Efficient data hiding scheme using lossless data compression and image steganography

Rahul Jain

U.I.E.T, Kurukshetra University,
Kurukshetra, Haryana 136119,India
eng.rahuljain@gmail.com

Naresh Kumar

Assistant Professor, U.I.E.T, Kurukshetra University,
Kurukshetra, Haryana 136119,India
naresh_duhan@rediffmail.com

Abstract: Steganography is an art of hidden communication in which secret message is embedded into a cover image. It has many applications like Online transactions, military communication etc. In this paper, we have proposed a data hiding scheme using image steganography and compression. This scheme can be applied to gray scale as well as color images. This scheme improves the data hiding capacity of the image as compared to other existing image steganography methods while retaining the quality of the image after embedding the secret message into it. The improved embedding capacity of the image is possible due to preprocessing the secret message in which a lossless data compression technique is applied.

Keywords: Steganography, Compression, Peak Signal to Noise ratio, Mean Square Error.

1. Introduction

Internet is the most popular medium that exchange information between parties. Most important factor of information technology and communication is the security of information. One of the aspects of information security is information hiding. Generally Information security means protecting information from unauthorized access, disruption, modification or simply illegal use. Moat widely used data hiding techniques are digital signature, cryptography and steganography. The word steganography comes from Greek word “steganos” which means covered or secret and the “graphy”s means writing or drawing. So, literal meaning of steganography is “covered writing” [9]. Generally steganography is known as invisible communication. Cryptography provides confidentiality, steganography on the other hand hide the message and there is no knowledge of the existence of the message. In simple words, it is hiding the information into other information. Steganography do not alter the message structure but hides the message inside a cover object. The cover object can be text, image, audio, video etc.

Basic steganography diagram is shown in figure 1. In the basic steganographic process, the secret message is hidden into a cover object. The cover object can be any of text, image, audio, video etc. A secret key is also used and the secret message is embedded into the cover object using the secret key. This new message obtained is called stego message. The stego message is transmitted over the public channel. The receiver gets the message and retrieves the message using the stego key which is same as used by the sender. In this way security is achieved by hiding the existence of the message.

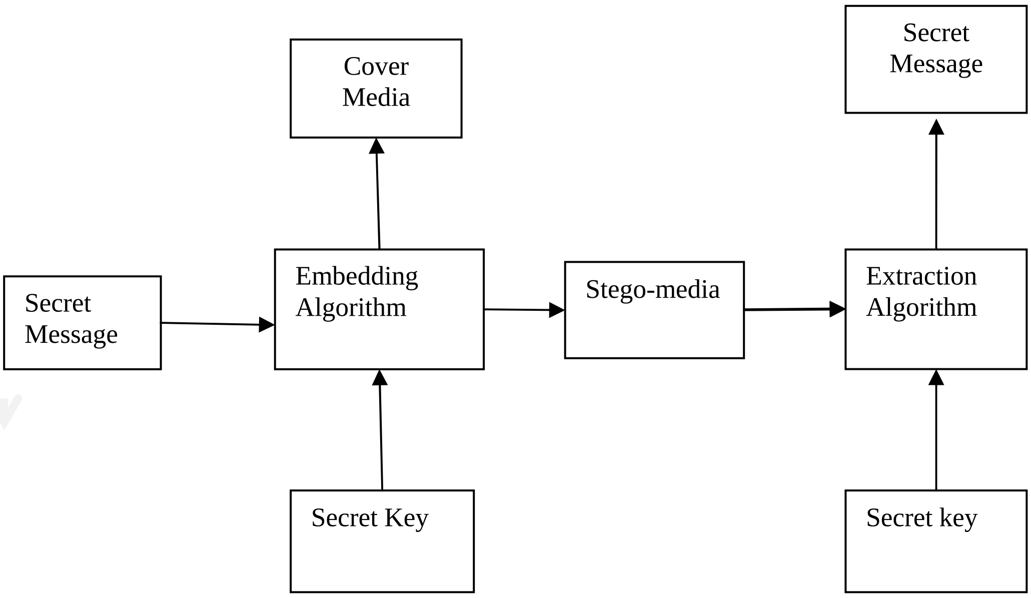


Fig 1 Basic steganographic process

Image steganography is a process that hides the message into cover-image and generates a stego-image. That stego-image then sent to the receiver without anyone else knowing that it contain the hidden message. The receiver can extract the message with or without stego-key that depends on the hidden scheme [5]. Image steganography techniques can be divided into two groups:

- Image Domain also called spatial domain and
- Transform Domain also called frequency domain [10].

Spatial domain techniques embed information in the intensity of the original image pixels directly. Basically least significant bit (LSB) method is used where it replaces the least significant bit of original pixel with the message bit.

Transform domain also known as frequency domain where images are first transformed then the message is embedded in the image. *Discrete cosine transformation* (DCT) technique is used in JPEG images to achieve compression. DCT is a lossy compression transform where the cosine values cannot be generated as original, because DCT alter values (e.g. 8.667 to 9) to hide the information.

2. Related Work

Steganography is an area of invisible communication. It is not a modern technique which is used for protecting the unauthorized access of the information but is an ancient technique which is in existence since 440 B.C.

The most basic and important image Steganographic Technique is Least Significant Bit [3,12] embedding technique. In this technique data can be hidden in the least significant bits of the cover image and the human eye would be unable to notice the hidden image in the cover file. This technique can be used for hiding images in 24-bit, 8-bit or gray scale format. In this technique, least significant bit of each pixel is replaced with secret message bit until message end. But this technique has very less embedding capacity and is easy to detect. K Suresh Babu [7] discussed an image Steganography technique that can verify the reliability of the secret message that is transmitted to the receiver through the network. If the attacker has tried to hack the secret information in the stego-image and changes it, this method can easily track that. In this technique, the hidden information is embedded in the spatial domain of the cover image and uses two special AC coefficients of the Discrete Wavelet Transform domain to verify the integrity of the secret information from the stego image.

Marvel [8] discusses spread spectrum image steganography technique. In spread spectrum techniques, hidden data is spread throughout the cover-image making it harder to detect. In spread spectrum image steganography the secret message is embedded in noise and then combined with the cover image which results into the stego image. Since the power of the embedded signal is much lower than the power of the cover image, the embedded image is not perceptible to the human eye or by computer analysis without access to the original image.

Swain & Lenka [4] discuss a technique which uses both cryptography and image steganography and hence provides more security. In this technique the secret message is encrypted by using a new cipher which is extended from Hill cipher. Then the cipher text of the secret message is embedded into the cover image in 6th, 7th and 8th bit locations of the darkest and brightest pixels. Swain and Lenka also discuss another technique [17] which provides more payloads as compared to their previous technique. In this technique, Firstly, the secret message is encrypted using new cipher algorithm called twelve square substitution cipher algorithm, and then embed the cipher text in the carrier image in 6th and 7th bit locations or 7th and 8th bit locations or 6th and 8th bit locations of the different pixels of the cover image depending on the value of an index variable whose initial value depends upon the length of the cipher text. [1]

In wu and Tsai's et al., proposed a method called pixel value differencing method. In this method, the size of the hidden data bits can be estimated by difference between the two consecutive pixels in cover image using simple relationship between two pixels. PVD method generally provides a good imperceptibility by calculating the difference of two consecutive pixels which determine the depth of embedded bits.

Modified Kekre's Algorithm (MKA) [6] is based on Least Significant Bit (LSB) method. MKA can be applied on 8 bit gray scale images or 24 bit Read Green Blue (RGB) color image. It uses up to four LSB's of a pixel to embed the data. The number of secret data bits that can be embedded in the pixels depends upon the pixel intensity of the pixels of the cover image. MKA uses 8 bit secret key to perform XOR operation to all the bytes of the secret message to achieve security. XOR operation is also performed using the same key while extracting the message.

3. Proposed Work

In the previous techniques discussed, if the data embedded in the image is increased, the image quality deteriorates. So, we cannot embed sufficiently large data into the cover image. In our proposed technique we overcome this problem. In this technique, secret data is preprocessed first and then the preprocessed secret data is embedded into the LSBs of the cover image depending upon the intensity of the pixel values of the cover image. For pre-processing a lossless data compression technique, LZW (Lempel–Ziv–Welch) technique is used for pre-processing the data. In this technique sequence of 8-bit secret data is encoded as fixed-length 12-bit codes. The code from value 0 to 255 represents one character sequences consisting of the corresponding 8-bit character. As the data is encoded, the codes with values 256 through 4095 are created in a dictionary depending upon the sequences encountered in the data. A dictionary is initialized to contain the single-character strings corresponding to all the possible input characters. At every step in the compression process, input characters are gathered into a sequence until the next character comes that will make a sequence for which there is no code in the dictionary. The code for the sequence without the character encountered is emitted, and a new code for the sequence with the character encountered, is added in the dictionary. The algorithm works by scanning the input secret data for successively longer substrings until a string is found that is not in the dictionary. When such a string is found, the index for the string without the last character is fetched from the dictionary and sent to output, and the new string including the last encountered character is added to the dictionary with the next available code. The last input character is then used as the next starting point to scan for substrings. In this way, successively longer strings are added in the dictionary and made available for subsequent encoding as single output values. This compression technique gives best results on the secret data with repeated patterns. Steganography technique used is Modified Kekre Algorithm (MKA). In this technique, firstly 8-bit secret bit is selected. The secret bit is XORed with all the bytes of the secret message that is to be embedded into the cover image. For the pixels of the cover image having intensity value greater than 239; if the bit 1 is to be embedded then 5 bits of secret text are embedded and for embedding bit 0, 4 bits are embedded in LSBs of that pixel. For a pixel having intensity from 224 to 239; if bit to be embedded is 0, 5 bits of the secret message are embedded and for bit 1, 3 bits are embedded. For a pixel having intensity value in the range of 192 to 223, 2 bits are embedded otherwise only one bit is embedded. Embedding and extraction Algorithms are:

A. Embedding Algorithm

i. Extract secret message from file into an input_array.

ii. Pre-process the secret message by applying the compression technique as follows

  1. Initialize the dictionary_array to contain all strings of length equal to one.
  2. Find the longest string L in the dictionary_array that matches the current input data from the input_array.
  3. Emit the dictionary_array index for L to output_secret_data and remove L from the input message.
  4. Add L followed by the next symbol in the input to the dictionary_array.
  5. Go to Step 2.
  6. Repeat the steps from 2 to 5 upto end of the array.

iii. Represent the total size of the secret message (represented in bytes) into 16 bit binary form.

iv. Maintain a single array named encoded_message for the total size of the secret message generated in step (iii) and encoded message generated in step (ii).

v. Embed the elements of encoded_messgae into the pixels of the cover image from 2nd pixel by analysing as follows
   - If the pixel intensity of the cover image is greater than 239 we embed four elements of encoded_messgae into first four LSBs of that pixel.
   - Otherwise if it is less than and equal to 239 and greater than 223 then three elements are embedded into first three LSBs of that pixel.
   - Else if it is less than and equal to 223 and greater than 192 then 2 elements are embedded into first two LSBs of that pixels.
   - Otherwise only one element is embedded into very first LSB of that pixel.

vi. Repeat step (v) until last element of encoded_messgae is embedded.

vii. Obtained image is the stego image.

B. Extraction algorithm

i. Extract LSBs of the pixels, starting from the first pixel of stego image as follows
   - If the pixel intensity of the stego image is greater than 239 then extract first four LSBs of that pixel.
   - Otherwise if it is less than and equal to 239 and greater than 223 then extract the first threes.
   - Else if it is less than and equal to 223 and greater than 192 then extract the first twos.
   - Otherwise only the first LSB is extracted.

ii. Repeat step (i) until we get 16 bits. Its decimal equivalent represents the length of the secret message.

iii. Repeat step (i) until we get the bits equal to the size of the message. The obtained data is named encoded_input.

iv. Now using the information obtained in step (ii), decode the encoded_input obtained in step (iii) as follows.
   - Read a value from the encoded_input and output the corresponding string from the initialized dictionary.
   - Obtain the next value from the encoded_input, and adds to the dictionary the concatenation of the string just output and the first character of the string obtained by decoding the next input value.
   - Then proceed to the next input value (which was already read in as the "next value" in the previous pass).
   - Repeats the process until there is no more input, at which point the final input value is decoded without any more additions to the dictionary.

v. Secret message is obtained at the receiver site.

By using the compression technique first and image steganography technique then, sufficient payload is achieved and without compromising with the quality of the image which is used to embed the secret data. Sufficiently large amount of data can be transferred in a more secure way by using our proposed technique.

4. Experimental Results

We carry out experiments by taking most widely used images and some other images for evaluating their performances and compared them with some existing techniques. The image quality metrics used are Peak Signal to Noise Ratio (PSNR), Mean Square Error (MSE) and Root Mean Square Error (RMSE). The reason of using MSE and PSNR as quality metrics in our experiments is that these are the most widely used in the literature. High PSNR value and low MSE value signifies good quality image. PSNR is measured in decibel (db). The data hiding capacity is calculated in bytes.

We take various color images for hiding the secret data. The images taken in our experiments include Lena, Pepper, and Baboon each of size 512x512 pixels and some other images of different dimensions. The reason for making comparison with these methods is that they are more recently developed and have good performance. The secret data taken in our experiments is Abraham Lincoln's letter to his son's teacher that is embedded into each of these images which is of size 1785 bytes. The resultant stego images with hidden secret message, employing our proposed method are shown in Figures below. The performance results are shown in Table 1, 2, 3 and 4.

Table 1 PSNR values of Different Approaches on different Images.

Cover Image	**Modified Kekre Algorithm**	**Proposed Algorithm**
Barbara.jpg	60.9421	63.7030
Baboon.jpg	73.3934	76.0391
Football.jpg	68.2248	70.8976
Peppers.png	72.0161	74.8332
Lena.bmp	73.2938	76.0108

Table 2 MSE values of Different Approaches on different Images.

Cover Image	**Modified Kekre Algorithm**	**Proposed Algorithm**
Barbara.jpg	0.0523	0.0277
Baboon.jpg	0.0030	0.0016
Football.jpg	0.0098	0.0053
Peppers.png	0.0041	0.0021
Lena.bmp	0.0030	0.0016

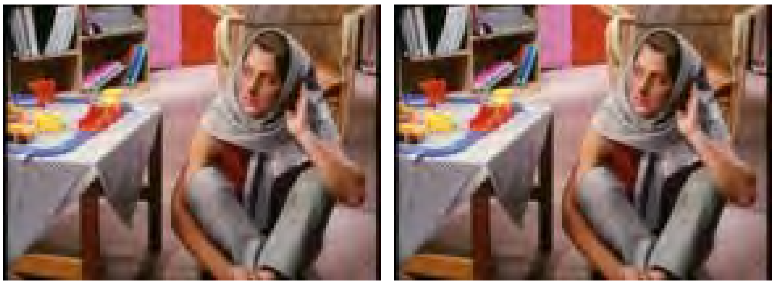
Table 3 RMSE values of Different Approaches on different Images.

Cover Image	**Modified Kekre Algorithm**	**Proposed Algorithm**
Barbara.jpg	0.2288	0.1665
Baboon.jpg	0.0546	0.0402
Football.jpg	0.0989	0.0727
Peppers.png	0.0639	0.0462
Lena.bmp	0.0552	0.0404

Table 4 CAP (Maximum Embedding Capacity) values in bytes of Different Approaches on different Images.

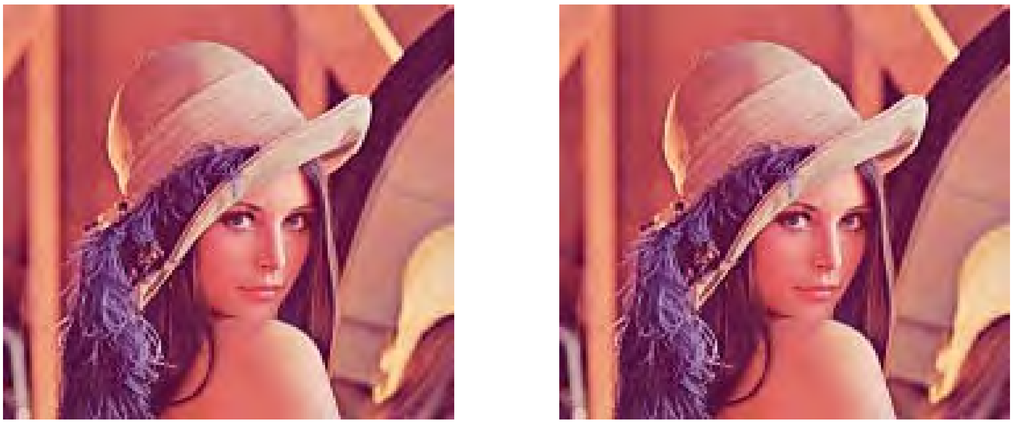
Cover Image	**Modified Kekre Algorithm**	**Proposed Algorithm**
Barbara.jpg	7371	13104
Baboon.jpg	117631	208550
Football.jpg	39780	70720
Peppers.png	114718	204355
Lena.bmp	127402	226026

It is evident from the above tables that the proposed technique is better than the existing technique and produces better results. For every image the value of PSNR, MSE and CAP i.e. maximum embedding capacity of our proposed technique is more than the MKA technique [6]. The Capacity of all the cover images to embed the secret data increases by applying the proposed technique. The security of the secret data also increases due to its preprocessing.



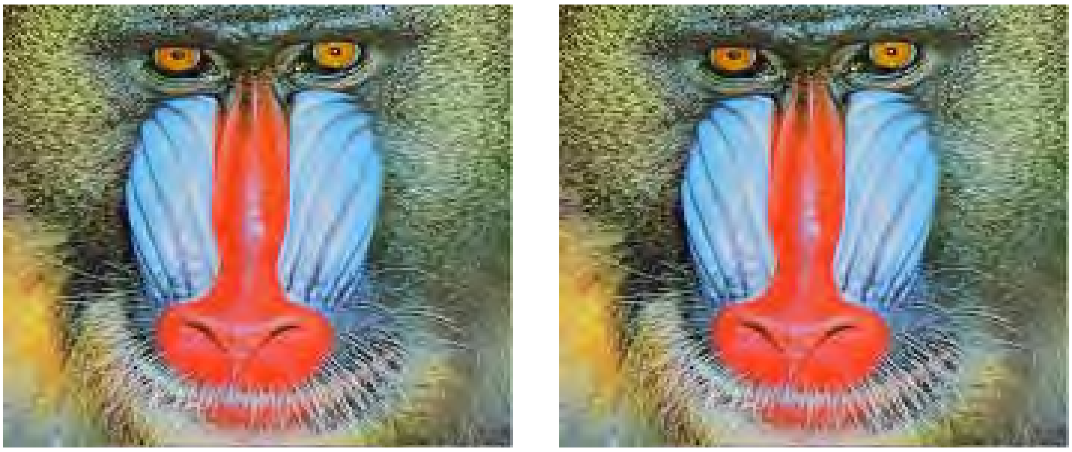
(a) Barbara (b) Barbara

Fig 2 Barbara (a) Cover image and (b) Stego image



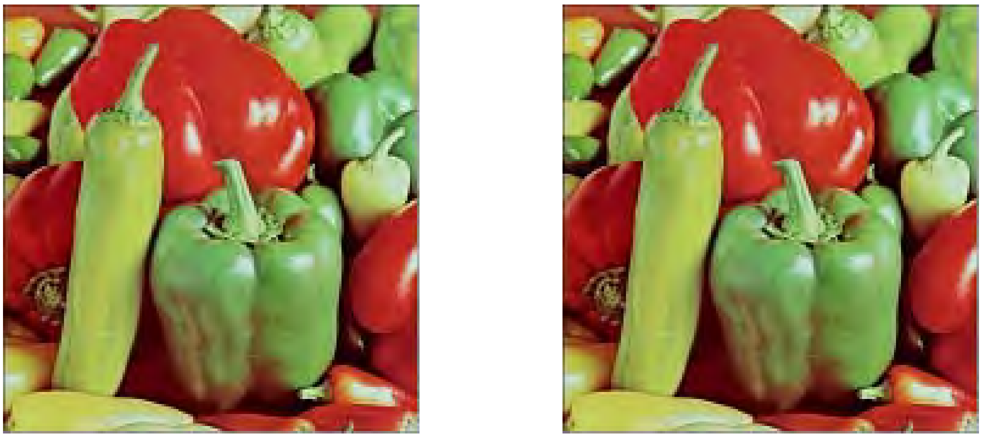
(a) Lena (b) Lena

Fig 3 Lena (a) Cover image and (b) Stego image



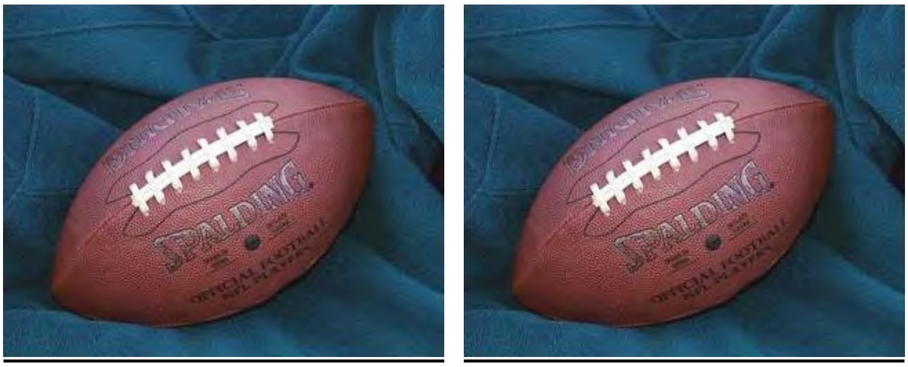
(a) Baboon (b)Baboon

Fig 4 Baboon (a) Cover image and (b) Stego image



(a) Pepper (b) Pepper

Fig 5 Pepper (a) Cover image and (b) Stego image



(a) football (b) football

Fig 6 football (a) Cover image and (b) Stego image

5. Conclusion and Future Scope

In this work we explored the existing image steganography techniques. We proposed an efficient image steganography technique. In image steganography, image is used as a carrier for transmission of the secret information or data. The image used can be either gray scale or color image. In this technique data is firstly preprocess. This preprocessing reduces the size of the data by a significantly great amount. This preprocessed data is then embedded into the LSBs of the pixels of the image depending upon the intensity of the pixel values. Our proposed algorithm is targeted to achieve very high image embedding capacity into the cover image and more security of the secret data.

The proposed technique performs better than MKA [6]. It has high PSNR value and low MSE value as compared to MKA. This preprocessing reduces the size of the secret data by a significant amount and thus permits more data into the same image. The embedding capacity of the proposed technique is very high as compared to MKA. This method has good imperceptibility, sufficient payload and has high security. Data security and high embedding capacity is there due to the pre-processing of the data before embedding into the cover image. This method does not require the original image while extracting the secret data from stego image.

The work can be extended to provide an alternative strategy for data compression which can further reduce the size of the data. Efficient cryptographic techniques can also be used along with the steganographic techniques to provide more security to the data.

6. References

[1] Ahlfeldt; R.M., (2006) “Information Security in a Distributed Healthcare Domain”. Ph.D. thesis, University of Sk¨ovde, Department of Communication and Information.

[2] Chandramouli R. and Memon N. (2001), "Analysis of LSB based Image Steganography Techniques", Proceedings of ICIP 2001, Thessaloniki, Greece, October 7−10.

[3] Deshpande N, Snehal K., ”Implementation of LSB Steganography and Its Evaluation for Various Bits” K.K.Wagh Institute of Engineering Education & Research, Nashik India

[4] http://www.appliedtrust.com/resources/security/every-company-needs-tohave-a-security -program

[5] Johnson, N.F. & Jajodia, S. (1998), “Exploring Steganography: Seeing the Unseen”, *Computer Journal*.

[6] Kekre H.B, Athawale A., Halarnkar P.N, (2009) “Performance Evaluation of Pixel Value Differencing and Kekre’s Modified Algorithm for Information Hiding in Images”, International Conference on Advances in Computing, Communication and Control, pp 342-346

[7] K Suresh Babu etal. (2005) “Authentication of secret information in image steganography”, *Computer Journal*.

[8] Marvel L.M (1999) “Spread Spectrum Image Steganography,” IEEE transactions on image processing, vol. 8, no. 8, pp. 1075-1083.

[9] Moerland, T., “Steganography and Steganalysis”, Leiden Institute of Advanced Computing Science, www.liacs.nl/home/ tmoerl/privtech.pdf

[10] Mazurczyk W., Smolarczyk S., Szczypiorski K.:(2009) “Hiding Information in Retransmissions”, In: Computing Research Repository (CoRR), abs/0905.0363,arXiv.org E-print Archive, Cornell University, Ithaca, NY (USA).

[11] M. Hussain, M. Hussain., (2010) “Pixel Intensity Based High Capacity Data Embedding Method”, Information and Emerging Technologies, International conference 978-1-4244-8003

[12] Morkel, T., Eloff, J.H.P & Olivier, M.S., (2005) "An overview of Image Steganography", Proceedings of Information Security South Africa (ISSA) Conference.

[13] N. Tiwari and M. Shandilya, (2010) “Secure RGB Image Steganography from Pixel Indicator to Triple Algorithm-An Incremental Growth”, International Journal of Security and Its Applications Vol. 4(4)

[14] Tanembaum, A. S., Computer Networks, 4th Edition. Prentice Hall, 2003\

[15] Wu D. C and Tsai W. H. (2003), “A steganographic method for images by pixel-value differencing", Pattern Recognition Letters, vol. 24, no. 9-10, pp. 1613-1626.

[16] Wu H.C., et al. (2005), “Image Steganographic scheme based on pixel-value differencing and LSB replacement methods”, VISP(152), No. 5

[17] Swain G, Lenka S.K (2011),”Steganography Using the Twelve Square Substitution Cipher and an Index Variable”.

[18] Swain G, Lenka S.K (2010), “A Hybrid Approach to Steganography Embedding at Darkest and Brightest Pixels”, International Conference on Communication and Computational Intelligence .