SPACE EFFICIENT IMAGE STEGANOGRAPHY: A NOVEL DATA HIDING SCHEME FOR TAMIL TEXT DOCUMENT

K. Manimozhi, S. Abiramasundari, V. Kalaichelvi, H. Manikandan and P. Meenakshi
Department of Computer Science Engineering, SASTRA Deemed University, Thanjavur, Tamil Nadu, India
E-Mail: kmanimozhi77@src.sastra.edu

ABSTRACT
Hiding secret information in a cover, which is of same type or different type is called Steganography. The main goal of the Steganography is that the sharing of secret information should not be shown to the intermediaries during communication. Several algorithms have been developed having different kinds of cover media, such as text, image, audio or video. We propose an approach to convert a secret message, which is available as Tamil text document into a gray-scale image that can be created of any size based on the size of the secret message by the sender, and that is transmitted over communication channel. At the hiding end, the characters in the tamil text document are encoded after encryption, then the image is scanned row by row, and in each non black pixel four characters are hidden. Only if the color of a pixel is not black, then the characters are hidden into it. At the extraction end, the characters are extracted after finding that the pixel is not in black color. This proposed method shows an agreeable experimental result with the cover image chosen. The efficiency of this proposed method is that the size of the stego image is lesser than the size of the secret message. Security can be increased by encrypting the secret text before embedding into the cover image.

Keywords: information security, image steganography, encryption, compression.

INTRODUCTION
Steganography is the form of security technique for ensuring the secret messages unread. The main objective of using steganography is to embed and hide the important information in a cover text/image. This can be done by various encoding techniques. However, the secret communication can be achieved through cryptography and steganography as well. Nowadays, communication through medium is more wide-ranging that necessitates hide secret information.

Steganography can be broadly categorized into linguistic steganography and technical steganography in which technical steganography is classified into digital images, video, audio and text steganography. Image steganography embeds the given information into a cover image. Even though many different image steganography algorithms have been developed, the breaking of secret messages still happens. Therefore, an urge of developing the effective algorithms is a great challenge for the developers. In connection to that, developers either derive a new algorithm or upgrade the existing one.

The secret messages can be made invisible by tattooing inside the cover image/text through steganography leaving without any trace of suspension [1-3]. Anyone suspect that the image could have any secret messages results in the failure of steganography technique used [4]. The tamil text document can also be used for hiding as cover text in text steganography [5].

The prisoner’s rescue problem illustrates the concept of advanced representation of steganography [6]. Tremendous approach of steganography methods have been used in images [2-8], video frames [9-10] and audio [11-12]. Since text steganography is always being a great challenge for the cryptanalysis [13].

Context free grammars of rewriting rules are used as a base for structuring the text in syntactical steganography which are syntactically accurate [14]. NICETEXT algorithm [15-16] is one such algorithm that is used to identify the underlying information in a cover file. It applies the part-of-speech for generating the words for a sentence. In some cases, the result of NICETEXT would yield meaningless sentences.

In lexical steganography, the secret message is hidden by the relevant words of natural language. It chooses a dictionary word much similar to the original word without making the word suspicious and it is applied to the sentence.

Information hiding can also be done by picking the first letter of the alternate text. This approach is also extended to selected words [17]. Irrelevant text can also be used to hide the secret message in a HTML file [20]. In line shifting method, a line is rotated to some degree α. Whereas in word shifting method the information is stored by altering the gap between the words and moving words [18-21].

The other methods include feature coding method, which enforces the method of altering the complete formation of the text for steganography [22], [23]. The white spaces in between the characters of text is used for hiding the information that are mainly applied for Indian languages [24-26] using the technique of Open Spaces method, feature coding method and dynamic programming method. The various symmetry of reflection and shapes of the characters in the line of text are also considered for hiding the secret data bits [27-29].

PROPOSED WORK
We propose an image steganographic technique for Tamil text document. The Tamil text comprises of vowels, consonants, vowel consonants and some punctuators that are listed. The text is encoded using the look up table given. We take a secret Tamil text document and a 32-bit PNG cover image as input. The prerequisite to hide the secret text into the cover image is that the cover
image must have at least one fourth of the text length pixels.

At the hiding end, once the text and suitable image is chosen, the text can be hidden into the image. The text is segmented into number of blocks based on the block size. The block size is n x n, where n must be an even number and it may range from two to 16. When the text is to be hidden, the suitable block size is found based on the text length, it is ensured that the text length is a multiple of chosen block size n x n. If not, then the excessive characters can be grouped into block of nearest even size. Padding also can be used if necessary.

Once blocks are ready, we apply folding technique block by block. One block is taken at a time as a table that is horizontally folded based on the midway. Horizontally folded table is then folded vertically. After both the folding are applied, the size of the block is reduced to n/2 x n/2.

After folding, each cell in the table will have four characters. We encode all the characters in the table with the help look up tables such as Table-1, Table-2, Table-3 and Table-4.

Once the text is encoded, it can be hidden into the cover image using the Hiding algorithm.

At the extraction end, the stego image is received and all the non black pixels are scanned and the codes are extracted, decoded, then the folding techniques are applied reversely to get the hidden secret message. The proposed mechanism is shown in Figure-1.

**Hiding algorithm**

**Input:** Secret tamil text and a cover image (32-bit PNG image)

**Steps:**
1. Find the suitable block size ‘n’ such that it is suitable for folding.
2. If the text length is not a multiple of block size, then round it with the help of padding.
3. Take one block of characters from the input tamil document and store it in n x n table.
4. Apply horizontal folding so that the table is resized to n/2 x n and each cell contains two characters.
5. Apply vertical folding so that the table is resized to n/2 x n/2 and each cell contains four characters.
6. Encode the characters in the table by using the look_up table.
7. Take the cover image.
8. Apply raster scanning onto the cover image.
9. When a non black pixel is found, replace its byte values with next available cell’s content.
10. Repeat the steps 8 and 9 until all the characters in the cells of the table are hidden.
11. Return the stego image.

**Output:** Stego image in which the secret tamil text is hidden

**Extraction algorithm**

The reverse process of the hiding algorithm is applied for extracting the secret tamil text from the stego image.

**Table-1. Code for Tamil vowels.**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>ஐ</td>
<td>੐</td>
<td>஧</td>
<td>ந</td>
<td>ண</td>
<td>ப</td>
<td>ம</td>
<td>ர</td>
<td>ல</td>
<td>ள</td>
<td>ழ</td>
<td>வ</td>
<td></td>
</tr>
</tbody>
</table>
Table-2. Code for Tamil vowel consonants.

<table>
<thead>
<tr>
<th>ச</th>
<th>ஞ</th>
<th>ட்</th>
<th>த்</th>
<th>ந்</th>
<th>ன்</th>
<th>ப்</th>
<th>ப்</th>
<th>ப்</th>
<th>ப்</th>
<th>ப்</th>
<th>ப்</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>49</td>
<td>67</td>
<td>85</td>
<td>103</td>
<td>121</td>
<td>139</td>
<td>157</td>
<td>175</td>
<td>193</td>
<td>211</td>
<td>229</td>
</tr>
</tbody>
</table>

Table-3. Code for Tamil consonants.

<table>
<thead>
<tr>
<th>ச</th>
<th>ஞ</th>
<th>ட்</th>
<th>த்</th>
<th>ந்</th>
<th>ன்</th>
<th>ப்</th>
<th>ப்</th>
<th>ப்</th>
<th>ப்</th>
<th>ப்</th>
<th>ப்</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

Table-4. Code for punctuators.

<table>
<thead>
<tr>
<th>247</th>
<th>248</th>
<th>249</th>
<th>250</th>
<th>251</th>
<th>252</th>
<th>253</th>
<th>254</th>
<th>255</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blank</td>
<td>space</td>
<td>,</td>
<td>!</td>
<td>?</td>
<td>&quot;</td>
<td>(</td>
<td>)</td>
<td>New line</td>
</tr>
</tbody>
</table>

EXPERIMENTAL RESULT

A sample Tamil text without any punctuators shown in Table-5 is taken and filled in a 16 x 16 block. Table-6 is the output of horizontal folding and Table-7 shows the output of vertical folding. At last the characters shown is Table-7 are encoded using the codes given in the Tables Table-1, Table-2, and Table-3. The encoded result is shown in Table-8. The encoded data is then hidden into the 32-bit PNG cover image given in Figure-1 and the stego image given in Figure-2 is created after hiding the encoded data.
Table-5. The secret Tamil text.

Table-6. Text after horizontal fold.

Table-7. Text after vertical fold.
PERFORMANCE ANALYSIS

An algorithm that is newly proposed must be analyzed for its performance. Any algorithm’s efficiency is analyzed by two important factors such as Running time and memory space. Since this algorithm is applied in the network communication, both time and space are significantly important. Time efficiency of the hiding procedure is linear under best case. The hiding procedure includes folding, encoding and embedding techniques. The folding technique uses only the secret text. When we take the text length as n, the running time of folding technique is $O(n)$. After folding, the characters are encoded using the look up table. The best case running time of the encoding is $O(n)$, as each search takes $O(1)$ with any efficient data structure. The best case running time of the Embedding procedure is also $O(n)$, so that the best case total running time of the hiding procedure is linear. In worst case, the encoding procedure takes $O(n^2)$, since the worst case running time of each search is $O(n)$, so that the worst case total running time of the hiding algorithm is $O(n^2)$.

Generally the Space Efficiency deals about the amount of space occupied by an algorithm for its completion. This algorithm is space efficient since it uses no additional memory. Since four characters are hidden into a non black pixel, the size of the stego image is smaller than the size of the text document. A cover image with more non black pixels can be chosen to hide a document with more characters. Thus the proposed algorithms are a space efficient image steganographic technique.

CONCLUSION AND FUTURE ENHANCEMENT

An approach to convert a secret message, which is available as Tamil text document into a gray-scale image that can be created of any size based on the size of the secret message by the sender, and that is transmitted over communication channel. At the hiding end, the characters in the Tamil text document are encrypted and encoded, then the image is scanned row by row, and in each non black pixel four characters are hidden. Only if the color of a pixel is not black, then the characters are hidden into it. At the extraction end, the characters are extracted after finding that the pixel is not in black color. This proposed method shows an agreeable experimental result with the cover image chosen. The efficiency of this proposed method is that the size of the stego image is lesser than the size of the secret message. Security can be increased by encrypting the secret text before embedding into the cover image. In future, the security of the secret information can be enhanced by adding an efficient encryption technique.

REFERENCES


Circuits and Systems for Video Technology. 11(12): 1237-1245.


