Introducing Image Steganography in Bangla Language Communication

Sams Jarin
Lecturer
Department of CSE
World University Bangladesh, Dhaka, Bangladesh

Shah Mohazzem Hossain
Captain, Bangladesh Army
Instructor, Department of EECE
Military Institute of Science and Technology, Dhaka, Bangladesh

Md. Rabiul Islam, Ph.D.
Associate Professor
Department of CSE
Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh

ABSTRACT
Steganography is a modern technique of science by which any message can be hidden in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message like a form of security through obscurity. This technique has advantage over cryptography alone, is that messages do not attract attention to themselves where cryptography protects the contents of a message, steganography can be said to protect both messages and communicating parties. Now-a-days wide ranges of algorithms have been used using text, audio, video and images files as covering media. This paper represents a new approach for steganography in bangla texts considering the typical structure of bangla alphabets and also introduces a much easier process by which secret message is embedded in a 24 bit color image by using widely used 4LSB method. This paper shows a new path to provide an imperceptible stego-image for human natural vision by enhancing the embedding capacity of image steganography.

Keywords
Steganography, 4LSB, Image steganography, Steganolysis, SHA-1, Unicode.

1. INTRODUCTION
In these modern years of advancement, considerable progress has been made in the area of steganography system, a related way of networking security system by hiding the information in medium for increasing security level. The word steganography is derived from the Greek words "stegos" meaning "cover" and "grafia" meaning "writing" defining it as "covered writing" [1]. This steganography deals with hiding of data in images and steganalysis reveals the presence of data and also the hidden data in a secure way [2].This technique relies on a message being encoded and hidden in a transport layer in such a way as to make the existence of the message unknown to an observer. Most importantly, the transport layer and the carrier file are not secret and can therefore be viewed by observers from whom the secret message itself should be concealed. The first recorded uses of steganography can be traced back to 440 BC when Herodotus mentions two examples of steganography in The Histories of Herodotus [3]. When it was also reported that, the Nazis invented several steganographic methods during World War II such as Microdots, and have reused invisible ink and null ciphers. In 1945, Morse code was concealed in a drawing. Modern steganography entered the world in 1985 with the advent of the personal computer being applied to classical steganography problems [3]. Since then many steganographic techniques have been developed and implemented in the world of network security. In this paper one such new approach in image steganography has been introduced for bangle texts. This system has mainly three parts; the first one is about image taking and converting the image pixel values into the gray scale values. The second one is to encode and decode the information written in Bangla character into binary format from the corresponding Unicode, hexadecimal value and the third one is to hidden the information message into the taken preprocessed of the first stage road.

2. PROPOSED SYSTEM MODEL
2.1 Steganography
Steganography means hiding the data within a medium such that the very existence of data is hidden. It varies from encryption in the sense that its objective is secret communication, rather than data protection. Main goal of steganography is to avoid drawing suspicion to the transmission of a hidden message. The plain medium used for hiding information is called cover medium, and the medium generated after embedding the secret text is called stego medium. The Process of Steganography as shown in Fig. 1 can be explained by Simmons' "Prisoners' Problem" where if message is first inspected by intended recipient (Eve), who has access to their communication link. If Alice & Bob try to communicate through encrypted text, Eve will sense a conspiracy and block the message. This is the case where steganography come in. Alice and Bob can use steganography to hide the message such that Eve is unaware of existence of message [3].

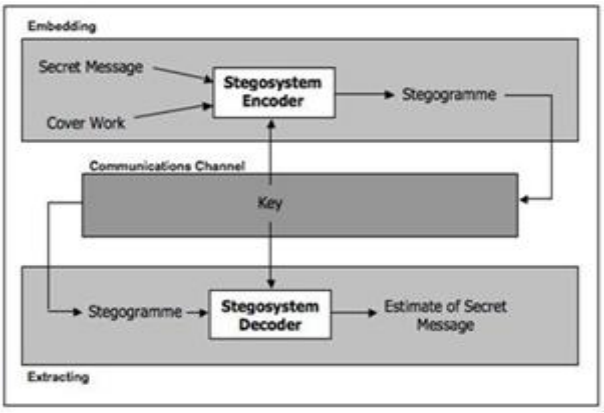


Fig. 1. Block diagram of the process of steganography

2.2 Process Description
Now-a-days there are many techniques for steganograpic purpose. In this paper some the popular digital image encoding system is discussed like least significant bit (LSB) encoding and masking and filtering techniques etc.

LSB: Least significant bit (LSB) encoding is by far the most popular of the coding techniques used for digital images. By using the LSB of each byte (8 bits) in an image for a secret message, one can store 3 bits of data in each pixel for 24-bit

images and 1 bit in each pixel for 8-bit images. So much more information can be stored in a 24-bit image file.

2LSB: Depending on the color palette used for the cover image (i.e., all gray), it is possible to take 2 LSB's from one byte without the Human Visual System (HVS) being able to tell the difference.[4] One of the most popular message hiding schemes for palette-based images (GIF files) has been proposed by Machado [5, 6]. Another is spatial domain techniques include contrast adjustment, noise insertion etc.

Transforms: Transform domain techniques modify the transform coefficients of the image. The transform coefficients are obtained by applying transforms, such as the Fourier transform, discrete cosine transform or the wavelet transform, to the image [7, 8].

In recent times, Text steganography is used for Telegu and Arabic Language [9, 10]. But working on Bangla information hiding using bangla text in carrier image has never done before.

2.3 Embedding Process

At first the receiving bangla text is converted to text format to its corresponding hexa-decimal values. After that 0985 is subtracted from each of the bangla letter in order to reduce the 16 bit hexadecimal Unicode value per letter for bangle into 8 bit value per letter. Then the reduced bits are hidden into the innocent image file using 4 LSB method which is described later. After that the stego image file via insecure channel is sent. In the decoding phase, sent bits from the stego image is retrieved accurately. Here, 0985 is added with each 8 bit of data in order to get the 16 bit hexadecimal Unicode value of corresponding bangla text.

2.4 Image Conversion

At first, an innocent image file is taken as an input in which bangla text had to hide. When using a 24-bit color image, a bit of each of the red, green and blue color components is used, so a total of 3 bits is stored in each pixel. For example, the following grid id considered as 3 pixels of a 24-bit color image, using 9 bytes of memory.

0 0 1 0 0 1 1 1 1 1 1 0 1 0 0 1 1 1 0 0 1 0 0 0

0 0 1 0 0 1 1 1 1 1 0 0 1 0 0 0 1 1 1 0 1 0 0 1

1 1 0 0 1 0 0 0 0 0 1 0 0 1 1 1 1 1 1 0 1 0 0 1

In this case, for storing data the last bits of every pixel can be changed without imperceptible in human eyes. The image format for this should be gray scale for insertion of data. For this every image file had to be changed into bmp format.

2.5 Methodology

In the new 4LSB method, the innocent image file is used as a carrier file. So at first the image is converted into bmp format. Then the Bangla Unicode character is changed from hexa-decimal value. From these values 0985 is subtracted and then converted into binary stream which is embedded into the last 4 bits of every image RGB pixels. A short block of this method is depicted in Fig.2.

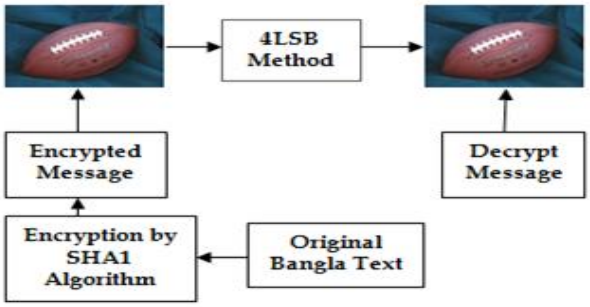


Fig. 2. Proposed methodology of the process

2.6 Algorithms

For Encoding:

Step 1: Reading the binary plaintext image from a file and compute the size I * J of the image.

Step 2: Decreasing 16 bit hexadecimal values of Bangla characters to 8 bit.

Step 3: Convolving the binary plaintext image with the pre-conditioned cipher and normalize the output.

Step 4: Binarize the output obtained in Step 3 using a thresh-old based on computing the mode of the Gaussian distributed cipher text.

Step 5: Inserting the binary output obtained in Step 4 into the lowest 4-bit layer of the host image and write the result to a file.

For Decoding:

Step 1: Reading the stego image.

Step 2: Regenerating the cipher using the same key used in Algorithm 1.

Step 3: Correlating the cipher with the input obtained in Step 1 and normalize the result.

Step 4: Quantize and format the output in bangla from Step 3 and write to a file.

2.7 SHA-1 Algorithm

Secure hash algorithm 1 (SHA-1) is that they are extremely sensitive to the message, i.e. changing even one bit of the input will change the output dramatically. They are added bit by bit through exclusive or with the encrypted binary data stream of the given secret message. It is clearly described in the whole paper.

3. MODEL IMPLEMENTATION

3.1 Data Encoding State

The user enters the text to be encoded and the password to be used for the encoding process, and clicks the encode data button.

3.2 Data Encoded State

After the details for the encoding purpose have been provided and the user clocks the Encode button, the data is encoded into a new image and displayed the encoded image. Here, the image with the encoded data seems very similar to the original with the encoded data also sounds almost similar. However, original image is modified by injecting a limited amount of bangla message data (or, in steganographic terms, “noise”) into the carrier image. If the amount of “noise” is too high, then the encoded image quality may become poor and visible

deteriorations may be observed. 3 pixels of a 24-bit image can be as follow in Table 1.

Table 1. Pixel values of Input Image

00101101	00011100	11011100
10100110	11000100	00001100
11010010	10101101	01100011

Corresponding binary value for "গ" and "ম" are "0000 1001 1001 0111" and "0000 1001 1010 1110". 0985's corresponding binary value is "0000 1001 1000 0101".

Now 0985 is subtracted from each of these values as follow:

0000 1001 1001 0111 - 0000 1001 1000 0101 = 0001 0010.

0000 1001 1010 1110 - 0000 1001 1000 0101 = 0101 1001.

In this paper, pixels is subtracted from another pixel values which is a preparation of stage for the decoding of the secret messages.

Now to hide "গ" and "ম" into the above 3 pixels of a 24 bit image 8 bit "0001 0010" and "0101 1001" is hidden respectively instead of their 16 bit values as follow in Table 2.

Table 2. 3 Pixel Value After Replacing Last Bit By Message

00101101	00011100	11011101
10100111	11000100	00001100
11010010	10101101	01100011

3.3 Data Decoding State

The user extracts the hidden bangla message in the image by specifying the correct password and clicking the decode button.

3.4 Invalid Password State

If the user tries to extract the hidden message in the image by specifying an incorrect password and clicking the decode button then a read-only message box appears showing that the password is invalid.

3.5 Encryption Process

The algorithm is the basic to demonstrate the concept. Any other algorithm may be substituted as long as basic information such as the length of the hidden data, the offsets of the hidden data in the message stream, and the actual data encryption key are encoded in the image. Another point to be considered is that this example specifically uses a bitmap image format; hence, the encoding algorithm is tied to the data format of the bitmap image format. However, these concepts can easily be applied to other graphic formats (JPEG, GIF, etc.) based on their respective internal data format specifications. The encoding process involves these steps:

- Input the innocent fresh image file which is 24 bit format (8-bit for red, 8 bit for green, 8 bit for blue) into a picture box such as jpeg, bmp, gif etc.
- Input the Bangla letter into text box. (By dragging from the menu bar)
- Determine the message length.
- Encrypt the message by hard encryption algorithm and correct password and clicking the decode button.

Input password not more than 8 characters.

- Determine the decimal value for each character of password and encrypted message.
- Convert every character of password and encrypt message into eight bit binary stream.
- Split the image pixel into three byte (R, G, and B components of a pixel) into binary stream.
- First inject the password not more than 8 characters then inject the message length. Inject every binary stream of text message to least significant 4 bit to every octet of each pixel.
- If message is finished then we would create a symbol to indicate the finishing of image
- If the end-of –stream is reached for the message stream, then save the file and exit.

3.6 4LSB in Bangla Character Hiding

In this paper Bangla characters is used as the hidden or secret messages which represents the Unicode value for bangla characters. For hiding this Unicode bangla character it is converted into binary format which is showed in the following figure. For hiding this binary coded value in the bmp image 4LSB method is used here. Here MATLAB 7 is used for doing this part and output is depicted in Fig.3.

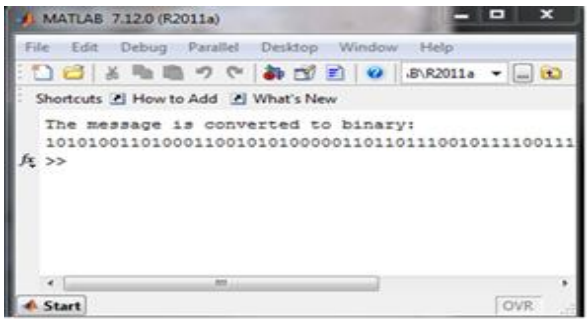


Fig. 3. The converted binary value of the corresponding Bangla secrete messages

3.7 Decryption Process

The decoding algorithm involves these following steps

- Input password into the password text box.
- If password is invalid. Not get original message (get a garbage message). Else get the encrypted message.
- Get the length of message hidden in the encoded image by splitting the R, G, and B components of the first several pixels in which we insert message length using bit-shift operator.
- Decode every four LSB of binary bit stream up to message length then
- Separate every 16 bit and form its corresponding decimal value.
- Form character decimal value from the Unicode character according to Unicode value and form the original message.

4. PERFORMANCE ANALYSIS

4.1 For Encoding Data

- An area that displays the standard image, which contains no hidden message.
- A text box for specifying the password (up to eight characters) for encrypting the secret message.
- A text box for inputting the actual secret Bangla message.
- A button that invokes the process of encoding the secret message into the image (that contains no hidden message) using the password specified. A read-only message box appears showing that the password is invalid.

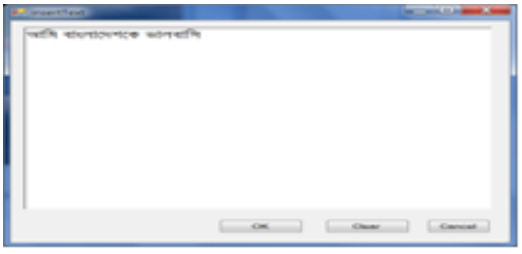
4.2 For Decoding Data

- An area that display the encoded image which contain the hidden message.
- A text box for specifying the password (up to eight characters) for decrypting the secret message.
- A button that invokes the process of decoding the secret message from the encoded image (which contains the hidden message) using the password specified.

4.3 Sample Input

The image in Fig.4 is defined as the innocent raw image which is fresh i.e. it has not contain any secret data. Then after pressing the encode button the text box will appear in which the input is required secret message in bangla. After pressing the ok button the system will inject the bangla characters into the image.

(a)



(b)

Fig. 4. Sample input for the performance analysis

4.4 Corresponding Output

At the output, bangla message containing image is sent to sender of another pc shown in Fig.5. Then, the image is browsed and the decoding button is pressed from the menu item for inserting the password. Then by providing the correct password and decoding key of the original secret message written in bangle is received.

Fig. 5. Sample output of the analysis

4.5 Comparison

The proposed algorithm of the image steganography system is tested by taking 32,000 bangla characters and hiding them in 50 standard images of different sizes. The result that is got from these experiments is recorded and is showed in the following graph in Fig.6. From this graph it is clear that the peak signal to noise ratio is much less in the new proposed method than the old existing method of LSB.

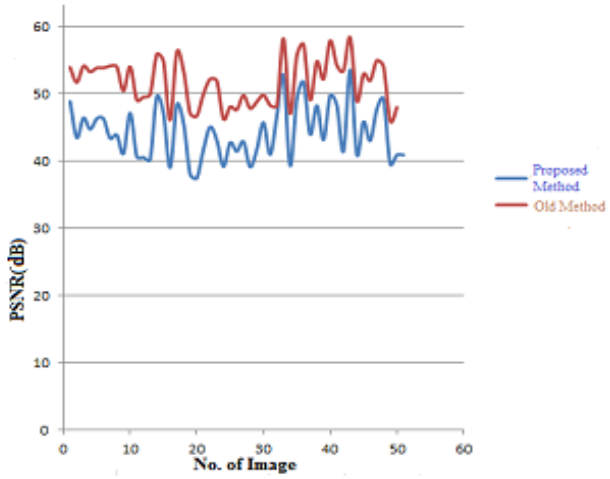


Fig. 6. PSNR(dB) comparison old vs. proposed method

5. FUTURE SCOPE

The future of this work is depended in extending it to hide all file formats. This allows for a broader spectrum of uses that one would be able to encode .exe, .doc, .pdf, mp3 and video file. Then would also be possible to implement batch image processing and statistical analysis so that it can be implemented through a dataset of images and detect data hiding. This designed system can also be implemented for hiding bangle texts within audio or video files.

6. CONCLUSION

This paper is a fraction of the steganalysis approach in that field. To date general detection techniques as applied to steganography have not been devised and methods beyond visual analysis are being explored. Too many images exist to be reviewed manually for hidden messages so development of a tool to automate the process will be beneficial to analysts. The ease in use and abundant availability of steganography tools has law enforcement concerned in trafficking of illicit material via web page, images, audio, and other transmissions over the Internet. Methods of message detection and understanding the thresholds of current technology are under

investigation. Success in steganographic secrecy results from selecting the proper mechanisms. However, a stego-medium which seems innocent enough may, upon further investigation, actually broadcasting the existence of embedded information for development in the area of covert communications and steganography must be continued. Research in building more robust methods that can survive image manipulation and attacks continues to grow. The introduction of this paper has made it possible to combine the future research areas of modern world such that it can act as a useful tool for demonstrating the impact of embedding message of different languages data according to an efficient newly invented stego-systems.

7. REFERENCES

[1] Sara Khosravi, Mashallah Abbasi Dezfoli Mohammad Hossein Yektaie, "A new steganography method based HIOP (Higher Intensity of Pixel) algorithm and Strassen's matrix multiplication", Journal of Global Research in Computer Science, Vol. 2, No.1, 2011.

[2] S. Katzenbeisser, and F. Peticolas, "Information Hiding Techniques for Steganography and Digital Watermarking", Artech House Inc, USA, 2000.

[3] Petitcolas, FAP; Anderson RJ; Kuhn MG (1999). "Information Hiding: A survey". Proceedings of the IEEE (special issue) 87 (7): 1062–78. doi:10.1109/5.771065. Retrieved 2008-09-02.

[4] Bret Dunbar, "A detailed look at Steganographic Techniques and their use in an Open-Systems Environment", SANS Institute, 2002.

[5] Jiri Fridrich and Du Rui. "Secure steganographic methods for palette images.", In Inter'l Workshop on Information Hiding, pages 47–60, 1999.

[6] N.F. Johnson and S. Jajodia. "Exploring steganography: seeing the unseen." In IEEE Comput., pages 26–34, February 1998.

[7] X.-G. Xia, C. G. Boncelet, and G. R. Arce, "A muldresoludon watermark for digital images," Proeedings of the IEEE International Conference on image Processing, Santa Barbara, CA, October 1997, pp. 548-551.

[8] Ki-Hyeok Bae, Sung-Hwan Jung, "A New Information Hiding Using Wavelet Coefficient Relation in JPEG 2000", Dept. of Computer Engineering, Changwon National University, Korea,pp.224-228

[9] Sravani Alameti, Sake Pothalalah , Dr.K Ashok Babu,"A New Approach to Telegu Text Steganography by Shifting Inherent Vowel Signs", International Journal of Engineering Science and Technology, Vol. 2 (12), 2010, 7203-7214.

[10] M.Hassan Shirali-Shahreza, Mohammad Shirali-Shahreza "Arabic/Persian Text Steganography Utilizing Similar Letters with Different Codes", The Arabian Journal for Science and Engineering, Volume 35, Number 1B.