High Capacity Information Hiding System for Encrypted Text Message Using Pixel Intensity Based LSB Substitution Technique

Savita [1], Dr Mamta Juneja[2]

[1]ME, Computer Science and Engineering, University Institute of Engineering and Technology, Panjab University, Chandigarh

[2]Assistant Professor, Department of Computer Science and Engineering, University Institute of Engineering and Technology, Panjab University, Chandigarh

Abstract

This paper presents the spatial domain image steganography based embedding system using different pixel intensities. Here the efforts have been done to combine cryptography with the steganography for imparting better security to the proposed system. Various performance measures like peak signal to noise ratio, mean square error and the percentage of total pixels used has been calculated for checking the efficiency and capacity of the system. Also, steganalysis has been performed to check if the proposed system is intact to various statistical attacks like histogram analysis, chi-square analysis and RS analysis.

Keywords: *Information hiding, Steganography, Cryptography, Steganalysis.*

1. Introduction

In today's world digital communication has expanded a lot by the means of various communication media like internet and mobile networks. So, information hiding plays a very important role for safe guarding the data against intruders, eve droppers and hackers. For many decades, cryptography [1] has been used as the best option for information hiding. Cryptography scrambles the message in an open environment in such a way thereby making it meaningless and unreadable until and unless the decryption key is available. Now, cryptography alone is not sufficient as a number of techniques have been developed that find the loopholes in the cryptography system thereby cracking them. So Steganography along with cryptography is the best solution for information hiding. Steganography [2] is a word from Greek origin meaning covered writing. Steganography hides the existence of the message by concealing it in some carrier like image, audio or video file. Various steganography techniques [3] have been developed in both spatial domain [4] and transform domain [5]. In spatial domain, one works on the pixel intensities and in transform domain one makes use of the frequency components. Both the domains are efficient with their own pros and cons. In this paper, spatial domain has been used. One of the most efficient techniques in the spatial domain is least significant bit (LSB) embedding [6-8] as changes make in the least bits make a very small changes in the overall color of the pixels. However, steganalysis [9] researchers gradually found that some characteristics of the carrier/cover used for concealing secret message are changed even if we embed a small message to the LSB of the cover. Westfeld and Pfitzmann [10] designed a technique to successfully identify sequential LSB embedding method, which based on pair-of-value distributions called chi-square analysis. Fridrich et.al [11] proposed the RS analysis method to detect LSB steganography in gray (8-bit) and color (24-bit) images. Other techniques available in the spatial domain include masking and filtering [12], multi bit plane image steganography (MBPIS) [13] and multiple-based notational system (MBNS) [14].

In this paper, a modified version of the scheme proposed by Kekre et.al. [15] based on variable least significant bits data embedding along with the cryptography applied on the text message has been presented. The work proceeds by encrypting the secret data using data encryption standard (DES) algorithm [16] rather than just applying the 8 bit secret key XOR method. Then the encrypted text message is embedded base on different pixel intensities. The proposed system successfully resists the stastical attacks like histogram analysis [17], chi-square analysis and RS analysis. The remaining parts of this paper are organized as follows. Section 2 summarizes the existing steganography techniques like the least significant bit (LSB) embedding and pixel-value differencing (PVD) [18], and statistical steganalysis techniques like RS steganalysis and histogram analysis. In Section 3, we present our method for escaping the detection of the steganalysis methods. Our experimental results and comparison anlysis are in Section 4. Finally, the conclusions and discussion about future work of our research are in the last section.

2. Steganography and Steganalysis techniques

2.1 Steganography techniques in spatial domain

In [6-8] are LSB based techniques, which replaces the least significant bits of cover image with message bits. LSB substitution scheme is the simplest one to hide message in a cover image. But the major problem it has is of sequential substitution, hence eavesdroppers easily detect the presence of message inside the image .To overcome this problem, random LSB substitution technique was introduced [19]. The multi bit plane image steganography (MBPIS) was proposed by Nguyen, Yoon and Lee which is an extension of the simple LSB substitution to the multiple bit planes. Zhang and Wang also presented an adaptive steganographic scheme with the multiple-based notational system (MBNS) based on human visual system (HVS) which converts secret data into symbols by representing variable bases in a notational system. To achieve high imperceptibility aspect [20] is proposed to embed the information into the edges of the cover- object. For achieving high capacity in LSB domain pixel indicator technique [21] was introduced that makes use of any one channel among RGB and sequentially embeds data into two least significant bits of chosen channel. An adaptive least-significant bit (LSB) steganographic method was proposed. This method includes pixel value differencing (PVD) which uses the difference value of two consecutive pixels to estimate the total number of secret bits that can be embedded into the two pixels. This approach helps to differentiate the smooth and edge areas. Another novel adaptive data hiding scheme proposed in [16] to utilized edge area with k-LSB method and the smooth area with PVD method.

2.2 Steganalysis Techniques [10,11,17]

No doubt different steganography techniques like LSB, transform functions such as DCT, DWT in the transform domain and the adaptive steganography have emerged as practical ways for concealing

information but at some point these are still vulnerable to attacks. A lot of work has been done on steganalyzing different steganography techniques and it has been seen that these techniques sometimes lead to compromise in the statistical properties of the image and hence results in failure of the steganographic system. There is a significant amount of work done in the area of specific detection. The focus here is on the detection of the LSB embedding techniques. Various techniques available are as follows:

1. Histogram Analysis
In statistics, a histogram is a graphical representation of the distribution of data. It is an estimate of the probability distribution of a continuous variable and was first introduced by Karl Pearson. A histogram is a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval. The total area of the histogram is equal to the number of data. A histogram may also be normalized displaying relative frequencies. It then shows the proportion of cases that fall into each of several categories, with the total area equaling 1.

2. Chi-square Analysis
Pfitzman and Westfeld introduced a powerful statistical attack based on histogram analysis of Pairs of Values (PoVs) that are swapped during message embedding process. The PoVs can be formed by pixel values, quantized DCT coefficients, or palette indices that differ in the least significant bit. The idea of this attack is to test for the statistical significance of the fact that the occurrences of both values i.e. the theoretically expected frequency distribution and some sample distribution observed in the attacked image in each pair are the same. In the original, the theoretically expected frequency is the arithmetic mean of the two frequencies in a PoV. Once the observed sample distribution and the theoretically expected frequency distribution are determined, the Chi-square test can be used to determine the degree of similarity between them.
Chi-square test works as follows:
1. Suppose that there are *k* categories and that we have a random sample of observation. Each observation must fall within only one category. For example, for a palette image there are at most 256 colors c i in the palette, which means at most 128 PoVs and *k* = 128.
2. The theoretically expected frequency in *i*th category, *i* = 1, 2, . . *k* after embedding an equally distributed message bits is defined as:
N_i= number of indices in the pair $\{C_{2i},C_{2i+1}\}/2$
3. The actual frequency of occurrence in the sample is:
N_i' = number of indices to C_{2i}
4. The Chi-square statistic is calculated as:
$$CHS_{k-1}= \sum_{i=1}^{k}(Ni - Ni')^2/2$$
With k-1 is degree of freedom.
5. *p* is the probability of the above statistic under the condition that the distributions of N_i and $N_{i'}$ are equal. If the distribution of N_i, tends to be equal to that of $N_{i'}$,

CHS_{k-1} will approach to zero, accordingly *p* value will be close to one.

3. RS Analysis

To reliably and accurately detect the hidden message that is randomly scattered through the image, Fridrich et al. introduced a powerful steganalysis method that utilizes the spatial correlation in the stego image. The basic idea is to discover and quantify the weak relationship between the LSB plane and the image itself. The principle used behind is that since the LSB flipping simulates the act of adding pixel noise, it more frequently results in an increase in the value of the discrimination function *f* rather than a decrease. Thus the total number of regular groups will be larger than that of singular groups. Let R_M and SM be the relative number of regular groups and singular groups for a non-negative mask M, respectively. The following zero-message hypothesis is true for the typical cover images, that is

$$R_m \cong R_{-m}$$

$$\text{and} \qquad S_m \cong S_{-m}$$

which mean that the value of R_M is approximately equal to that of R_{-M} if no hidden message exists. The same should be true for the relationship between S_M and S_{-M}. However, the above assumption does not hold if the LSB plane is randomized. In such case R_M and S_M have the following relationship:

$$R_m \cong S_m$$

RS analysis works as follows:

1. Given an M-by-N image whose pixel values belong to the set *P*, the image is first partitioned into groups of n adjacent pixels (X_1,....., X_n) along the rows or columns. To capture the spatial correlation, a discrimination function *f* is defined as the mean absolute value of the differences between adjacent pixels.

$$F\ (X_1,.....,X_n) = \frac{1}{n-1}\sum_{i=1}^{n} |Xi + 1 - Xi|$$

This measures the smoothness or regularity of the pixel group G= (X_1,....., X_n).

2. The LSB embedding can be described using standard flipping function Fi and dual flipping function F_{-1} as follows:

F_1: 0 ↔ 1, 2 ↔ 3… 254 ↔ 255

F_{-1}: -1 ↔ 0, 1 ↔ 2 ……... 255 ↔ 256

F_0: $F_0(x)$= x, for all x belonging to P.

3. The pixel group *G* can be classified into three different types: R, S, and U depending on how the flipping changes the value of the discrimination function:

Regular groups R ↔ $f(F(G) > f(G))$

Singular groups S ↔ $f(F(G) < f(G))$

Unchanged groups U ↔ $f(F(G) = f(G))$

Here, F (G) applies the flipping function *F* to each component of the group G=(X_1………X_n).

For a group to be regular, the pixel noise within this group should be increased after LSB flipping. Similarly, for a singular group, the pixel noise becomes decreased after the flipping operation.

For very noisy images, the difference between the number of regular and singular pixels in the cover image is small. Finally, the RS steganalysis is more accurate for messages that randomly spread over the stego image than for images with sequentially embedded message.

The RS analysis provides very reliable detection for the secret message randomly embedded in the stego images only under certain assumptions. If these assumptions do not hold, this detection will fail.

3. Proposed Work

In this section, the proposed work has been described, which is then compared with the method proposed by Kekre et.al. The modified Kekre algorithm (MKA) method made use of up to five least significant bits for embedding information. In this paper, the work proposed by Kekre has been further modified for increasing capacity security and temper resistance. The cover image that is used as carrier is a 24 bit RGB bitmap color image. And the secret data used is the text message. Before embedding the secret message file into the cover-image the process of encryption has been used to encrypt the secret data using DES algorithm for increasing the security of the system rather than just applying the 8 bit XOR operation on the secret data there by making it less vulnerable to attacks. DES is the symmetric block cipher. It has 64 bit block size and uses a secret key known only to the sender and receiver so that encryption and decryption can be performed by them only. In it there are 16 stages of processing, called as rounds. There is also an initial and final permutation, termed IP and FP that are inverses of each other. The main rounds take place after division of the block into two 32-bit halves and processed according to Feistel scheme also known as criss-cross scheme. In Feistel structure, the process of decryption and encryption are almost similar, only difference between them is that while decrypting the sub keys used are applied in the reverse order. This all simplifies the implementation to great extent as there are no separate encryption and decryption algorithms required. After this the encrypted message is embedded into the cover image using the process embedding algorithm. The embedding algorithm divide the pixel intensities of the color image into different pixel intensity ranges ranging from high intensity pixel values to low pixel intensity values. Other modification proposed in the work is that in the entire image where ever the value of any two red, green or blue values is zero the remaining one's all the pixel value bits i.e. 8 bits are utilized for embedding secret data. After embedding the stego-image so formed is sent to the receiver. The original message at receiver side is extracted from the stego-image using the decryption process. In addition to being more secure, the proposed work also successfully resists to the stastical attacks like Rs analysis, chi-square analysis and histogram analysis. Proposed work architecture and the embedding algorithm are described below:

3.1 Proposed architecture:

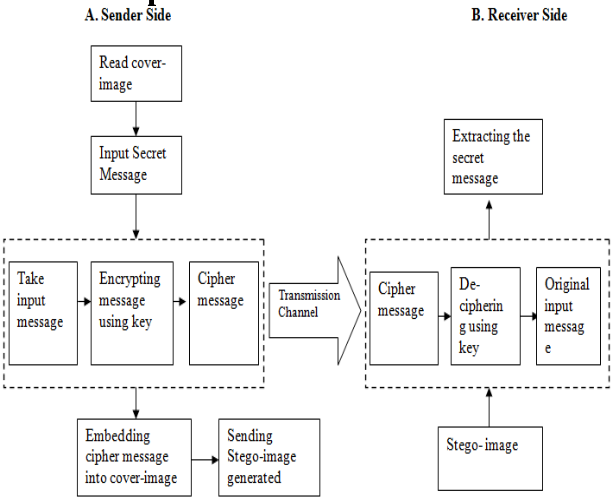


Fig. 1 Proposed architecture

3.2 Proposed Algorithm

A. Embedding module:

Step 1: Read cover image for embedding secret text message.
Step 2: Input encrypted secret message.
Step 3: Embedding secret message inside cover image using proposed embedding scheme.
Step 4: Sending generated stego-image to the receiving party.

B. Extraction module.

Step 1: Receiving stego-image.
Step 2: Entering the stego-key for de-ciphering the stego image.
Step 3: applying reverse embedding procedure for extraction of original text message.

3.3 Proposed Embedding Scheme

Every pixel value in the cover image is analyzed and then following embedding process is applied:
1. If the value of the pixel say pi, lie in the range $240 \leq pi \leq 255$, then we use the 4 LSB's of the corresponding pixel for embedding secret message bits. It means if all the first 4 Most Significant Bits (MSB's) are 1 then the remaining 4 LSB's are used for hiding secret data.
2. If the value of the pixels lies in the range $231 \leq pi \leq 239$, then we utilize the 3 LSB's of the corresponding pixel for embedding secret message bits.
3. If the value of the pixels lies in the range $224 \leq pi \leq 230$, then we uses the 5 LSB's of the corresponding pixel for embedding secret data bits. It means if all the first 3 Most Significant Bits (MSB's) are 1 then the remaining 5 LSB's are used for hiding data.
4. If the value of the pixels lies in the range $199 \leq pi \leq 223$, then we uses the 2 LSB's of the corresponding pixel for embedding secret data bits.
5. If the value of the pixels lies in the range $192 \leq pi \leq 198$, then we uses the 5 LSB's of the corresponding pixel for embedding secret data bits.
6. If the value of the pixels lies in the range $51 \leq pi \leq 191$, then we uses the least significant bit of the corresponding pixel for embedding secret message bits.
7. If the value of the pixels lies in the range $32 \leq pi \leq 50$, then we uses the 2 LSB's of the corresponding pixel for embedding secret message bits.
8. If the value of the pixel say gi, is in the range $16 \leq pi \leq 31$, then we utilize the 3 least significant bit of the pixel for embedding secret message bits.
9. If the value of the pixels lies in the range $0 \leq pi \leq 15$, then we uses the 4 LSB's of the corresponding pixel for embedding secret data bits.
10. If any one of the red, green or blue component is equal to 0 and other two components are 255 then we utilize least 4 bit values of the zero value component.
In table 1, Pixel intensity represents the pixel value and utilized bits represents the total number of bits embedded into a pixel.

Table1: Proposed Work

S.No.	***Pixel Intensity***	Utilized Bits
1	240-255	4
2	231-239	3
3	224-230	5
4	199-223	2
5	192-198	5

6	51-191	1
7	31-50	2
8	16-30	3
9	0-15	4
10	R=255, G=255, B=0	8
11	R=255, G=0, B=255	8
12	R=255, G=0, B=255	8

4. Experimental results and comparison analysis

The experimental results have been evaluated based on two criteria. First, on the basis of imperceptibility and payload capacity. Second, on the basis of resistance to stastical attacks.

4.1Evaluation based on Imperceptibility/Stego-image Quality and Payload Capacity:

Imperceptibility is the factor that is used to evaluate the stego-image quality. It results in high value when the difference between the cover image chosen and the stego-image generated are less. For evaluating stego-images based on this criteria peak signal to noise ratio (PSNR) and mean square error (MSE) values are calculated. Value for PSNR should be high and for MSE it should be low. Second thing that is considered is the payload capacity which is defined as the capacity of the image to hide details within it without any distortion to the original image. Value for payload should be as high as possible. Four colored bitmap cover images lena, baboon and Mahalaxmi are used, each of size 512x 512 in our experiment as shown in fig. 2. 24 bit RGB color image is used as cover image. Text message is Abraham Lincoln's letter to his son's teacher that is to be hidden into the cover image. They were compared to work done by Kekre table 3 and the results obtained are shown in Table 2 .

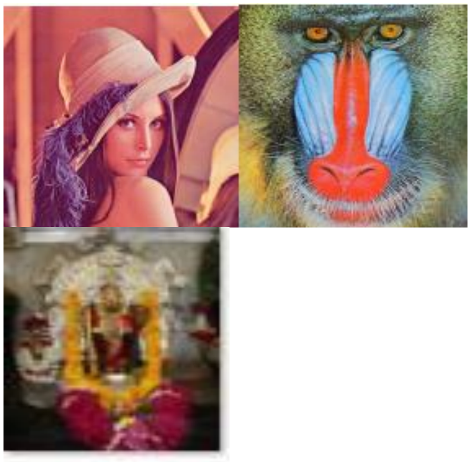


Fig. 2 Cover-images (Lena, Baboon, Mahalaxmi)

Table 2: Value of MSE, PSNR and percentage of pixels used in image to store data

Image	*%age of used pixels*	*MSE*	*PSNR*
Lena	30.0464	0.0038	43.13
Baboon	47.9922	0.0041	44.47
Mahalaxmi	17.7162	0.0093	45.05

Table 3: Comparison value of MSE, PSNR and percentage of pixels used in image to store data

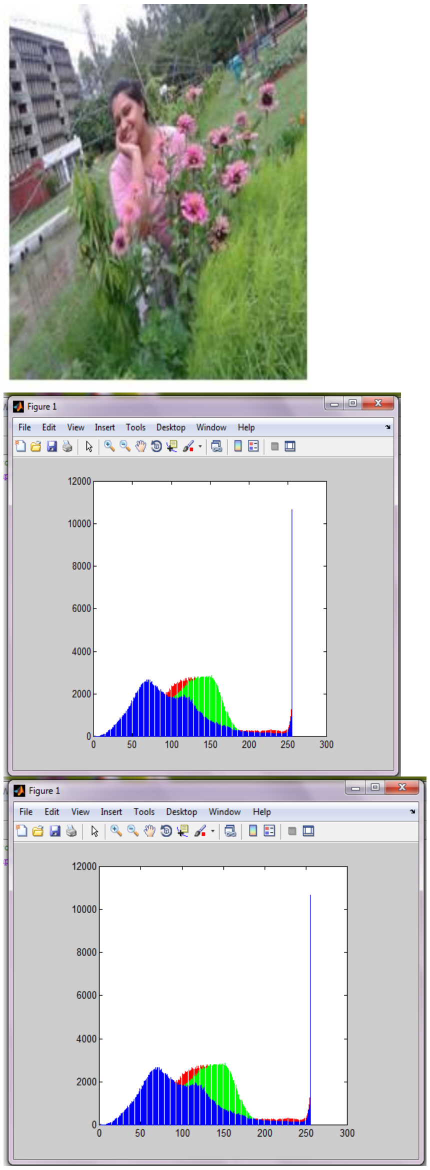
Image	*MSE*	*PSNR*
Lena	0.0038	43.13
Baboon	0.0041	44.47
Mahalaxmi	0.0093	45.05
Kekre's Algorithm using text message		
Lena	4.77	41.34
Baboon	4.18	42.39
Mahalaxmi	3.74	41.91

4.2 Evaluation based on Resistance to Stastical Attacks:

Stastical attacks are the application of steganalysis which focuses on finding the details hidden in the stego-image. Different types of stastical attacks are there like visual analysis, histogram analysis, chi-square analysis and RS analysis etc. The proposed system is also checked for resistance to the histogram analysis, chi-square analysis and RS analysis. Results on “Mypic” for histogram analysis, chi-square analysis and Rs analysis have been shown in fig.1, fig.2 and fig.3 respectively.

1. Histogram Analysis

The results of Histogram analysis technique proposed in [3] are shown in Figure 3. As there are no differences found in histograms of original and stego image so could not be attacked.



(a) Cover image (b) Cover image histogram
(c) Stego-image histogram
Fig. 3(a, b, c) Histogram attack on mypic.bmp

2. Chi-Square Analysis

The results of chi-square analysis technique proposed in [18] are shown in Figure 4.And, the proposed system successfully sustain this attack.

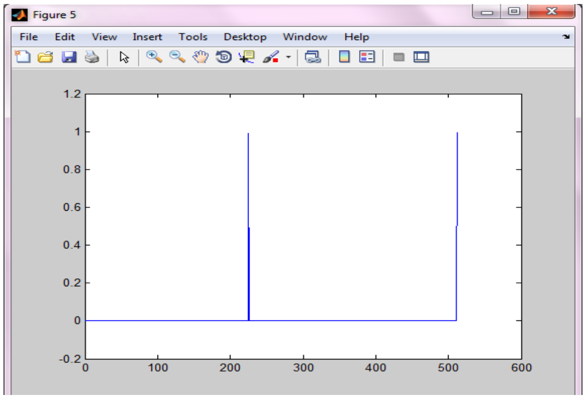


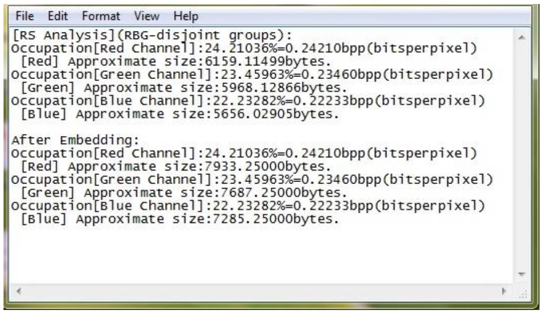
Fig. 4 Chi-square attack on mypic.bmp

3. RS Analysis

The results of RS analysis technique proposed in [19] are shown in Figure 5.And, the proposed system successfully sustain this attack.

Table 4: RS ana;lysis for mypic.bmp

Mypic.bmp	***Initial Value***	***After embedding***
R_m-R_{-m}	0.02380	0.00119
S_m-S_{-m}	0.01960	0.00258



```
[RS Analysis](RBG-disjoint groups):
Occupation[Red Channel]:24.21036%=0.24210bpp(bitsperpixel)
 [Red] Approximate size:6159.11499bytes.
Occupation[Green Channel]:23.45963%=0.23460bpp(bitsperpixel)
 [Green] Approximate size:5968.12866bytes.
Occupation[Blue Channel]:22.23282%=0.22233bpp(bitsperpixel)
 [Blue] Approximate size:5656.02905bytes.

After Embedding:
Occupation[Red Channel]:24.21036%=0.24210bpp(bitsperpixel)
 [Red] Approximate size:7933.25000bytes.
Occupation[Green Channel]:23.45963%=0.23460bpp(bitsperpixel)
 [Green] Approximate size:7687.25000bytes.
Occupation[Blue Channel]:22.23282%=0.22233bpp(bitsperpixel)
 [Blue] Approximate size:7285.25000bytes.
```

0.09
0.07
0.05
0.03
0.01

Fig. 5 RS Analysis attack on mypic.bmp

5. Conclusions

The main advantage of our proposed work is that if we use more bytes of cover-image for hiding the secret data bits, then our proposed work results in better peak signal to noise ratio values and improved security as compare to work proposed by kekre. In our proposed work, variable pixel intensities have been used to store data in variable bits depending upon that intensity value and it has resulted in good statistical result for peak signal to noise ratio (PSNR), mean square error (MSE) values and for the percentage of pixels utilized. The proposed steganographic system also makes effective use of the lower intensity pixels to increase the capacity of data embedding. The additional attractive features of this steganographic system are that it is less vulnerable to external attacks like attacks by intruders as cryptography has been applied to deal with such issues and is also less prone to other statistical attacks like histogram analysis, chi-square analysis and RS analysis. And also this proposed system utilizes those intensity area pixels of the colored component where pixel intensity of any of the color component red, green or blue is zero. The results obtained after implementing the proposed system show the efficiency of the proposed system.

6. Future Scope

There are various ways of further improving our proposed work. The security level may be increased by using more reliable and more secure method for encrypting the data. Also the approach of compressing data before encryption can also be employed. As far as extending this research goes, pixel intensity ranges can be refined further and methods like PRNG i.e. pseudo random number generation can be combined with the basic LSB substitution technique for improving the efficiency of the work.

REFERENCES

[1]. John L.Manferdeli and David A.Wagner, “Cryptography and cryptanalysis”, Springer, 2013.

[2]. Provos et.al. , “Hide and seek: an introduction to steganography”, IEEE Security & Privacy Magazine, Volume 1, 2003, pp. 32-44.

[3]. Nagham Hamid et.al, “Image steganography techniques: an overview”, International Journal of Computer Science and Security (IJCSS), Volume 6, Issue 3, 2012.

[4]. P.Kruus et.al, “A survey of steganography techniques for image files”, Advanced Security Research Journal, Volume 5, Issue 1, pp.41-52, 2003.

[5]. N. F. Johnson et.al, “A survey of steganographic techniques” in Information Hiding Techniques for Steganography and Digital Watermarking, Ed. London: Artech House, pp.43-78, 2000.

[6]. G.C. Kessler, "An Overview of Steganography for the Computer Forensics Examiner", Forensic Science Communications, Volume 6, Issue 3, July 2004.

[7]. D. Artz, "Digital Steganography: Hiding Data within Data", IEEE Internet Computing: Spotlight, June 2001, pp 75-80.

[8]. K. Bailey and K. Curran, "An Evaluation of Image Based Steganography Methods", Multimedia Tools & Applications, Volume 30, Issue 1, July 2006, pp 55-88.

[9]. Neil F. Johnson and Sushil Jajodia, “Steganalysis: The investigation of hidden information”, in proc. of IEEE conference, pp. 113-116, September 1998.

[10]. A. Westfeld and A. Pfitzmann , "Attacks on steganographic systems", in Lecture Notes in Computer Science, Springer, Berlin, pp.61-75, 2000.

[11]. Fridrich and M.Goljan, "Practical steganalysis of digital images — state of the art", in proc. of SPIE, Security and Watermarking of Multimedia Contents IV, E.J. Delp III and P.W. Wong , pp.1-13, 2002.

[12]. Johnson et.al, “Exploring steganography: Seeing the unseen”, IEEE Computer Journal, Volume 31, Issue 2, pp. 26–34, 1998.

[13]. B.C.Nguyen et.al, “Multi bit plane image steganography”, in proc. of 5th International Workshop (IWDW), Springer, Volume 4283, pp.61–70, 2006.

[14]. Xinpeng Zhang and Shuozhong Wang, “Steganography using multiple-base notational system and human vision sensitivity”, IEEE Signal Processing Letters, Volume 12, Issue 1, pp.67-70, 2005.

[15]. H. B. Kekre et.al, “Performance evaluation of pixel value differencing and Kekre’s modified algorithm for information hiding in images”, in proc. of International Conference on Advances in Computing, Communication and Control, 2009, pp342-346.

[16]. DES – Wikipedia, http://en.wikipidea.org/wiki/DES.

[17]. N. Provos and P. Honeyman, “Detecting steganographic content on the Internet”, in proc. of Network and Distributed System Security Symposium (NDSS), pp.1-13, 2002.

[18]. D. C. Wu and W. H. Tsai, “A steganographic method for images by pixel-value differencing,” Pattern Recognition Letters, Volume 24, Issue 9-10, June 2003, pp. 1613-1626.

[19]. S.Venkatraman et.al. , "Significance of Steganography on Data Security", in proc. of International Conference on Information Technology: Coding and Computing (ITCC), April 2004.

[20]. Kathryn Hempstalk, "Hiding Behind Corners: Using Edges in Images for Better Steganography", in proc. of the Computing Women's Congress, February 2006, pp 11- 19.

[21]. AdnanGutub et.al. , “Pixel indicator high capacity technique for RGB image based Steganography”, in proc.

of IEEE 5th International Workshop on Signal Processing and its Applications (WoSPA), March 2005.

Savita student in University Institute of Engineering and Technology , Panjab University, Chandigarh. She did Bachelors in Computer science from Punjab Technical University, India in 2011and currently pursuing masters in the same. Her interest areas are Image Processing, Steganography, information hiding and Information Security.

Mamta Juneja Assistant Professor in University Institute of Engineering and Technology, Panjab University, Chandigarh, India. She did masters and doctorate in Computer Science from Punjab Technical University, India. Her interest areas include Image Processing, Steganography, Information Hiding and Information Security.