New Steganography Method Using Litter Manipulations Frequency

Song Feng Lu, Omer Farooq and Haider Ali

ABSTRACT

Steganography the science tricks to hidden the information from the curious eyes when transmit it between sender and receiver, the Steganography digital algorithm split into four type text, image, audio and video it's very difficult to use the text cover to hide the information with regard to the other type because there is no disposition to add the information to the text without observed it. In this paper we propose the steganography text by using synonymic dictionary to hidden one bit of the secret information in every word, the technology worked by using word to hide the bit value of the secret represent (1,0,nothing) and every word from the cover text continue on bit value maybe (1,0,nothing), and make all the difficulty on the sender and can send the cover text by the E-mail or publish it in the newspaper or even make it in the propaganda and no one don’t have the stego-key can observe it’s have hidden important information inside this cover, and we can insert the stego-key in the same cover text to make it easily to the reserve to read the hidden information, all what the receiver want to know is the stego-key and the way to convert between the binary and hexadecimal.

INTRODUCTION

In the last years the information security is the first concern for a lot of searchers they want to find techniques and ideas to be sure that the transformation is safeness without any penetration from the snooper. The steganography is the science of how
to hide the digital information inside the electronic or un-digital cover without make any distribution or real effect in the cover. For this, since a lot of great application that working by the different way depend on the cover type, the important applied is to transfer the information safely between the two side without any doubt from the third side, this used in war between the country like the second great war they sent the secret letter to the army its look like normal messages for the enemy but it hold another meaning no one can understand it only how in the same army, or used by the spy to send the messages without any doubt in the written word and for another praxis is to protect the royalty and regalia for all the electronic file by use the Watermarking we can prove that you are the real owner for the text or photo or the video or the audio file the Watermarking means added special information to the cover without make any distribution to these added information on the original cover.

**Related Works on Text Steganography**

A number of techniques have been presented in detecting hidden data against text steganography. Based on the explorer that have been carried out completely, most of the text steganalysis methods are either based on ordered or pure statistical framework that procure from a detail and complex difficult analysis. Detection methods that used classifier to fined stegano-text have been defined in few research papers. Among all types, Support Vector Machine (SVM) is the most popular type that is used in text steganography to suppurate between natural text and stegano text. There are various attributes known to be used as the input for the SVM like the based on translation machine to translate the ngram attribute, the suitability role computed using TF-IDF weight (term frequency-inverse document frequency) and the syntactic structure. The other types are based on statistical following, which includes the variable of detection entropy (DE) and the variance (Var) of word experience, by using statistical variable of word place; and based on format-based features such as punctuation marks, and the distance of font attributes midway of two adjacent latter as the features that being draw forth from the text. There were also a few types that employed other than SVM classifier like the mechanism that is based on immune clone (IC) technique as a classifier to optimize the preference of the meta-features that taken away from the text attributes, and a classifier which is based on Perplexity mechanism to classify n-gram language model feature (also known as statistical language model (SLM)). A few mechanisms that are make based on statistical framework and have been suggested in previous works. The statistical analysis is hold out based on the text attributes or text features that have been known in the text. One mechanism is based on the statistical analysis to calculate the length of two consecutive spaces (neighbor difference). The diffident situation are based on the probability of distribution of the space latter and based on the allotment of first character of words to get the similarity coefficient as the key needle to different between natural text and stegano text.
The new Idea

All the text steganography depends on the added the space between the world or change the font or change the world position or use the character effect like in Arabic language or used the table to put the synonymic world value all this way to hidden the information. In our idea is to hide the information without make any clear effect to fell we hide information and without need to use the original cover text to compare and extraction the information, the old steganography text have a lot of problem like its un safe and must have the original cover text for the two side to find the change and extraction the information and in our idea we encroachment this problem.

The Steganography Text Elements:

For elements must be available to convert to the stego-Text this elements is The Secret message, Cover text, Secret character and word long.

Secret Message

We must convert the secret message to the binary and we have two kinds of this converter the first one support (5) bit but use only the English alphabetic without the number or the special character as in the table below, the second support (8) bit its include all the character, if the secret message contain number we use the (8) bit if it’s just character without number we use the (5) bit to make the cover text take more of secret information Table beside.

<table>
<thead>
<tr>
<th>Character</th>
<th>Binary 5-bit</th>
<th>Binary 8-bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-Z</td>
<td>00000</td>
<td>0000000000</td>
</tr>
<tr>
<td>a-z</td>
<td>00001</td>
<td>0000000001</td>
</tr>
<tr>
<td>0-9</td>
<td>00010</td>
<td>0000000010</td>
</tr>
<tr>
<td>!-.</td>
<td>00011</td>
<td>0000000011</td>
</tr>
</tbody>
</table>

The title and author information at the top of the first page are also in 10pt font, and span both columns. Use bold typeface and capital letters for the title, and ensure that the title does not exceed 2 lines. Leave one blank line between the title and the list of authors. Indicate the affiliation of each author using superscript numbers.
Secret Character

After we scan the text cover and find the frequency for every character and how many times it's repeated in the text cover and then we will decide what secret character we will use as stego-key, from [fig-1] we concluded the character (E,e) have the higher repeated in the cover text, and character (Z,z) is the lower repeated in the cover text. If we want to choose the secret character we must find the one how don’t change the balance of the cover file, for example if we take the character (E,e) we must convert every word don’t have this character in the synonym dictionary with one have this character to hide value (1) and the adverse to hide the value (0) so in this case we will make the character (E,e) not the higher repeated in the cover text and we will make it unstable, so we must choose one character have middle repeating like (H,I,N,O,R,S) to make the cover text stable.

Cover Text

In order to maintain the camouflage of the system the input cover-texts must be composed using different writing styles. Cover-texts can also be generated using natural language generation (NLG) systems. There are several fully implemented natural language generation (NLG) systems freely available for research purposes. Their main function is constructing natural language output from non-linguistic information representations according to some communication specifications.

Word Long

We must ignore the words like (it, is, are, the, they, our,….. etc.) and just use the word how contain 5 character and more.

Formulation

For the purpose of studying the natural distribution of secret letter frequencies of cover-text blocks, the probability density function of the classical cell occupancy problem is used to express the secret letter frequency distribution. Assuming that the cover-text consists of infinite number of words, the probability of words having (n) secret letters can be approximated by Poisson probability density function:

\[ P = \frac{e^{-n} n^n}{n!} \]

- \( P \) is the probability
- \( n \) is the number of word in the cover text
- \( e \) is Euler’s number (\( e = 2.71828…. \)).
\[ P(n) = \frac{\lambda^n e^{-\lambda}}{n!} \]

\( \lambda \) is the average number of events per interval of word change.

**Experimental Result**

We will try to hide (5) character in text cover the secret message is “start” and the text cover is: When you really want something, it is natural to ask the lord of the universe.

The secret message “start” (5 character) \((5*5\text{bit} = 25\text{bit})\) we must hidden if we use the (5) bit converter, \((5*8\text{bit} = 40\text{bit})\) we must hidden if we use the (8) bit converter, in this case we don’t have number we can use (5)bit converter.

**Sample Text**

In this case we take the character (I,i) , if the chosen for the word size was take as four letter minimum in every word, the text would divided in to seven words.

When1 you really2 want3 something4, it is natural5 to ask the lord6 of the universe7.

That’s mean we can hide (7) bit in this cover text. Assume that the chosen secret letter frequency function has the following simplest form: where \(L_1\) is the frequency of the chosen secret letter (i). If the one's priest value was 1 and the zero's priest value was 0, it would be possible to embed seven binary digits in this text. For example, to embed the bit string (1111111), the stego-text would be:

Albeit(L1 =1) you indeed(L2 =1) wish(L3 =1) something(L4 =1), it is typical(L5 =1) to ask the king(L6 =1) of the universe(L7 =1).

To embed the bit string 0000000, the stego-text would be:

When(L1 =0) you really(L2 =0) want(L3 =0) matter(L4 =0), it is natural(L5 =0) to ask the lord(L6 =0) of the cosmos(L7 =0).

**CONCLUSIONS**

During design, implementation and analysis, many facts and aspects were found. Some of these aspects were taken into consideration in redesigning the system's adjacent stages. The following are some of the considered and concluded system-related characteristics:

1. The size of stego-key is very small so that it may easily be transmitted through trusted communication channels.
2. Best quality of camouflage can be achieved if the synonym substitution stage is appropriately designed since there is always a choice of not substituting words that might decrease the level of camouflage quality.
(3). The security level can be considered high since there is no theoretical limit for the word size.

(4). It is necessary to encode the stenographic data before embedding to keep the system security level high.

(5). The embedding capacity is limited by the nature of the host language and the required levels of security and camouflage quality.

(6). Using one secret letter as a stego-key element, for the English language, is possible.

ACKNOWLEDGEMENTS

This work is supported by the Applied Basic Research Program of Wuhan Science and Technology Bureau under Grant No. 20160101010003 and the Natural Science Foundation of Hubei Province Grant No. 2016CFB541

REFERENCES


