Presentation a new method for Steganography in Persian text of an electronic document

Somayeh Behjati ARDAKANI¹, Ali Mohammad LATIF²*, Kamal MIRZAIE³

¹Department of Computer Engineering, Yazd Branch, Islamic Azad University, Yazd, Iran
²Department of Electrical and Computer Engineering, Yazd University, Yazd, Iran
³Department of Computer Engineering, Maybod Branch, Islamic Azad University, Maybod, Iran

Received: 22.03.2015; Accepted: 29.05.2015

Abstract. In this study, new method is presented for text steganography based on a feature orthography of Persian language. The proposed method have make embed a secret information in the text by changing font from end of the sentence. Point recognize and change its font according to secret information is the main stages of this method. The results shows that proposed method have an appropriate performance and efficiency.

Keywords: Hiding the information, Persian text steganography, word document, change the font

1. INTRODUCTION

Recently, because of increasing world communication and devising various communication channels such as the Internet, satellite communication and telecommunications, information is easily available in use a wide range of people around the world. Some years ago, main and prestigious international journals articles, papers, confidential data or official correspondence was physically and paper, there were using by a limited range of users. So illegal use or misuse or forgery of these kind of documents were difficult and limited because of inaccessibility and also need for special techniques.

Nowadays, entrance of digital technology to administrative system around the world and people’s life and the World Wide Web and other channels of communication, it has provided easier access to the information. Administrative systems have been developed without paper that preparation and offer many documents and information in the digital data form. These data can be in different formats such as text, electronic books, images, animations, videos, databases, software, and computer games. Digital nature of the data necessitated that create, modify, update, amendments, share, save and distribution of information are easier than ever before. Therefore access and free communicating information, equal of that it might be going higher possibility of data misuse. Therefore, the importance and necessity of hiding the information contained is clear in this condition.

Penhan negary is the Persian meaning of steganography which is essentially a Greek word and consists of two words, Steganos means "hide" and Graphy means "writing". So steganography is used for hiding data in a media as a cover.

However, in steganography and watermarking, code is hide into the text, difference between steganography and watermarking is, in steganography code is very important and content is not seem very important but in watermarking, it is important that unauthorized copying does not on content and watermark is not important.

* Corresponding author. Email: ALatif@yazd.ac.ir

Special Issue: Technological Advances of Engineering Sciences

http://dergi.cumhuriyet.edu.tr/ojs/index.php/fenbilimleri ©2015 Faculty of Science, Cumhuriyet University
Some cases are important in a steganography method:

1. **Transparency:** System transparency suggests that host before and after embed and arranging should not have significant difference in information because purpose is make non-sense of convey the information and in fact, security of a system are steganography in transparency issue.

2. **Resistance:** System Resistance is the meaning of hidden information will be have necessary resistance against intentional and unintentional changes.

3. **Capacity:** amount of data that can be hidden in the host that is capacity. Algorithm have higher performance and efficiency with more capacity.

These three features are related to each other, that changes of one of them have effect on two other features.

Methods that have been proposed for text steganography base on components of a shape, we can say like this: ShirAli were used of dots method to text steganography in 2006 [1]. In this method, information first pressed which should be hidden. Then point’s first-letter of the text can be found. Hidden information that we read bit to bit, there are pressed information and form of zero and one bits. If the bit was zero, a character remain unchanged which we selected on text for steganography. But if it was one, point of the character is moving some upward. Sample of dots method shape can be seen in Figure 1.

![Figure 1. Vertical displacement of dot in dots method [1].](image)

ShirAli were used of steganography method by Unicode features for hide of text [2]. He found in the research of Unicode standard that in Unicode standard for Arabic letters, there are two storage code. A code is a representative of word and it has not depending on the letter position of the in the word. Another type of code has depends on the letter position in the word and there is four types and for each type of letters, there is a code. Code of representative letter in each word is storing and saving as a steganography of zero bit. Code form of letter in word is storing and saving as a steganography of one bit [2].

ShirAli were used improved method of La as a steganography [3]. He research in Arabic and Persian text and found that “\( \text{ي} \)” word is a character with “FEFB” code in Unicode Standard. We can show “\( \text{ي} \)” word another way, that are shown by (\( \text{ل} \)+space character between letters and hexadecimal code of 0640 + 1 letter) that their combination is as shown “\( \text{لا} \)”. We used steganography of a bit from (La) “\( \text{ي} \)” character and for zero bit from (La) “\( \text{لا} \)” character. This method has some problems such as increasing file size and because of gap and space between (Lam) “\( \text{ل} \)” and “\( \text{ا} \)” is become wider and make abnormal appearance of the text. In improved method instead of entering space code between (Lam)”\( \text{ل} \)” and “\( \text{ا} \)” were used form’s code of them and word is shown like (La) “\( \text{ي} \)” [3]. Sample of improved method shape of (La) “\( \text{ي} \)” can be seen in Figure 2.
Presentation a new method for Steganography in Persian text of an electronic document

<table>
<thead>
<tr>
<th>Watermarking bits</th>
<th>1001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cover-text</td>
<td>کلا لازم است قبل از وقوع بلاهای طبیعی اقدامات ایمنی لازم انجام شود.</td>
</tr>
<tr>
<td>Original “LA” method</td>
<td>کلا لازم است قبل از وقوع بلاهای طبیعی اقدامات ایمنی لازم انجام شود.</td>
</tr>
<tr>
<td>Our Improved method</td>
<td>کلا لازم است قبل از وقوع بلاهای طبیعی اقدامات ایمنی لازم انجام شود.</td>
</tr>
</tbody>
</table>

Figure 2. Sample of improved method shape of (La) “ی” [3].

In another study ShirAli were used of symbolic distance method between words as a text steganography[4]. He describes his method that in Persian and Arabic, furthermore that there is a normal space between words, in some words like (rafteh-am) "رفته ام" there is a little gap and space between two parts of one word. This space is known as a symbolic space. Normal space in Unicode system has hexadecimal code of 0020 but symbolic space that also called “zwnj”, it has 200c code. Hiding information in this method is that if there is a symbolic space in a word, one of two forms is selected depending on the information that should be hidden in the text. For hiding a bit. There is normal space after symbolic space in a word, for hiding zero bits, symbolic space does not changed.

Davazrani was used of letter’s form to steganography in 2009 [5]. This method used on shapes of letter and there are four letters of (Re-Ze-Zhe-Waw) (ر ز ژ و). These words have special slope in their forms and we can used of this slope to steganography. As between four letters, three of them are in Arabic alphabets. It can also use of Arabic texts by this method except for Persian text. Three letters (Re-Ze-Zhe) (ر ز ژ) are different just in number of dots. So in this method, a parameter that is considered to slope of these three letters, separated from slope parameter of (Waw) (و) letter.

Algorithm is uses four parameters that first two parameters related to slope of (Re-Ze-Zhe )، (ر ز ژ) letters and (و) shapes of letter into main text and two other parameters related to changed slope of (ر ز ژ) letters and (Waw) (و) letter is to text steganography shape.

Steganography method is beginning with slope letters (letters mentioned above) from shape into main text. it has not any changes in slope if your bit to steganography will zero. But if it has one bit, slope of letter will- change. This method will continues when all of the information can become secret. In Figure 3, letters slope of {ر ، و} is defined and in Figure 4, sample shape can be seen by this method.

According to 3-A figure, slope of {ر} letter is described by the following equation:

\[
\begin{align*}
\mathit{m}_r & = \tan(\hat{R}) = \\
& = \frac{L_r}{W_r}
\end{align*}
\]

Equation (1)

In this equation, \(m_r\) is slope \(L_r\) and \(W_r\) are Vertical length and horizontal width of {ر}, respectively, \(R\) angle is a range from connected line between beginning and end points of {ر} letter relative to horizontal axis. By Fig. 3b, we have been obtaining similar relationship to slope of {و} letter. Eq.

\[
\begin{align*}
\mathit{m}_v & = \tan(\hat{V}) = \\
& = \frac{L_v}{W_v}
\end{align*}
\]

Equation (2)
In Eq.(2), \( m_v \), \( L_v \), \( W_v \) and \( \hat{V} \) are slope, vertical length and horizontal width and connected line between beginning and end of (Waw) \( \{\text{j}\} \) letter relative to horizontal axis.

**Figure 3.** Definition of slope for \( \{\text{j}\} \) letters in letters form method [5].
Presentation a new method for Steganography in Persian text of an electronic document

Figure 4. Compare shape of main text and shape of text steganography [5].

Alnazer were used of stretching character method to text steganography in 2009 [6]. There is a character with hexadecimal code of 0640 in Unicode standard. This character as a extra character is used only for Structural purposes and do not used in every situation and just used in a space where connecting Arabic letters. In other words, do not used after the last letters and before the first letters of words. When this character is placed between two connecting letters, there is a little stretch between two letters. For hiding a bit range for example "110 010" value, it begins and starts from the least significant. If bit was zero, in first letter without a point (dot) after that there will a possibility of adding stretch character, it makes bit hidden by adding stretch character. If bit was one, in first letter that have a point (dot) after that there will a possibility of adding stretch character, it makes bit hidden. In other words it makes hiding zero bit by adding stretch character after non points letters and it makes hiding bit one by adding stretch character after dots letter. Furthermore, we can added stretch character before letters [6]. Figures 5 and 6, a work sample of Alnazer shows by adding stretch character before and after the letters.

<table>
<thead>
<tr>
<th>Watermarking bits</th>
<th>110010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cover-text</td>
<td>من حسن اسلام المرء تركه مالا يعنه</td>
</tr>
<tr>
<td>Output text</td>
<td>من حسن اسلام المرء تركه مالا يعنه</td>
</tr>
</tbody>
</table>

Figure 5. Steganography method by adding stretch character after the letters [7].

<table>
<thead>
<tr>
<th>Watermarking bits</th>
<th>110010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cover-text</td>
<td>من حسن اسلام المرء تركه مالا يعنه</td>
</tr>
<tr>
<td>Output text</td>
<td>من حسن اسلام المرء تركه مالا يعنه</td>
</tr>
</tbody>
</table>

Figure 6. Steganography method by adding stretch character before the letters [7].

ShirAli were utilizing similar letters with different codes to text Steganography in 2010 [8]. It can do this due to the existence of two different codes for “Ya” «ی» and “Kaf” «ک» characters in Persian and Arabic languages. For example, in the Persian language “Ya” «ی» is shown «ی» and in Arabic language “Ya” «ی» is shown «ي» and Kaf” «ک» is shown «ک» in Persian language and Kaf” «ک» is shown "ک" in Arabic language. He was used them to Steganography that If the
character is Persian «ک» or «ی», it means that the bit 0 is hidden in the text. If the character is Arabic «ق» or «ي»), it means that the bit 1 is hidden in the text.

Yazdani was used of letters shape to Steganography in 2013 [9]. This method is applied based on letters shape by using of four letters (h, kh, g, ch) (چ-ج-خ-ح). In this method, if hide message bit was one, these letters have been changed in base line. if hide information bit was one, these letters have not been changed in base line. Figure 7 shows an example of shape method.Figure 7-A shows چ letter situation without changing in base line. Figure 7-B shows چ letter situation with changing in base line.

Unlike the English and Chinese texts, there are not much special technicus for Persian text Steganography based on unique features orthography of Persian language. The main purpose of this study is proposed of a new method for Persian text Steganography in file document of their text.

2. THE PROPOSED METHOD

The study of Persian letters in a word file and check the font letters, it was resulted that selection of letters that are not connected to characters before and after. A problem with changing font letters in a word file is that letter of font changed is separated from before and after letter and it has contrast with hiding of information. Because separate of letters can be seen in text. For example if we write word of «گل» by different fonts, «گ» and «ل» are separated and meaning of «گل» word will be completely change. So letters and symptoms of Persian language, end point was chosen that change the font on it, there was no discernable change. When end dots of sentence in the text was found and identified, algorithm compared end dots of sentence in the text with number of hidden information bits. If the number of bits hidden information is less than the number of dots, it changes fonts of dots equal to the number hidden information bit that has number of one and it does not changes fonts of dots equal to the number hidden information bit that has number of zero and dots that are more than equal to the number hidden information bit, it is equal to hidden information 0 bit and its font does not change.
Presentation a new method for Steganography in Persian text of an electronic document

Figure 8. Shows main and original text and text Steganography with set of hide information (1101).

A) Main text

In this figure that is a part of the word electronic document there are 4 dots in end of sentence. In Figure 7-B, font of end dot changes in first, second and fourth sentences. Because first second and fourth bit of hide information are one but end font of third sentence does not change because third bit of hide information is zero.

To extract of hide information in every space that equivalent of font of changed dot in hide information bit is number of one and otherwise, it is zero and last hidden information is extracted. In this method, transparency that is one of Steganography standards, it is 100 percent. Because there is not any change in appearance in the text. In terms of resistance, this method is resistant against of attacks resize text attacks, change of text color, font changes (italics and bold and underlined), insert a word or phrase in the text, but it has not resistant against attacks by changing the font, copying and cutting attack.
3. CONCLUSION

In this paper, a new method for steganography in Persian texts is presented. In this method, information of hide message are prepared and arranged by using font changing feature of end dots of sentence in a word file. Transparency of this method is 100 percent. Some benefits of this method are as follows: in all languages can be used. There is not increase in the volume of text files. In Unicode character do not change. This method is resistant against of attacks resize text attacks, change of text color, font changes (italics and bold and underlined), insert a word or phrase in the text, but it has not resistant against attacks by changing the font, copying and cutting attack.

REFERENCES