STEGANOGRAPHY FOR TWO AND THREE LSBs USING EXTENDED SUBSTITUTION ALGORITHM

R.S. Gutte¹, Y.D. Chincholkar² and P.U. Lahane³

¹Department of Electronics and Telecommunications Engineering, Dnyanganga College of Engineering and Research, India
E-mail: granjeetsinh@gmail.com

²Department of Electronics and Telecommunications Engineering, Sinhgad College of Engineering, India
E-mail: ydc2002@rediffmail.com

³Department of Electronics and Telecommunications Engineering, RMD Sinhgad School of Engineering, India
E-mail: prashlahane@gmail.com

Abstract
The Security of data on internet has become a prior thing. Though any message is encrypted using a stronger cryptography algorithm, it cannot avoid the suspicion of intruder. This paper proposes an approach in such way that, data is encrypted using Extended Substitution Algorithm and then this cipher text is concealed at two or three LSB positions of the carrier image. This algorithm covers almost all type of symbols and alphabets. The encrypted text is concealed variably into the LSBs. Therefore, it is a stronger approach. The visible characteristics of the carrier image before and after concealment remained almost the same. The algorithm has been implemented using Matlab.

Keywords:
Steganography, Cryptography, Encryption, Decryption, Extended Square Substitution Algorithm

1. INTRODUCTION

Steganography is a method of secret communication in which one kind of information is embedded into other information. The word Steganography has a Greek origin. It means covered writing. Numerous evidents from the ancient history may be found as hiding messages within wax tablets and writing the message on messenger's body by Greece people. The common example of the Steganography is, writing secret message on a paper with onion juice or ammonia salts and the secret message can be then exposed by heating the paper. Once the cover object has material concealed in it, it is called Stego-object. Various multimedia files can be used as carriers to hide the data. Steganography can be classified on the basis of carriers used for concealing the data as image, audio and video Steganography. It is the age of the internet we get the service just on a click. We use the internet for searching information, downloading and uploading of multimedia, to check E-mails, E-banking. Online reservations, therefore all our online information must be kept secure. Encryption and Steganography are used for protecting the transmitted data over internet. The advantage of Steganography over encryption is it keeps the existence of the data secret. The proposed method provides two layered security. First, the data is encrypted and second, the encrypted text is then concealed into the LSB plane of carrier image, thus the strength of Steganography can be enhanced by combining it with cryptography. Concealing the data into LSB plane of carrier image does not much affect its quality.

2. EXISTING APPROACHES

Various methods of image Steganography have been proposed. The Steganography algorithm should be able to keep the stego media as same as the original carrier and it should conceal more amount of data. In other words it should not violate the visual characteristics of the carrier. The Steganography methods can be classified as Spatial Domain Based, Pallete based, Transform domain based Steganography Algorithms. The simplest and mostly used spatial domain Steganography is LSB Steganography in which the LSBs of the Carrier are replaced with the binary data bits. In Palette based Steganographic methods a color image is transformed into palette based color representation and the data message is hidden within the bits of palette. Transform based techniques hide data in the coefficients of the represented domain. The transforms like Discrete Fourier transform, Cosine transform, Wavelet transforms can be used to generate the coefficient and the obtained coefficients are altered or replaced [1].

Mohmmad Ali Bani Younes et. al. proposed Steganography approach based on Image encryption exchange using LSB method. In Image encryption approach we can encrypt the image and insert the secret information in LSBs and after embedding if the entropy and correlation values of stego image and original image are the same then the process is a secure one. The insertion position of the information is randomly selected depending on the key [2]. Ross J. Anderson and Fabien A.P. Petitcolas has discussed about the unified terminologies used in Steganography and its state of the art, also they shown that public key Steganography may sometimes be possible in presence of an active warden[3]. H. Motameni et.al. proposed Labeling Method in Steganography in which different colors are labeled, dark places in the image are recognized and then text is inserted at the dark places [4]. Tanmay Bhattacharya et.al. proposed Steganography technique using DWT and Spread Spectrum in which decomposition of the cover image is done using DWT and the secret images are concealed into each band using a pseudo random sequence and a session key [5]. P. Mohan Kumar and D. Roopa proposed a Steganography method for tamper-proofing. They have suggested an application that the Photographers working in enemy area can use this method to send the spy photographs [6]. Joachim J. Eggers et.al. investigated about two different schemes. The first one is derived from blind watermarking and second is a Steganography method based on noiseless communication channel. The format of stego image was JPEG [7]. Mohammed A.F. AlHusainy et.al. used enough number of bits from each pixel in an image to map 32 characters [8]. Lisa M Marvel et.al. proposed Spread Spectrum
Steganography. This method uses inherent noise to hide information within the digital image. A binary signal is variably embedded within samples of a low power white Gaussian noise sequence consisting of real numbers. Since the power of the embedded signal is much lower than the cover image therefore it is difficult to detect [9]. Two way block matching for image in Image Steganography was proposed by Ran-Zan Wang and Yeh-shun Chen. This method can embed relatively large important images into relatively small cover images [10]. Xinpeng Zhang and his colleagues proposed an approach called ‘multibit assignment Steganography for palette images’; secret data can be embedded at the same colour pixels of the gregarious palette images also they introduced a smarter Optimal Parity Assignment (OPA) method to have smaller distortion [11]. Weiming Zhang et.al. presented Double Layered ±1 Data Embedding Scheme in which the secret message is embedded into two LSB planes using binary covering codes and wet paper codes [12]. Alvaro Martin et.al. did the statistical analysis of some popular Steganography techniques, they found that the fundamental statistics of natural images are altered by that hidden non-natural information. If the change lies within the intrinsic variability of the statistics the detection is difficult [13]. H. Rifa-Pous and J. Rifa presented a steganographic protocol based on hamming code [14]. Abbas Cheddad et.al. did the brief review of digital image Steganography and compared between Steganography, Watermarking and Encryption. Also, they presented some of the applications of the Steganography such as Smart Identity cards, Medical Imaging Systems, Copyright. They provided a brief survey about Steganography methods such as Steganography exploiting the image format, Steganography in image spatial domain, Steganography in image frequency domain, Adaptive Steganography [15]. Hardik J. Patel and Preeti K. Dave proposed LSB based image Steganography method for JPEG images. The binarised values of the pixels of secret image were inserted at the one, two, four LSBs of the carrier image and they did the statistical analysis with Histograms. They concluded that inserting the secret data at one, two LSB positions provides good stego-image quality but the fewer payloads whereas inserting data at four LSBs gives increased payload but the degraded image quality [16]. S. Gurusubramani and his colleagues introduced the concept of Multi Layer Data Security (MLDS) [17].

Piyush Marwaha, Paresh Marwaha proposed Visual Cryptographic Steganography in an image which uses a block or a grid cipher for cryptography and this cipher containing the data is mapped in a 3-D matrix of RGB [18]. Mohammad Shirali-Shahreza presented Steganography in MMS to hide secret text and encryption. He used J2ME (Java 2 Micro Edition) for implementation [19]. The six square substitution algorithms were consisting of only the alphabets [20]. The twelve square substitution algorithm covered alphabets as well as the special characters but the alphabet ‘q’ was missing, they did not consider the letters like μ, π, Ø, ×, ð, þ, ø, ÷, ¯, ƒ etc also they had not encrypted the space character [21]. This algorithm also considers the characters ‘q’, ‘space’ as well as the above mentioned symbols. The proposed technique uses both cryptography and Steganography for more secured technique. Firstly, the secret message(M) is encrypted using new cipher algorithm called extended square substitution cipher algorithm(E) with a secret key(k) ‘axbwC’, and then the cipher text(M) in the carrier image is inserted at LSB places. After embedding the stego-image(S) is sent to the receiver and receiver retrieves the text from the stego image using the same cipher key which was used at transmitter. The embedding locations are not same in all pixels, so it is more secure. The Block diagram describes the process of formation of the stego-Image and retrieval of text message from it. Function f(E) indicates transmitter and (f(E))^(-1) indicates receiver.

3. EXTENDED SUBSTITUTION CIPHER

The Extended Substitution Cipher Algorithm includes almost all type numerals, special characters and symbols. There are 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 and another 12 by 6 matrices arranged in squares for digits and special characters and mathematical symbols, as shown in Table.2. All the special characters and digits from desktop/laptop keyboard are included in this table. There are two Tables. Table.1 covers all the small and capital alphabets. Table.2 covers almost all type of special characters and symbols. The matrices of Table.1 have 9 columns and 6 rows, the matrices of Table.2 have 12 columns and 6 rows. We tried to arrange the alphabets, special characters and symbols in such a way that it will not have a repeated substitution.

While extracting the secret message, if the special character and digits appeared it is referred to Table.2 and if the capital or small alphabet appears then it is referred to Table.1. It is decided according to its ASCII code. Now we will see the working of this algorithm. This can be elaborated using a simple example suppose one word ‘Jay Hind’. Its first alphabet is ‘J’. Therefore it is referred to Table.1 and Matrix-1. Its corresponding row and column is checked and this alphabet ‘J’ is replaced with the alphabet having same row and column in Matrix-5. The next alphabet is ‘a’ therefore it is referred to first table and Matrix-2. Its corresponding row and column is calculated then it is then replaced with the alphabet having same row and column position from Matrix-6. When a symbol or special character appears, it is referred to Table.2 and the similar procedure is followed. The Matrix-1 has its cipher in Matrix-5; Matrix-7 has its cipher in Matrix-11 and so on. The following examples can illustrate the working of the algorithm.

i. Plain-text: Jay Hind!!
   Cipher text: IUH2UgAu(

ii. Plain-text: Qwertyu!@#$%^&*()\_
   Cipher text: 47dFgfym1234567890

The algorithm is able to hide almost all the alphabets as well as special characters and mathematical symbols.
Table.1. Plain Text and Cipher Text (Alphabets)

<table>
<thead>
<tr>
<th>Matrix-1</th>
<th>Matrix-2</th>
<th>Matrix-3</th>
<th>Matrix-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a X b W c V d U e</td>
<td>a X f T k p u p z</td>
<td>x y z v w s t u A</td>
<td>s B K T a j t C L</td>
</tr>
<tr>
<td>f T g S h R i q j</td>
<td>L b g H i D q v @</td>
<td>B C D E F G H I J</td>
<td>U b k u D M v c l</td>
</tr>
<tr>
<td>k P I o m N Y M o</td>
<td>c W h S m r w O</td>
<td>K L M N O P Q R S</td>
<td>V E N X d m w F o</td>
</tr>
<tr>
<td>p L q K r j s i t</td>
<td>K d i G Y C s x n</td>
<td>T U V X W Y Z @</td>
<td>W e n y G P Y f O</td>
</tr>
<tr>
<td>u H v G w F x E y</td>
<td>e V j R o t y N Z</td>
<td>g h i d e f a b c</td>
<td>x H Q Z g p z l R</td>
</tr>
<tr>
<td>z D @ C ? B n A Z</td>
<td>J F B U Q M I E A</td>
<td>j k l m n o p q r</td>
<td>@ h q A J S ? i r</td>
</tr>
</tbody>
</table>

Table.2. Plain Text and Cipher Text (Digits and Special Characters)

<table>
<thead>
<tr>
<th>Matrix-9</th>
<th>Matrix-10</th>
<th>Matrix-11</th>
<th>Matrix-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 (\mu) (\beta) (\delta)</td>
<td>8 9 (\sim) ! # $ \ldots &amp; % &amp; * ( ) + - _ ^ _ - | { }</td>
<td>8 9 (\sim) ! # $ \ldots &amp; % &amp; * ( ) + - _ ^ _ - | { }</td>
<td>8 9 (\sim) ! # $ \ldots &amp; % &amp; * ( ) + - _ ^ _ - | { }</td>
</tr>
<tr>
<td>(8 9 \sim ! # \ldots &amp; % &amp; * ( ) + - _ ^ _ - | { }</td>
<td>(8 9 \sim ! # \ldots &amp; % &amp; * ( ) + - _ ^ _ - | { }</td>
<td>(8 9 \sim ! # \ldots &amp; % &amp; * ( ) + - _ ^ _ - | { }</td>
<td>(8 9 \sim ! # \ldots &amp; % &amp; * ( ) + - _ ^ _ - | { }</td>
</tr>
</tbody>
</table>

4. DATA EMBEDDING

The image of size \((p \times q)\) can be represented in Eq.(1) as

\[
I = \left\{ x(i, j) \mid 0 \leq i \leq p, 0 \leq j \leq q \right\}
\]  

(1)

The secret message \(m\) of length \(l\) can be represented in Eq.(2) as

\[
M = \{ m(i) \mid 0 \leq i \leq l, m(i) \in \{0, 1\} \}
\]  

(2)

The encrypted message \(\overline{m}\) after using the extended substitution algorithm can be represented as

\[
\overline{M} = \overline{m(i)} \mid 0 \leq i \leq l, \overline{m(i)} \in \{0, 1\}
\]  

(3)

Further, the process of embedding message into an image can be explained as: Firstly, the carrier image is transformed into binary form. We took only red plane out of Red, Blue and Green to insert the secret message in color image. Each pixel value is then converted into 8 bits also cipher text of the secret message is binarised. In case of Two LSB Steganography Two bits of the binarised cipher text are inserted into the carrier image.

bits of the binarised cipher text are inserted into the carrier image in case of Three LSB Steganography. The process of inserting the data at LSB position is random and depends on variable \(x\).

Number of pixels needed to embed the available cipher can be calculated using following formula.

\[
\text{Number of Pixels Needed} = [n \times l]
\]  

(4)

where, \(n = 4, 3\) for Steganography at two LSBs and Steganography at three LSBs respectively. \(l\) is length of cipher.

The insertion of the data at LSB positions depends on the variable \(x\). The variable \(x\) takes the values 0, 1, 2, 3 for the both Steganography schemes. For Steganography at Two LSBs, if the value of \(x = 0\), eight bit binary cipher text is embedded at 6th and 7th bit locations of first four binary pixel value of the carrier image. If \(x = 1\) next eight bit cipher text is embedded at 7th and 8th bit locations at next four binary pixel value and in same fashion for \(x = 2, 3\).

For example, suppose alphabet ‘A’ is a cipher. Its ASCII value is 65. Its binary value is 01000001. First two bits from left are 01; second pair is 00, third 00 and fourth 01. If value of \(x = 2, \ldots\)
then to hide this eight bit data at every two LSBs of the binary pixel value, four subsequent pixels are required. Suppose we call those subsequent pixels as \( p, q, r, s \); then the first pair ‘01’ is inserted at 6\(^{th}\) and 8\(^{th}\) bit location of pixel \( p \). The second pair ‘00’ is inserted at 6\(^{th}\) and 8\(^{th}\) bit location of \( q \) and so on. Table.3 illustrates this concept.

Table.3. LSBs Selection for Two bit Steganography

<table>
<thead>
<tr>
<th>Cipher Text</th>
<th>Binary pixels</th>
<th>Variable ( x )</th>
<th>Insert</th>
<th>LSB positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ((01000001))</td>
<td>( p )</td>
<td>2</td>
<td>01</td>
<td>6(^{th}), 8(^{th})</td>
</tr>
<tr>
<td></td>
<td>( q )</td>
<td>2</td>
<td>00</td>
<td>6(^{th}), 8(^{th})</td>
</tr>
<tr>
<td></td>
<td>( r )</td>
<td>2</td>
<td>00</td>
<td>6(^{th}), 8(^{th})</td>
</tr>
<tr>
<td></td>
<td>( s )</td>
<td>2</td>
<td>01</td>
<td>6(^{th}), 8(^{th})</td>
</tr>
<tr>
<td>B ((01000010))</td>
<td>( t )</td>
<td>3</td>
<td>01</td>
<td>5(^{th}), 8(^{th})</td>
</tr>
<tr>
<td></td>
<td>( u )</td>
<td>3</td>
<td>10</td>
<td>5(^{th}), 8(^{th})</td>
</tr>
<tr>
<td></td>
<td>( v )</td>
<td>3</td>
<td>00</td>
<td>5(^{th}), 8(^{th})</td>
</tr>
<tr>
<td></td>
<td>( w )</td>
<td>3</td>
<td>10</td>
<td>5(^{th}), 8(^{th})</td>
</tr>
</tbody>
</table>

Data embedding at three LSBs can be explained with Table.4. It needs only three consequent pixels for embedding single cipher. For Steganography at three LSBs, If the value \( x = 2 \), six bits of binary cipher text are embedded at 7\(^{th}\), 8\(^{th}\), 6\(^{th}\) bit locations of first two pixels and remaining two binary bits of cipher are embedded at 7\(^{th}\), 8\(^{th}\) bit locations of the third pixel. If \( x = 3 \), the six bits of the new cipher are embedded at 8\(^{th}\), 7\(^{th}\), 6\(^{th}\) bit locations at next two pixels and remaining two bits of the cipher are embedded at 8\(^{th}\) and 7\(^{th}\) bit location of the next pixel. The other two possibilities \( x = 0, x = 1 \) are considered in the Table.4.

Table.4. LSBs Selection for Three bit Steganography

<table>
<thead>
<tr>
<th>Cipher Text</th>
<th>Binary pixels</th>
<th>Variable ( x )</th>
<th>Insert</th>
<th>LSB positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ((01100001))</td>
<td>( e )</td>
<td>0</td>
<td>011</td>
<td>6(^{th}), 8(^{th}), 7(^{th})</td>
</tr>
<tr>
<td></td>
<td>( f )</td>
<td>0</td>
<td>000</td>
<td>6(^{th}), 8(^{th}), 7(^{th})</td>
</tr>
<tr>
<td></td>
<td>( g )</td>
<td>0</td>
<td>01</td>
<td>6(^{th}), 8(^{th})</td>
</tr>
<tr>
<td>B ((01100010))</td>
<td>( h )</td>
<td>1</td>
<td>011</td>
<td>6(^{th}), 7(^{th}), 8(^{th})</td>
</tr>
<tr>
<td></td>
<td>( i )</td>
<td>1</td>
<td>000</td>
<td>6(^{th}), 7(^{th}), 8(^{th})</td>
</tr>
<tr>
<td></td>
<td>( j )</td>
<td>1</td>
<td>10</td>
<td>6(^{th}), 7(^{th})</td>
</tr>
</tbody>
</table>

5. THE PROPOSED ALGORITHM

The whole explanation of the implementation can be easily understood with the flowchart shown in Fig.2. It follows the steps like checking of carrier length for concealment, conversion of image into binary bits, plain text encryption using extended substitution algorithm, conversion of encrypted text into binaries and concealing these binaries of encrypted text into the carrier.

6. RESULTS

Verification of the results has been done for various image formats and the proposed method works fine. Fig.3 shows the images the images before embedding and after embedding the text message of 300 characters length. Whereas Fig.5 & Fig.6 show Length of Text Message verses PSNR graphs. It is observed from Fig.4, Fig.5 & Fig.6 that there is slight difference between the PSNR for Two-LSB and Three LSB Steganography. Also, the overall PSNR is more than 70 dB for both the Steganography schemes when the character length is considered upto 300. PSNR has been calculated using the formula from Eq.(5). In the equation, \( R \) is the maximum fluctuation in the input image data type. \( R \) is 255 for 8-bit unsigned integer data type that is the maximum possible value of a pixel.

\[
PSNR = 10 \log_{10} \left( \frac{R^2}{MSE} \right) \text{dB}
\]

The experimentation of proposed method has been done with the help of Graphical user interface and it provides convenience to the user. The PSNR, MSE are 73.4008 dB, 0.0029716 for the length of 300 characters. These results have been displayed in Fig.7.
Fig. 3(a-c). Images before embedding; Images (d-f) after embedding the text message.

Fig. 4. Length of Message versus PSNR for Image (a).

Fig. 5. Length of Message versus PSNR for Image (b).

Fig. 6. Length of Message versus PSNR for Image (c).

Fig. 7. Graphical user interface of Three LSB Steganography for 300 characters.
7. CONCLUSION

The secret communication system has two layered security levels. First level is through encryption of the text using Extended substitution algorithm and second one is through embedding the encrypted text into LSBs variably. The verification of both the Steganography schemes along with Extended Substitution Algorithm has been done and it is clear from the experimentation that inserting the data at three LSB positions does not change image parameters like PSNR, Mean, Standard deviation, Entropy in much extent. Therefore, it retains the image quality similar to two LSB scheme. This system is able to conceal almost all types of alphabets (small as well as capital), special characters and mathematical symbols. The variable \( x \) takes values as 0, 1, 2, 3. Embedding the cipher at LSBs is decided by variable \( x \). As the LSB in each pixel are not same but decided according to variable value, it is stronger approach and helps in minimizing the error. The proposed secret communication system can be used for military application where a soldier in the enemy territory wants to send the secret message to the head-quarters of his country.

ACKNOWLEDGEMENT

All the authors work under Pune University. They are thankful to the teachers and management of Dnyanganga College of Engineering, Sinhgad college of Engineering and RMD Sinhgad School of Engineering for permitting them to pursue their research work.

REFERENCES


