Efficient Data Hiding Scheme Using Fourteen Square Substitution Algorithm

Mr. Shatendra Dubey  
Dept. of Computer Science & Engg.  
NRI Institute of Information Science & Tech.  
Bhopal- 462021, India

Deepali V. Patil  
Dept. of Computer Science & Engg.  
NRI Institute of Information Science & Tech.  
Bhopal- 462021, India  
Deepalipatil86@yahoo.co.in

Abstract—this paper describes techniques about secret communication between two people. We have used both cryptography and steganography to implement this. Firstly, we need to encrypt the message using new cipher algorithm called as fourteen square substitution cipher algorithm. Inside this algorithm, first 8 tables are used in upper case or smaller case or combination of both and remaining 6 tables are special symbols or digits. At the time of converting simple text into cipher text, it converts that cipher text into bytes and divide it by 2 and the values which come is called as index variable. Depending upon the index variable value, insert that cipher text in the career image at 6th and 7th bit location or 7th and 8th bit location. Change those bits depending upon index variable x. Index variable x should have value 0 or 1. The index variable value will be changed from 0 to 1 or 1 to 0 after each embedding. The first value of the index variable depends upon the length of the cipher text. After completion of embedding, the resultant image should be sent to the receiver and receiver should retrieve the cipher text from the said locations and then decrypt by using the fourteen square cipher algorithms to get the secret message. The embedding locations are not same in all pixels, so it is a stronger approach. We have implemented this algorithm using c#.net programming language. The twelve square substitution algorithm covered alphabets as well as special character but alphabet ‘q is missing’ over here. Fourteen square substitution algorithms is used to overcome this drawback. In this algorithm, both upper case and lower case alphabets are used. It will provide more payload too.

Index Terms—Steganography, Cryptography, Fourteen square substitution, encryption, decryption.

I. INTRODUCTION

Introduction about Network Security

Network security is the collection of tools designed to protect data and to prevent from hackers. Nowadays the word Network Security is somewhat misleading, because virtually all business, government and academic organizations have their interconnected networks. Such a collection is pronounced as Internet and therefore it called as Internet Security.

Introduction about Steganography and cryptography

Since the start of internet, one of the most important factors of Information Technology and Communication has been the security of information, for that two techniques are used cryptography and steganography.

Steganography

Steganography is not actually a method of encrypting message but hiding them within something else to enable them to pass undetected. The word steganography comes from Greek word where “Steganos” means “Covered” and “Graphy” means “Writing” The innocent files can be referred to as cover text, cover image or cover audio as appropriate. After embedding the secret message it is referred steg medium.

Cryptography

Cryptography encodes information in such way that nobody can read it, except the person who holds the key. Cryptography comes from Greek word where “Crypto” means “Secret” and “Graphy” means “Writing”.

Steganography versus Cryptography

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Context</th>
<th>Steganography</th>
<th>Cryptography</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Host files</td>
<td>Image, Audio, Text etc.</td>
<td>Mostly Text Files</td>
</tr>
<tr>
<td>2</td>
<td>Hidden Files</td>
<td>Image, Audio, Text etc.</td>
<td>Mostly Text Files</td>
</tr>
<tr>
<td>3</td>
<td>Result</td>
<td>Stego File</td>
<td>Cipher Text</td>
</tr>
<tr>
<td>4</td>
<td>Type of Attack</td>
<td>Steganalysis</td>
<td>Cryptanalysis</td>
</tr>
</tbody>
</table>

Need of Steganography and Cryptography

Encryption allows secure communication and it requires a key to read the information. An attacker is not able to remove the encryption but it is easy to modify the file and making it unreadable for the intended recipient.

Steganography allows secure communication It cannot be removed and it requires significantly altering the data in which it is embedded. The embedded data will be confidential until an attacker is able to find a way to detect it.
II. LITERATURE SURVEY

Following are the different techniques which can be used for secure communication:

When hiding the information inside images Least Significant Bit (LSB) method is used. In this method 8\(^{th}\) bit of every byte of the carrier file is substituted by one bit of the secret information. Although this technique is simple, the disadvantage is it causes noticeable distortion when the number embedded bits for each pixel exceeds three. There are several adaptive methods for steganography have been proposed to reduce the distortion caused by LSBs substitution [1].

In Image Encryption approach, Least Significant Bit (LSB) is being used to hide the data within encrypted image data and embed the information in LSB. After embedding if the entropy and correlation values of stego image and original image are same then the process is a secure one. To implement this, number of horizontal and vertical blocks will be generated at the sender side, and before transmitting it to the receiver, it mixed with the encrypted image. The receiver will need this information to reconstruct the same secret transformation table, after extracting the secret information from the encrypted image. Instead of sending whole secret transformation table, only secret information is sent. The binary representation of the hidden data is used to overwrite the LSB of each byte within the encrypted image data randomly based on the secret key before transmission. The values of the correlation and entropy before and after the insertion process are expected to be the same. The advantage of this is it will be used to reduce the chance of the encrypted image being detected and then enhance the security level of the encrypted image. [2]

This topic deals with three main challenges security, capacity and imperceptibility. This is achieved by hybrid data hiding scheme which incorporates LSB techniques with key-permutation method. Also this topic proposes an optimal key permutation method using genetic algorithm for best key selection. In here, a method has been proposed based on LSB substitutions, where random key is generated and distributed to the communicating parties. The secret data is embedded in the k LSBs of the Host image, where k is less than or equal to 3. It works poorly if k is greater than 3 because the number of all possible keys permutations will grow as k increases. To obtain the optimal embedding result the simplest method is to calculate the PSNR for each substitution, and select the one having the maximum PSNR value for each substitution, for the optimal result. But it is very impractical and time consuming, for that genetic algorithm is used to solve this problem, where genetic algorithm is a randomized search procedure that is commonly used to solve optimization problems. The advantage of this method is it provides good image quality and large message capacity as well as increase the system immunity. When increasing number of keys, it decrement in computation time also, at the same time system security improves. [3]

Here a new steganography techniques is used which embeds the secret message in frequency domain by using Discrete Wavelet Transform (DWT). The frequency domain transform is Haar-DWT. It consists of 2 operations—horizontal operation and the other is vertical operation. At first, scan the pixels from left to right in horizontal direction then perform the addition and subtraction operations on neighboring pixels and then store the sum on the top and the difference on the bottom. Benefit of this is security is maintained as well as no message can be extracted without the “key matrix” and decoding rules. This technique can be used to reduce the extra data in the stego-image compress the size of key matrix as far as possible. [4]

Here new method of digital steganography spread spectrum image steganography (SSIS) to inherent noise places is proposed. In this technique error-control coding, image restoration, and those similar to spread spectrum communication, are combined within the SSIS system. Fundamental concept is the embedding of the hidden information within noise which is then added to a digital cover image. This method could be extended to color imagery and audio signals. [5]

A new algorithm to hide data inside image using steganography technique is proposed. Here a bitmap (bmp) image will be used to hide the data. Data will be embedded inside the image using the pixels. Then the pixels of stego-image can be accessed back in order to retrieve back the hidden data inside the image. Two stages are involved. The first stage is to come up with a new steganography algorithm in order to hide the data inside the image and the second stage is to come up with a decryption algorithm using data retrieving method in order to retrieve the hidden data that is hided within the stego-image. Advantage of this technique is SSIS maintains privacy, confidentiality and accuracy of the data. [6]

Here a highly efficient steganography protocol is proposed. It is based on hamming codes, the embedding and the retrieval algorithm which have the same computational cost. The main idea behind this technique is to use a product code of two hamming codes with goal of improving the embedding rate. [7]

This technique gives two different schemes are investigated. The first one is derived from a blind watermarking scheme. The second scheme is designed for steganography such that perfect security is achieved, that means the relative entropy between cover data and stego data tends to zero. In this technique, information embedding has been investigated in particular in the context of digital watermarking. For digital watermarking, information embedding techniques have to be designed such that subsequent processing does not destroy the embedded information. This property makes digital watermarking technology also attractive for steganography when information embedding is followed by lossy compression. Advantage of this technique the performance of the schemes is compared with respect to security, embedding distortion and embedding rate. [8]

In this technique defines the goal of steganographic method is to minimize the visually apparent and statistical
In next method a text steganography using two square reverse ciphers is proposed. In this, there are two steps for encryption. In the first step, get the first step cipher by using table1 and table2. After getting the first step cipher we divide that into 2-2 characters. After that swap the 2-2 characters positions and get the final cipher. The cover image is divided into bytes. In each byte do not embed; but in few selected bytes based on the bit pattern of the cipher text. In the selected bytes the 7th bit position that is LSB minus one are to be embedded. The 8th bit means the LSB bit position. Advantage here it provides two level security- cryptography and steganography. As compared to LSB method this algorithm is much better in terms of intrusion prevention. But disadvantage is Table1 and Table 2 is divided in to alphabets where ‘q’ is missing also digits are not included. In future this approach can be extendable to send secret image in cover image. This approach can be extendable to audio and video carrier files. [10]

A method for steganography using six square substitution ciphers which includes only alphabets. [11]

A method for steganography using twelve square substitution ciphers which includes both alphabet and digit but the alphabet ‘q’ and some special character ‘space’ is missing. [12]

III. FOURTEEN SQUARE SUBSTITUTION ALGORITHM

In previous paper [11] author represented a six square substitution cipher which includes only alphabets. After that author made some changes in six square substitutions cipher and came with twelve square substitution ciphers [12] which include numerals and special characters. But in that some character and special symbols are missing to overcome that a new algorithm is proposed that is fourteen-square substitution that encrypts alphabets, digits and special characters. It uses eight 9 by 6 matrices each arranged in a square, as shown in table-1. Each of the 9 by 6 matrices contains the letters of the alphabet (upper case and lower case) and another six 6 by 7 matrices arranged in squares for digits and special characters, as shown in table-2. All the special characters and digits from your desktop/laptop keyboard are included in this table.

So the following describes that how the table 1 is prepared: In square-1, we have taken fifty two alphabets and two special characters, out of which twenty six are capital letters and twenty six are small letters. In each row we have arranged nine alphabets and each column contains six alphabets.

2 is made from square-1 by taking the first row of square-1 to sixth row place and other rows one position up. Similarly square-3 is created from square-2 by taking the first row of square-2 to sixth row place and other rows one position up. The same thing about square-4 which is created from square-3 by taking the first row of square-3 to sixth row place and other rows one position up. In square-5, We have converted rows into column and inter changed first and last alphabets. The same steps follows in square-6 to square-8 by taking first row of previous square and to sixth row place and other rows one position up. The same method we have followed in table 2.

Table 1: Plain Text and Cipher Text (Alphabets)

Table 2: Plain Text and Cipher Text (Digits and Special Characters)
of square-12. For second special character (including digits), the plain text is in square-10 and cipher text is in same row and column location of square-13. For the third special character (including numbers) the plain text is in square-11 and cipher text is in same row and column location of square-14. Similarly fourth special character (including numbers) corresponds to square-9 and square-12, 5th special character (including numbers) corresponds to square-10 and square-13, 6th special character (including numbers) corresponds to square-11 and square-14 and so on.

For example if the plain text is: Abc
Its cipher text would be: mCI

IV. THE EMBEDDING PROCESS

The carrier image is transformed into binary form. Each pixel becomes 1 byte. The cipher text of the secret message is converted into bytes. Now calculate the number of bytes, suppose it is n. Divide it by 2, say it is x. The x called as the index variable. The value x=0, corresponds to 6th and 7th bit locations of any pixel (byte) of the digital image. If present value of x=0 hide the two bits of cipher text in 6th and 7th bit locations of the present pixel (byte), and next value of x is 1 for the next pixel. If present value of x=1 hide the two bits of cipher text in 7th and 8th bit locations of the present pixel (byte), and next value of x is 0 for the next pixel.

Table 3: BYTE Selection using Index Variable

<table>
<thead>
<tr>
<th>Carrier File Byte</th>
<th>Operation</th>
<th>Location</th>
<th>Index Variable, x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Byte A</td>
<td>Embed(11)</td>
<td>6th and 7th</td>
<td>1</td>
</tr>
<tr>
<td>Byte B</td>
<td>Embed(00)</td>
<td>7th and 8th</td>
<td>0</td>
</tr>
<tr>
<td>Byte C</td>
<td>Embed(10)</td>
<td>6th and 7th</td>
<td>1</td>
</tr>
<tr>
<td>Byte D</td>
<td>Embed(11)</td>
<td>7th and 8th</td>
<td>0</td>
</tr>
<tr>
<td>Byte E</td>
<td>Embed(01)</td>
<td>6th and 7th</td>
<td>1</td>
</tr>
<tr>
<td>Byte F</td>
<td>Embed(11)</td>
<td>7th and 8th</td>
<td>0</td>
</tr>
<tr>
<td>Byte G</td>
<td>Embed(10)</td>
<td>6th and 7th</td>
<td>1</td>
</tr>
<tr>
<td>Byte H</td>
<td>Embed(10)</td>
<td>7th and 8th</td>
<td>0</td>
</tr>
<tr>
<td>Byte I</td>
<td>Embed(10)</td>
<td>6th and 7th</td>
<td>1</td>
</tr>
<tr>
<td>Byte J</td>
<td>Embed(10)</td>
<td>7th and 8th</td>
<td>0</td>
</tr>
<tr>
<td>Byte K</td>
<td>Embed(10)</td>
<td>6th and 7th</td>
<td>1</td>
</tr>
<tr>
<td>Byte L</td>
<td>Embed(10)</td>
<td>7th and 8th</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>So on</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example:
Consider the cipher text to be sent is:
11001011 01111010 10101010 10011001 01010101.
This data is five bytes. So n=5 and x=1. Suppose the different bytes of the digital image are A,B,C,D,E etc. From table-3 we can see that in byte A of the carrier file we embedded the data bits 11 in 6th and 8th bit locations, and next value of x becomes 0. We embed the next data bits 00 into byte B in 6th and 7th bit locations, next value of x becomes 1. Now we embed the next two bits 10 into C in 7th and 8th bit locations and so on. See table-3. In every image there will be some bytes representing the image features which should not be altered. In JPEG images of size more than one Mega Bytes, there will be a maximum of hundred bytes carrying the image characteristics, if we modify these bytes the image will be disturbed. So these bytes should not be altered. For different image formats like BMP, JPG, TIF it is different. For JPG it is around 100 bytes. Normally these are the first 11 bytes of image.

V. RESEARCH FINDING

Steganography is method of hiding the secret message in to image, audio, or video. There are many methods for hiding image, video, or audio.

This section tries to give an overview of the most important steganographic techniques in digital images. Graphics interchange format (GIF), Joint Photographic Experts Group (JPEG), and to a lesser extent—the portable network graphics (PNG) these are most popular image formats on the internet. Most of the techniques developed were set up to exploit the structures of these formats with some exceptions in the literature that use the bitmap format (BMP) for its simple data structure.

To hide image there are two techniques are present-
1) Spatial Domain Techniques which generally use a direct least significant bit (LSB) replacement technique.
2) Frequency Domain Techniques based methods such as discrete cosine transforms (DCT), Fourier transforms (FT) and discrete wavelet transforms (DWT).

Steganography in the image spatial domain

Spatial Domain Techniques embed a secret message by substituting insignificant parts of the cover-media with secret message bits; the receiver can extract the secret message if he/she has knowledge of the positions where the secret message bits have been embedded. Since only small changes are made in the embedding process, the sender assumes that they will not be noticed by a third person. It is also called as substitution techniques.

The very commonly used technique for embedding hidden information in digital images is the LSB Substitution, where the LSB of each image byte is replaced with secret message bit.

The embedding process consists of choosing a subset of cover-image bytes and performs the substitution operation on them, which replaces the LSB of a cover image byte with a secret message bit. The substitution operation is able to also be changing more than one bit of the cover then for instance by storing two secret message bits in the two LSB of the one cover-image byte. In the extraction process, the LSBs of the selected cover-elements are extracted and lined up to reconstruct the secret image. The advantage of this method is easy and computationally fast.

Steganography in image frequency domain

For frequency domain methods, the first step is to transform the image data into frequency domain coefficients by some mathematical tools (e.g. FT, DCT, or DWT). Then, according to the different data characteristics generated by these transforms, embed the information into the coefficients in frequency domain. After that the coefficients transformed...
back to the spatial domain and the whole embedding procedure is completed. The main disadvantage of this method, it is computationally complex and hence slower.

The emerging techniques such as DCT, DWT and FT are not prone to attacks too, especially when the hidden message is small. This causes because they alter coefficients in the transform domain thus image distortion is kept to a minimum. Generally, these methods suitable to a lower payload compared to spatial domain algorithms.

VI. CONCLUSION

It is observed that the algorithm works fine. It provides two levels of security One at the cryptography level and the other at the steganography level. If at all the intruder suspects it is very difficult for him to steal the data. The degradation in image quality cannot be noticeable. The image size does not increase after embedding the cipher text.

REFERENCES