International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013

A Text based Steganography Technique with Indian Root

Souvik Roy\textsuperscript{a,*}, P.Venkateswaran\textsuperscript{b}

\textsuperscript{a}Jadavpur University, India
\textsuperscript{b}Jadavpur University, India

Abstract

Steganography is the art and science of covered writing. Modern digital steganography uses text, images, audio, video etc. as a cover medium. This paper presents a text based steganography technique based on the Vedic Numeric Code. Frequency of the letters in English alphabet in conjunction with Vedic Numeric Code are used for the steganography technique. No separate importance is given for vowels and consonants.

Keywords: Steganography; Text Steganography; Information Security; Vedic Numeric Code

1. Introduction

Steganography is the art of hiding of a message within another so that presence of hidden message is indistinguishable. The key concept behind steganography is that message to be transmitted is not detectable to the casual eye. This is also the advantage of steganography over cryptography. An unhidden encrypted message, no matter how unbreakable, raises suspicion but steganography is used not to keep other from knowing the hidden information, but it is to keep other from knowing the hidden information [1]. There many steganography methods which use images [2, 3], video [4], audio [5] as a cover media.

Text steganography uses text as a cover media for hiding message. Message can be hidden by shifting word and line [6, 7], in the open spaces [8], in word sequence [9]. Properties of a sentence such as number of words, number of characters, number of vowels, position of a vowel in a word are also used to hide secret message. The advantage of preferring text steganography over other steganographic technique is its smaller memory requirement and simpler
communication [10]. But due to lack of large scale redundancy of information in text file, in compared to other
medias, text steganography seems to be most difficult kind of steganography [11].

This paper presents a different approach to the English text based steganography with Indian root. In the propose
method, no properties of a sentence are used rather characteristics of the English language is used. This gives
flexibility and freedom from the point view of the sentence construction but it increases computational complexity.

2. Related work

In [1] by using the properties of a sentence and the presence of redundant feature code able characters in Indian
Languages, a message is hidden into an innocent cover file containing Indian texts. Using the existence of too many
points in Persian and Arabic phases, information is hidden in the Persian and Arabic texts [12]. Hindi letters and its
diacritics and numerical code are used in [13] for hiding message into Hindi text. Generating a random sequence of
characters or words, specific information can be hidden in sequence [9] but it often results in meaningless words or
sentence which is susceptible to raise suspicion. In method [14], some specific characters from certain words, placed
in certain sequence in sentence, are used as hiding place for secret message but the method is time consuming and
takes a lot of mental power. By placing punctuation signs in specific places, information is hidden [9] but
information hiding capacity of the method is low. Using synonym of certain words in a sentence, information can be
hidden [15]. By altering the features of a text information is hidden in text [16].

3. Suggested Algorithm

Some of the outstanding characteristics, inflexion, fixed word order and use of periphrases, of the English
language are used for the steganography technique. Inflexion means that it can indicate the relationship of the words
into a sentence with a minimum change of shape. In fixed order, the place of each word in a sentence decides it
relationship with the others. Periphrases are the different ways to express something.

Sri Bharati Krishna Tirthaji in [17] described a particular code called Vedic Numerical Code used in deciphering
Sanskrit text. The coding is based on the tongue position. For applying the Vedic code to the English alphabet,
frequency of letters in English vocabulary [18] is used as the basis of assigning numbers to the letters in the English
alphabet. Frequency of letters is shown in table 1. No discrimination is made for assigning coding number to vowels
and consonants as compared to [13].

Each letter in the alphabet is assigned a number in the range of 0 to 15 as shown in table 1.

3.1. Encoding

For the proposed text based steganography technique, each letter in the secret message is represented by its
ASCII code and obtained ASCII code is expressed in 8 bit binary number. The 8 bit binary number is then divided
into two 4 bit parts. Each 4 bit part, representing a number in the range 0 to 15, is then used to choose corresponding
suitable letters from table 1. A meaningful sentence is constructed by using letters obtained as the first letters of
suitable words. To give flexibility in sentence construction, articles, pronoun, preposition, adverb, was/were,
is/am/are, has/have/had, will/shall, would/should are not taken into consideration for encoding during sentence
construction.

3.2. Decoding

For decoding the cover message, first letter of each word is taken and represented by the corresponding 4 bit
binary number and these 4 bit binary numbers are then combined to obtain 8 bit binary numbers and corresponding
ASCII codes for letters in the secret message are then obtained from the 8 bit numbers. Finally the secret message is
recovered.
3.3. Table

**Table 1. Table of letter frequency and number assignment**

<table>
<thead>
<tr>
<th>Letter</th>
<th>Frequency of letter</th>
<th>Number assigned</th>
<th>Letter</th>
<th>Frequency of letter</th>
<th>Number assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>11.1607 %</td>
<td>15</td>
<td>M</td>
<td>3.0129 %</td>
<td>7</td>
</tr>
<tr>
<td>A</td>
<td>8.4966 %</td>
<td>14</td>
<td>H</td>
<td>3.0034 %</td>
<td>7</td>
</tr>
<tr>
<td>R</td>
<td>7.5809 %</td>
<td>13</td>
<td>G</td>
<td>2.4705 %</td>
<td>6</td>
</tr>
<tr>
<td>I</td>
<td>7.5448 %</td>
<td>13</td>
<td>B</td>
<td>2.0720 %</td>
<td>5</td>
</tr>
<tr>
<td>O</td>
<td>7.1635 %</td>
<td>12</td>
<td>F</td>
<td>1.8121 %</td>
<td>4</td>
</tr>
<tr>
<td>T</td>
<td>6.9509 %</td>
<td>11</td>
<td>Y</td>
<td>1.7779 %</td>
<td>4</td>
</tr>
<tr>
<td>N</td>
<td>6.6544 %</td>
<td>11</td>
<td>W</td>
<td>1.2899 %</td>
<td>3</td>
</tr>
<tr>
<td>S</td>
<td>5.7351 %</td>
<td>10</td>
<td>K</td>
<td>1.1016 %</td>
<td>3</td>
</tr>
<tr>
<td>L</td>
<td>5.4893 %</td>
<td>10</td>
<td>V</td>
<td>1.0074 %</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>4.5388 %</td>
<td>9</td>
<td>X</td>
<td>0.2902 %</td>
<td>2</td>
</tr>
<tr>
<td>U</td>
<td>3.6308 %</td>
<td>8</td>
<td>Z</td>
<td>0.2722 %</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>3.3844 %</td>
<td>8</td>
<td>J</td>
<td>0.1965 %</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>3.1671 %</td>
<td>7</td>
<td>Q</td>
<td>0.1962 %</td>
<td>0</td>
</tr>
</tbody>
</table>

4. Experimental Result

To implement the above text based steganography method, a secret message is considered. Suppose it is “text”. The message has 4 letters and by representing the letters in ASCII code and converting it into binary form, 32 bits are obtained. The result of encoding is shown below.

**Promod Yadav has gone to Bangalore for the marriage of his daughter to Pormash Yadav.**

Here first letter of each word is used for encoding. Encoding is not case sensitive. In another example, secret message is “Steganography”. Encoded result is shown below.

**Byomkesh Kumar Hazra, father of Govindo Byomkesh of Gobordanga Pally in Golpukur Jela, has given an application to George’s executive group for a medical history of Zygomyosis in Govindo, Joint Medical Group also urgently placed it to the committee.**

5. Information Security

A brief description about the application of above technique in information security during online shopping is given. The steganography technique in conjunction with visual cryptography technique [19] can provide a secure online shopping experience by ensuring customer data privacy in terms of providing only limited information both to the shopping service provider and payment gateway but still ensuring smooth fund transfer. An authorization code such as password can be hidden inside a cover text and a snapshot of the cover text is taken. Now from the snapshot two shares are generated by using (2,2) visual cryptography scheme. One of the shares is kept by the bank and other share is given to the customer. Now during fund transfer in online shopping, submission of its share by the
customer to the bank side ensures fund transfer from the customer account to the shopping service provider account without providing any credit or debit card information to the third party. Visual cryptography ensure the security of the authentication data during from customer side to certified authority (CA) side from where only the needed information is given to shopping service provider side and security of the rest of the information from the CA side to the bank side is provided by steganography technique in conjunction with traditional cryptography technique like Triple DES, RC4 etc. Fig.1 is the snapshot of the cover message, Fig.2 is generated share 1, Fig.3 is generated share 2 and Fig.4 is the reconstructed cover message.

6. Conclusion

As compared to [13], which gives separate importance to vowels and consonants, this technique does not give any separate importance to vowels and consonants providing better flexibility in hiding data in case of English language. As certain words are not included during encoding to create better cover medium, results in large number of words as compared to [13]. In the first example, to hide a 4 letter word, 8 words are required excluding the words that are added to provide flexibility in sentence construction. In the second example, 25 words are required for hiding a 13 letter word. So to hide a larger message, this technique requires large no of words and creates a complexity in sentence construction. However in banking and online shopping application where one required to send only account no and password, this disadvantage of the technique is used in its advantage by creating spam messages to hide one’s account no and password or any other personal information and avoid rising suspicion. In case of mobile banking, transaction can be made via SMS securely using the proposed technique.
In conjunction with other cryptography technique like visual cryptography, the proposed text steganography technique can provide a two layer of authentication and security system in physical and online banking as well as online shopping as shown in section information security.

References