High Capacity and Statistical Attacks Resistant Reversible Image Data Hiding Scheme using LSB Substitution

Savita
ME, Computer Science and Engineering
University Institute of Engineering and Technology,
Panjab University,
Chandigarh

Mamta Juneja, Ph. D
Assistant professor, Computer Science and Engineering
University Institute of Engineering and Technology,
Panjab University,
Chandigarh

ABSTRACT
Reversible data hiding technique is advancing with each passing day. It is a form of steganography in which the hidden data can be recovered or extracted by the authorized user. In this paper, we have proposed the variable least significant bit based digital image hiding system based on LSB substitution. The proposed scheme has high security, better embedding capacity and increased imperceptibility. Also, the proposed scheme successfully resists the stastical attacks like Rs analysis, histogram analysis and chi-square analysis which adds to the efficiency of the proposed scheme.

General Terms
Steganography, Cryptography, Steganalysis.

Keywords
Reversible data hiding, steganography, cryptography, variable least significant bit embedding, steganalysis.

1. INTRODUCTION
Reversible data hiding [1] means lossless data hiding. For many years it has captured the attraction of researchers. Digital watermarking and Steganography [2] techniques are used for data hiding in a reversible manner. Many techniques have been proposed to hide data in an image [3] like steganography and cryptography. Steganography comes from the Greek and literally means writing in secret. It proceeds by making use of a cover image to hide data rather than attempting to alter the structure of the message to be hidden. Cryptography [4] contrast to steganography, changes the structure of a message to make it meaningless and unreadable until and unless the decryption key is available. Steganography has developed a lot in recent years because of advancements in the digital techniques being used to hide data. Generally, main requirements of any image steganographic system are high capacity, high imperceptibility, robustness and temper resistance. High capacity means the ability to hide large amount of information into the system. Robustness is that the system must remains intact even if modifications like cropping, compression, filtering, noise addition etc are applied on it. High imperceptibility is maintaining the high quality of the stego-image that is the changes done to the cover image due to embedding should not be visible in the stego-image. Since the difference between the embedded image and original image is almost imperceptible from human eyes, reversible data embedding could be thought as a covert communication channel. Temper resistance means that it should be difficult to modify and detect the message with various attacks like stastical and visual attacks once it has been embedded. A lot of research have been made to aid in discovering the steganalysis techniques for the perception of hidden data and hence have lead to finding better and better methods for hiding data. Various steganalysis techniques are also available to detect steganography [5]. This paper presents a modified version of the scheme proposed by Kekre [6] and Hussain [7] based on variable least significant bits data embedding along with the cryptography applied on the text message rather than just applying the 8 bit secret key XOR method. Other features of the proposed work are that it successfully resists the stastical attacks like histogram analysis, chi-square analysis and RS analysis.

2. EXISTING WORK
In [8], [9] and [10] 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 [11].To overcome this problem, random LSB substitution technique was introduced [11]. The multi bit plane image steganography (MBPIS) was proposed by Nguyen, Yoon and Lee [12] which is an extension of the simple LSB substitution to the multiple bit planes. Zhang and Wang [13] 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 [14] is proposed to embed the information into the edges of the cover- object. For achieving high capacity in LSB domain pixel indicator technique [15] 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) [16] 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 [17] to utilized edge area with k-LSB method and the smooth area with PVD method.

2.1 Modified Kekre's Algorithm
Modified Kekre's Algorithm (MKA) is based on Least Significant Bit (LSB) method. MKA is applied on 24 bit RGB color image. It works according to the MSB's and utilizes up to five LSB's of a pixel to embed the data by taking into account the intensity of various pixels. For achieving security

MKA makes use of 8 bit secret key to perform XOR operation on all the bytes of message which is to be hided. The same key is used to recover the message at receiver side. In MKA the algorithm used for the embedding process keeps a matrix of pixels where 5 bits of message are used to embed, and this matrix is also required during the extraction phase.

Table1: Modified Kekre's Algorithm Scheme

S.NO.	Pixel Intensity	Data Bits to Embed	Matrix Entry	Utilize Bits
1.	240-255	1	1	5
2.	240-255	0	-	4
3.	224-239	0	1	5
4.	224-239	1	-	3
5.	192-223	x	-	2
6.	0-191	x	-	1

In Table1, x represents don't care bit (whatever bit value 0 or 1), pixel intensity represents the pixel value, data bit to embed represents the number of message bits to be embedded into the cover-image, matrix entry maintains a matrix which denotes the 5 LSB are embedded and utilize bits represents the total number of bits embedded into a pixel.

2.2 Method Proposed by Hussain

Mehdi Hussain did two improvements to the MKA. First, utilizing the lower intensity pixel for hiding data along with the least significant bits. Secondly, it makes maximum utilization of matrix which keeps the track of pixel where 5 LSBs are used for data embedding i.e. it also maintain a matrix for those pixels which will embed 5, 3 and 2 LSBs of data. Like MKA, it also uses 8 bit secret key to perform XOR operation on all the bytes of message which is to be hided and the same key is used to recover the message at receiver side.

Table2: Hussain's proposed work

S.NO.	Pixel Intensity	Data Bits to Embed	Matrix Entry	Utilize Bit/Bits
1.	240-255	1	1	5
2.	240-255	0	-	4
3.	224-239	0	1	5
4.	224-239	1	-	3
5.	192-223	0	1	3
6.	192-223	1	-	2
7.	32-191	0	1	2
8.	32-191	1	-	1
9.	16-31	0	1	3
10.	16-31	1	-	2
11.	0-15	0	1	5
12.	0-15	1	-	4

In Table2, pixel intensity represents the pixel value, data bit to embed represents the number of message bits to be embedded into the cover-image, matrix entry maintains a matrix which denotes the 5 LSB are embedded and utilize bits represents the total number of bits embedded into a pixel.

The main advantage of this algorithm over MKA is that it has very high capacity of data hiding as compare to MKA [16]. It also utilizes the lower intensity based pixel for high capacity data embedding.

3. PROPOSED WORK

In this section, we describe the proposed method, which is then compared with the method proposed by Hussain. The MKA method made use of up to five least significant bits for embedding information and then Hussain modified it by utilizing the lower intensity pixels for embedding information depending upon the intensities of pixels. In this paper the work proposed by Hussain has been further modified for increasing capacity security and temper resistance. In proposed work, the carrier cover image is a 24 bit RGB bitmap color image. And the secret data used is the text message. Also a matrix is maintained for those pixels that utilize 8, 5, 3 and 2 bits to store data. Before embedding the secret message file into the cover-image the process of encryption has been used to encrypt the secret data using data encryption standards (DES) [13] 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 Work Architecture

The proposed work architecture depicts the methodology that is followed in designing the proposed steganographic system. It gives general idea of the work methodology and the steps involved. The cover image is read first and then the secret message is taken as input from the user. The text message that has been used is 8 characters long and the key length is seven characters long. Then this text is encrypted using DES algorithm and then the embedding algorithm is applied to embed text file used as secret message into cover-image. After that the stego-image generated is sent to receiver where the cipher text is decrypted and the original secret text is extracted using the key by the receiver.

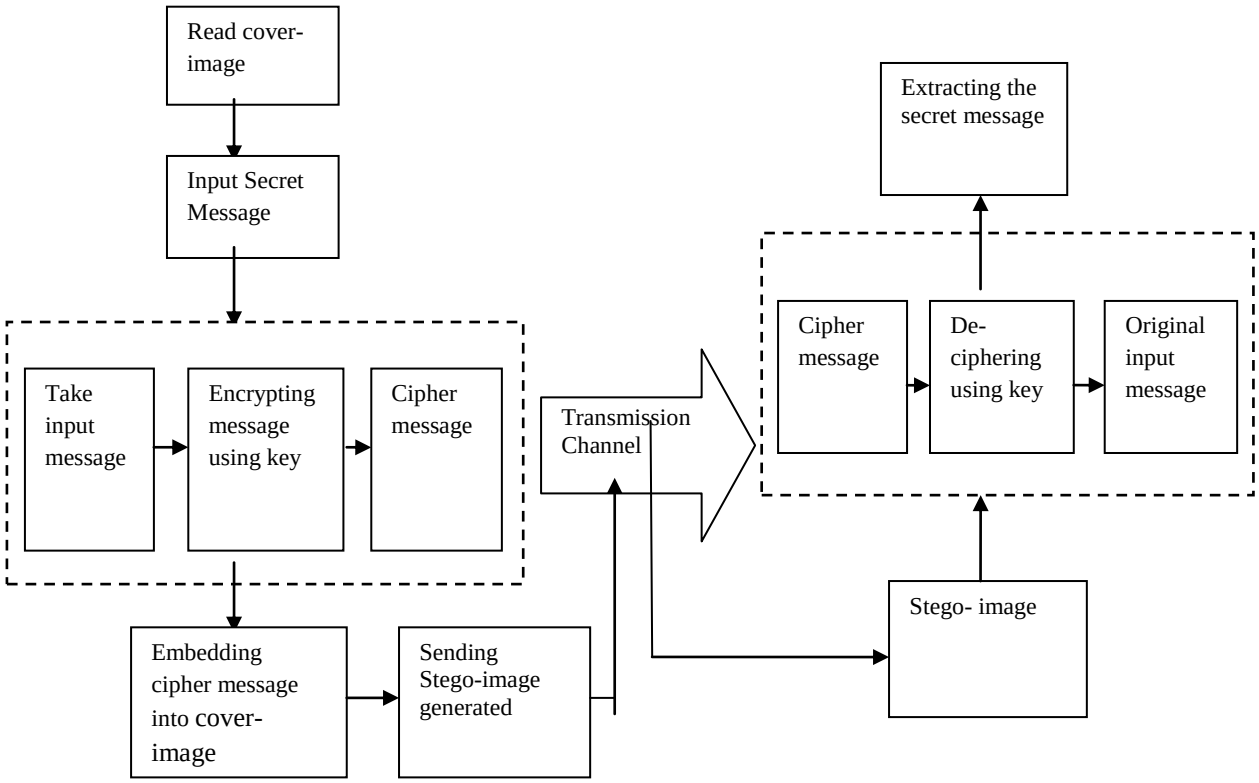


Fig 1: Proposed work architecture

3.2 Proposed algorithm

3.2.1 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.

3.2.2 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 ≤pi ≤ 255, if bit to be embedded is 0 then we use the 4 LSB's of the corresponding pixel for embedding secret message otherwise utilize 5 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≤pi ≤239, if bit to be embedded is 0 then we utilize the 3 LSB's of the corresponding pixel for embedding secret message bits otherwise utilize 4 lsb's.

3. If the value of the pixels lies in the range 224 ≤pi ≤ 230,whatever is the bit to be embedded we utilize the 5 LSB's of the corresponding pixel for embedding secret data bits.

4. If the value of the pixels lies in the range 199 ≤pi ≤ 223, if bit to be embedded is 0 then we utilize the 3 LSB's of the corresponding pixel for embedding secret data bits otherwise utilize 2 lsb's.

5. If the value of the pixels lies in the range 192 ≤pi ≤ 198, whatever is the bit to be embedded we utilize 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 ≤pi ≤ 191, if bit to be embedded is 0 then we uses the 2 LSB's of the corresponding pixel for embedding secret message bits otherwise use only least significant bit for embedding.

7. If the value of the pixels lies in the range 32 ≤pi ≤ 50, whatever the bit to be embedded is we uses the 2 LSB's of the corresponding pixel for embedding secret message bits.

8. If the value of the pixel lies in the range 16 ≤pi ≤ 31, if bit to be embedded is 0 then we utilize the 3 least significant bit of the pixel for embedding secret message bits otherwise 2 lsb's are utilized.

9. If the value of the pixels lies in the range 0 ≤pi ≤ 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 3, Pixel intensity represents the pixel value and utilized bits represents the total number of bits embedded into a pixel and x represents don't care bits (either 0 o 1).

S.NO.	Pixel Intensity	Data bit to embed	Matrix Entry	Utilized Bit/ Bits
1	240-255	0	-	4
2	240-255	1	1	5
3	231-239	0	1	3
4	231-239	1	-	4
5	224-230	x	1	5
6	199-223	0	1	3
7	199-223	1	1	2
8	192-198	x	1	5
9	51-191	0	1	2
10	51-191	1	-	1
11	32-50	x	-	1
12	16-31	0	1	3
13	16-31	1	1	2
14	0-15	x	-	4
15	R=255, G=255, B=0	x	1	8
16	R=255, G=0, B=255	x	1	8
17	R=0, G=255, B=255	x	1	8

4 EXPERIENTAL RESULTS AND COMPARISON ANALYSIS

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

4.1 Evaluation based on 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, pepper and mypic are used, each of size 512x 512 in the experiment. 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 Hussain and the results obtained are shown in Table 4.

Fig 2: Cover-images (lena, baboon, pepper and mypic)

Table 4: Value of MSE, PSNR, percentage of pixels and changed bytes percentage.

Proposed method using Abraham Lincoln's letter					
Cover Image	Embedded Data Bytes	% of used Pixel in Image	% of Changed Bytes	MSE	PSNR
Lena	1785	30.0464	58.3755	0.0038	43.13
Baboon	1785	47.9922	54.6375	0.0041	44.47
Pepper	1785	17.7162	53.1245	0.0093	45.05
Mypic	1785	31.5437	55.4325	0.0155	48.23
Hussain's method using Abraham Lincoln's letter					
Lena	1785	0.9666	55.5906	0.0174	65.7180
Bab0on	1785	1.1237	51.3975	0.0070	69.6609
Pepper	1785	0.8867	59.9455	0.0723	59.5402

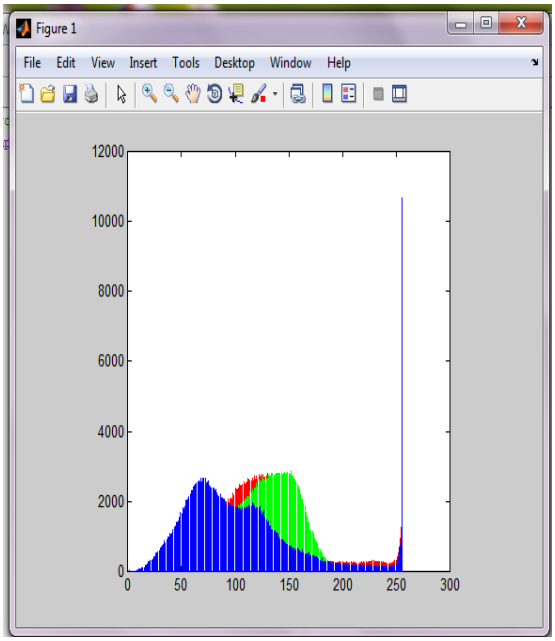
4.2 Evaluation based on Resistance to Stastical Attacks:

Statistical 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.3, fig.4 and fig.5 respectively.

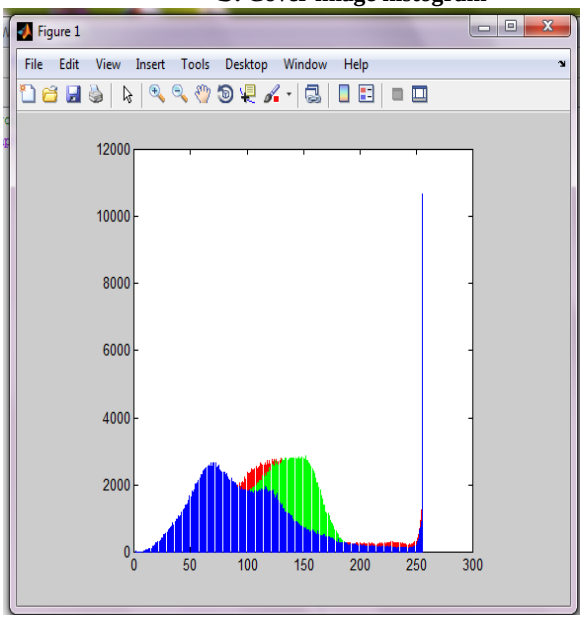
4.2.1 Histogram Analysis

The results of Histogram analysis technique [18] are shown in Figure 3. As there are no differences found in histograms of original and stego image so could not be attacked. So it is clear from the results that the proposed system successfully resists the histogram analysis as the histogram for both the cover image and stego-image comes out to be same.

a.Mypic.bmp



b. Cover image histogram



c. Stego image histogram

Figure 3(a, b, c): Histogram attack on mypic.bmp

4.2.2 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.

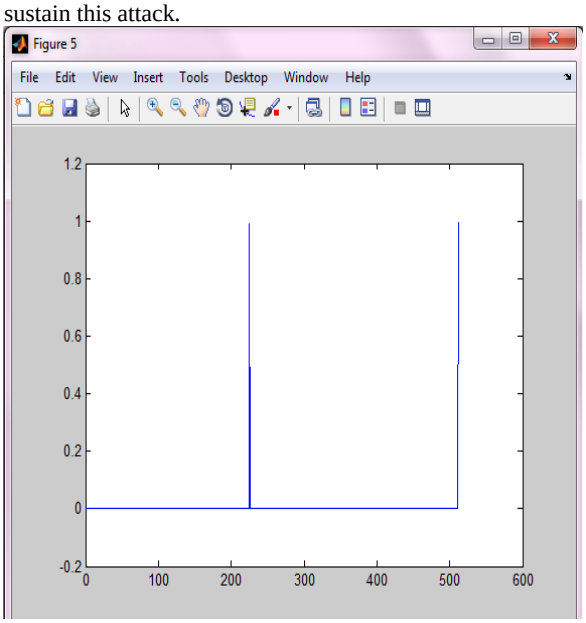


Fig 4: Chi-square attack on mypic.bmp

As the graph obtained for chi-square analysis fulfills the required range and expected results so the proposed system successfully resists this attack. The graph comes out to be same for both cover image and stego image.

4.2.3 RS Analysis

The results of RS analysis technique proposed in [19] are shown in Figure 5 (a, b) and table 5. And, the proposed system successfully sustains this attack. Table 5 shows the values of RS analysis results for before embedding and after embedding. Fig 5a shows the results for occupation of each channel for dis-joint groups and after embedding and fig 5b shows the graph obtained according to table 5 values.

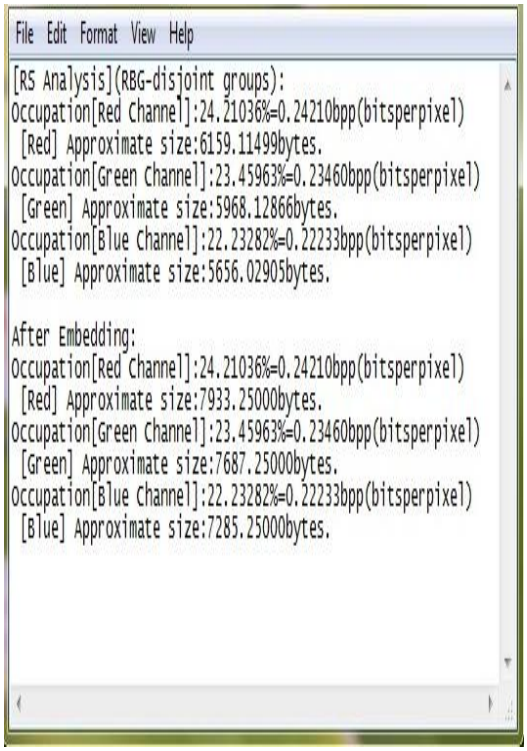


Fig 5a: Approximate per channel occupation

Table 5: Rs analysis initial and after embedding values for mypic.bmp

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

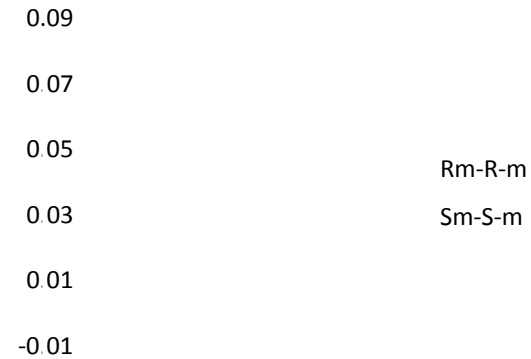


Fig 5 b : Graph for mypic.bmp

5. CONCLUSION

The main advantage of the proposed work is that has better results for peak signal to noise ratio (PSNR), mean square error (MSE) values and for the percentage of pixels utilized. Also it is more secure as better encryption algorithm has been used and is also less prone to other statistical attacks like histogram analysis, chi-square analysis and RS analysis. Effective use of the lower intensity pixels has been made to increase the capacity of data embedding so that more and more data can be stored. Also the 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. REFERENCES

[1] J. Fridrich, M. Goljan, and R. Du, "Lossless data embedding—new paradigm in digital watermarking," EURASIP J. Appl. Signal Processing, vol. 2002, no. 2, pp. 185–196, Feb. 2002.

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

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

[6] 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.

[7] Mehdi Hussain et.al, "Pixel intensity based high capacity data embedding method", IEEE, 2010.

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

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

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

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

[12] B.C. Nguyen et.al , "Multi bit plane image steganography", in Proc. of International Workshop on Digital Watermarking (IWDW) Springer, Volume 4283, November 2006, pp 61–70.

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

[14] 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.

[15] 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.

[16] 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.

[17] Cheng-Hsing Yang et.al. , "Adaptive Data Hiding in Edge Areas of Images with Spatial LSB Domain Systems", in proc. of IEEE Transactions on Information Forensics and Security, Volume 3, Issue 3, September 2008, pp 488-497.

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

[19] 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 , 2002 , pp.1-13.