A Proposed Algorithm for Text in Image Steganography based on Character Pairing and Positioning

Nitin Kaul
Electronics and Communication Engineering Department, Lovely Professional University
Punjab, India

Mrinal Chandra
Electronics and Communication Engineering Department, Lovely Professional University
Punjab, India

ABSTRACT
In a very less time internet has got heights. In every organization internet is the main employee without which the organization will doom. Most of the work is done, saved, edited online. With this such well-developed backbone, one important issue arises, and that is cybercrime. If communication is done online, that means data is being transferred, and security of that data is important. For this Cryptography and steganography plays a very important role. Number of algorithms are working to maintain the security one of which is steganography using image. The major problem with it is that on hiding large amount of data the image get distorted, to solve this issue we propose the idea of finding the data in the image rather than hiding it. In this paper an algorithm has been proposed to increase the capacity of data to be hidden. The proposed algorithm works on the bit level and is specifically text in image steganography. The main part of the algorithm is the key generated after searching text in image. The key will be position matrix where the text will be hidden. This algorithm will help to increase the size of data to be hidden.

Keywords
Text information, Image, character pairing, position matrix.

1. INTRODUCTION
Steganography is the art of hiding a file, message, image, or video within another file, message, image, or video. The word steganography combines the words steganos meaning "covered, concealed, or protected", and graphein meaning "writing". It is a way of hiding message or secret data into image which cannot be detected by Human Visual System (HVS) [3]. Steganography has four major parts i.e. message or secret data, carrier image, algorithm for steganography and the secret key. It has various useful application in online banking, voting, import or export of data for national or international government, military etc. As it is said coin has two faces so has this thing. Although of having positive side it has it negative side that It can be used for indecent activities; it is used by criminals and terrorists for their own secure communications also used for sending viruses. In the age when steganography was started, it was more like a physical hiding technique, like invisible ink, written document concealed in wax, but with the development in the technological era, physical steganography has been replaced by the digital media. In today's date, digital steganography is in the boom and is increasing day by day. How much data will be hidden, how much data will be transferred, all are the factors which need to be taken care of. The data which will be hidden in the cover data should be such that it should not affect the quality of the cover.

There are some parameters which need to be taken care always for a safe and secure data hiding system. Imperceptibility, Robustness, Embedding capacity.

- Imperceptibility: Imperceptibility is the property in which a person should be unable to distinguish the original and the stego-image.
- Embedding Capacity: Refers to the amount of secret information that can be embedded without degradation of the quality of the image.
- Robustness: Refers to the degree of difficulty required to destroy embedded information without destroying the cover image. [2]

In this paper, a very basic algorithm has been proposed to increase the capacity of embedding.

2. STEGANOGRAPHY
Steganography is the art of concealing an information without letting the intruder to know about the message. This is the basic difference between cryptography and steganography, in cryptography the data is changed by encrypting it with different algorithms and in steganography we hide the data without encrypting it or by encryption. Classification of steganography can be done on the basis of medium of cover message, on the basis of cover modification in the process of hiding secret message. W.r.t medium of cover message, steganography is categorized as [4]

- Text steganography
- Image steganography
- Audio steganography
- Video steganography

W.r.t cover modification in the process of data hiding, steganographic techniques has been categorized as

- Spatial domain steganography
- Transform domain steganography

Every steganographic system consists of an embedding algorithm and extraction algorithm. Embedding algorithm induces some changes in the cover data through which the secret data is embedded in the cover. The quality of the cover data is inversely proportional to the capacity of the secret information to be hidden. Less the capacity more the quality and vice versa. If the capacity of secret information is less the probability of detecting the changes in the cover data is low,

and is the capacity of secret information is high, the probability increases [1].

The techniques which are available for data hiding, uses some kind of algorithm to embed data in the cover, that is why the capacity limitation is there, but in the proposed algorithm secret data will not be embed in the cover but will be mapped from the cover data, because of which the capacity of secret data will increase.

3. PROPOSED ALGORTIHM

The proposed algorithm works like a search engine like Google. When we enter some information to be searched from a search engine, it searches that particular data on the whole web, as the web is the place where search engine can find anything. Similarly in our algorithm image is that web from where the data will be searched. The proposed algorithm works for text in image steganography. Where information to be hidden is text and image will act as the cover data. The whole algorithm will work on the bit level. Every character in text is of 8 bits and every pixel of an image on which we will work is also having 8 bits. The main objective of this algorithm is to increase the capacity of the data to be hidden. There are number of steganographic techniques, some provides good security, some provides good robustness, but in every technique the limitation is capacity, if one parameter is made perfect, other parameters are compromised. In this algorithm, character pairing will work but after splitting. Precisely, in this algorithm there is no need to hide the data Table 1 below gives the explanation about the above procedure.

Table 1 Splitting pattern of Text information

Text information	Binary format of Text	Splitting of 8 bits into 2 halves of 4 bits		Pairing of alphabets
HIDING	**H** 0 1 0 0 1 0 0 0	0 1 0 0	1 0 0 0	0 1 0 0 – 1st half of **H** 0 1 0 0 – 1st half of **I**
		1st half	**2nd half**	
	I 0 1 0 0 1 0 0 1	0 1 0 0	1 0 0 1	1 0 0 0 – 2nd half of **H** 1 0 0 1 – 2nd half of **I**
		1st half	**2nd half**	
	D 0 1 0 0 0 1 0 0	0 1 0 0	0 1 0 0	0 1 0 0 – 1st half of **D** 0 1 0 0 – 1st half of **I**
		1st half	**2nd half**	
	I 0 1 0 0 1 0 0 1	0 1 0 0	1 0 0 1	0 1 0 0 – 2nd half of **D** 1 0 0 1 – 2nd half of **I**
		1st half	**2nd half**	
	N 0 1 0 0 1 1 1 0	0 1 0 0	1 1 1 0	0 1 0 0 – 1st half of **N** 0 1 0 0 – 1st half of **G**
		1st half	**2nd half**	
	G 0 1 0 0 0 1 1 1	0 1 0 0	0 1 1 1	1 1 1 0 – 2nd half of **N** 0 1 1 1 – 2nd half of **G**
		1st half	**2nd half**	

The shaded array in the last column of the table 1 is one block of 2x4 bits which will be mapped from all the 2x4 bit blocks of cover image. Same will be applicable for all the 2x4 bit blocks.

NOTE: The pairing will be proper in case the characters are in even in number, if there are odd number of characters, then we will add a space in the end and then process the data.

Decryption

After character pairing and mapping of concern data from the cover, rows and columns of the new array will be the data actually, just searching will do the work. Basic idea behind the algorithm is that every information is made of bits, whether it is video, image or text. Manipulation in bits can do the wonders. Every bit can be found be in any kind of information. Thus in this algorithm text data will be searched in an image by comparing but after proper manipulation.

Encryption

The main part of any security algorithm is encryption. In this algorithm image and text information will be converted into binary format. The text which need to be secured is an 8 bit data. The algorithm for encryption is discussed below:

Step 1: Divide every 8 bit character into 2 halves of 4 bits

Step 2: Make an array of 4 bits formed after splitting the alphabets. Now pair the 2x4 bits such that 1st 4 bits will represent the 1 half of 1st alphabet and next 4 bits will represent 1st half of 2nd alphabet.

Step 3: Store the 2x4 bit data in a matrix. And search the matrix individually in the cover image.

Step 4: The searching of 2x4 matrix from the cover image will be done by matrix overlapping horizontally as well as vertically, so that none of the data will be left for searching.

Step 5: After searching, store the rows and columns where the same data was found in the cover image. That row and column will give the position where the data is hidden.

which will tell us about the position of the secret information. So that array will be acting as the key without which the data cannot be judged by the intruder. For the decryption process, once the receiver knows the position array, fetching of information at that particular position will take place, and the retrieved data will be the actual information to be transmit.

By this algorithm, the quality of the image will not be affected as the data hiding is not taking place, not even a single bit of cover image is getting changed, that means the quality is intact.

Example:

As an example, Lena image is considered as the cover image and the secret text data is "HIDING"

Text information	Pairing of alphabets
HIDING	0 1 0 0 – 1st half of **H** 0 1 0 0 – 1st half of **I**
	1 0 0 0 – 2nd half of **H** 1 0 0 1 – 2nd half of **I**
	0 1 0 0 – 1st half of **D** 0 1 0 0 – 1st half of **I**
	0 1 0 0 – 2nd half of **D** 1 0 0 1 – 2nd half of **I**
	0 1 0 0 – 1st half of **N** 0 1 0 0 – 1st half of **G**
	1 1 1 0 – 2nd half of **N** 0 1 1 1 – 2nd half of **G**

Fig 1 Mapping of text information from image data

The above diagram represents that the individual blocks are mapped from the individual blocks of the Lena image until the matching is done. After mapping, the matched positions are stored as given in table 2

Table 2 Position matrix of mapped data

Information data	Row	Column
0 1 0 0 – 1st half of **H** 0 1 0 0 – 1st half of **I**	13338	3-6
	13339	3-6
1 0 0 0 – 2nd half of **H** 1 0 0 1 – 2nd half of **I**	15007	3-6
	15008	3-6
0 1 0 0 – 1st half of **D** 0 1 0 0 – 1st half of **I**	3302	1-4
	3303	1-4
0 1 0 0 – 2nd half of **D** 1 0 0 1 – 2nd half of **I**	15035	3-6
	15036	3-6
0 1 0 0 – 1st half of **N** 0 1 0 0 – 1st half of **G**	10327	3-6
	10328	3-6
1 1 1 0 – 2nd half of **Ns** 0 1 1 1 – 2nd half of **G**	1147	2-5
	1148	2-5

The interpretation is like, the first block of 2x4 bits are found in 13338-13339 rows and 3-6 columns of the image. Similarly the remaining data is found in other locations. The disadvantage is the length of the matrix generated but the principle can work for increasing the capacity of hiding data.

4. FUTURE WORK

The proposed algorithm has worked for the text in image, and has increased the capacity of the data to be hidden but it has given us the limitation that the size of key which was generated was very much. Future work can be done to minimize the key and securely transmit it to the receiver. Furthermore the work can also be extended to image in image steganography as image size is more as compared to text and it will be more challenging to work on.

5. CONCLUSION

This paper discussed that the size of the data to be hidden can be increased. Character bit pairing and mapping has been done in the algorithm to find out the key which will help to trace out the secret data from the cover image. In this paper, work has been done on text secret data which was fetched out from an image. The results showed that, the quality of the cover data was not disturbed as no changes were done to cover image to hide the secret text, rather secret text was searched out from the cover image.

6. REFERENCES

[1] Jessica Fridrich, Miroslav Goljan, Dorin Hogea, "New methodology for breaking steganographic techniques for JPEGs", www.ws.binghamton.edu-/fridrich/research/ jpeg01.

[2] C.P.Sumathi, T.Santanam, G.Umamaheswari, "A Study of Various Steganographic Techniques Used for Information Hiding", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.4, No.6, December 2013.

[3] K. Muhammad , J.Ahmad , N. U. Rehman , Z. Jan , R. J. Qureshi, "A Secure Cyclic Steganographic Technique for Color Images using Randomization", Technical Journal, University of Engineering and Technology Taxila, Vol. 19 No. III-2014.

[4] Hedieh Sajedi, Mansour Jamzad, "Secure Cover Selection Steganography", J.H. Park et al. (Eds.): ISA 2009, LNCS 5576, pp. 317–326, 2009. © Springer-Verlag Berlin Heidelberg 2009.