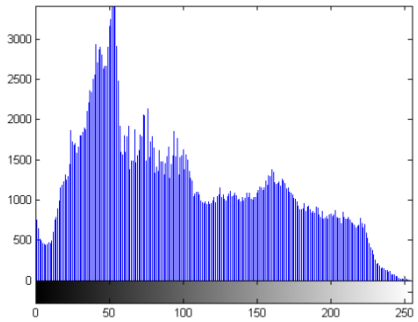


a) House

b) HouseSteg



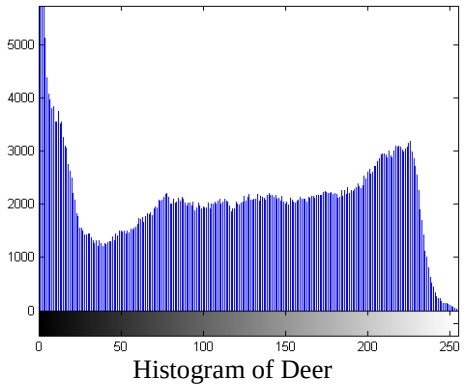
Histogram of House



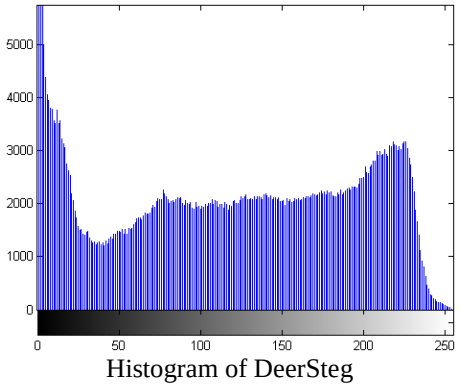
Histogram of HouseSteg

a) Deer

b) DeerSteg



Histogram of Deer



Histogram of DeerSteg

VI. REFERENCES

[1] Holotyak, T., Fridrich, J. and Soukal, D., 2005, March. Stochastic approach to secret message length estimation in±k embedding steganography. In *Electronic Imaging 2005* (pp. 673-684). International Society for Optics and Photonics.

[2] Mielikainen, J., 2006. LSB matching revisited. *IEEE signal processing letters*, *13*(5), pp.285-287.

[3] Zhang, L., Wang, H. and Wu, R., 2009. A high-capacity steganography scheme for JPEG2000 baseline system. *IEEE Transactions on Image Processing*, *18*(8), pp.1797-1803.

[4] Yu, L., Zhao, Y., Ni, R. and Zhu, Z., 2009. PM1 steganography in JPEG images using genetic algorithm. *Soft Computing*, *13*(4), pp.393-400.

[5] Chan, C.K. and Cheng, L.M., 2004. Hiding data in images by simple LSB substitution. *Pattern recognition*, *37*(3), pp.469-474.

[6] Wang, R.Z., Lin, C.F. and Lin, J.C., 2001. Image hiding by optimal LSB substitution and genetic algorithm. *Pattern recognition*, *34*(3), pp.671-683.

[7] Wang, R.Z., Lin, C.F. and Lin, J.C., 2001. Image hiding by optimal LSB substitution and genetic algorithm. *Pattern recognition*, *34*(3), pp.671-683.

[8] Wang, S., Yang, B. and Niu, X., 2010. A secure steganography method based on genetic algorithm. *Journal of Information Hiding and Multimedia Signal Processing*, *1*(1), pp.28-35.

[9] Tiwari, P., Madaan, N. and Nafis, M.T., 2015. Cryptographic Technique: Base Change Method. *International Journal of Computer Applications*, *118*(14).

[10] Tiwari, P. and Nafis, M.T., 2015. Error Patterns and Analysis of Hindi Shallow Parser. *International Journal of Computer Applications*, *110*(14).

[11] Dwivedi, S. (2016). An optimized steganographic technique. *CSI Transactions on ICT*, 1-9. doi: 10.1007/s40012-016-0116-x