Journal of Computer Science 9 (7): 898-904, 2013
ISSN: 1549-3636

doi:10.3844/jcssp.2013.898.904 Published Online 9 (7) 2013 (http://www.thescipub.com/jcs.toc)

Text Steganography Based on Font Type in MS-Word Documents

[1]Wesam Bhaya, [2]Abdul Monem Rahma and [3]Dhamyaa AL-Nasrawi

[1]Department of Information Networks, Information Technology Collage, Babylon University, Babil, Iraq
[2]Department of Computer Science, University of Technology, Baghdad, Iraq
[3]Department of Computer Science, Science Collage, Kerbala University, Kerbala, Iraq

Received 2013-01-03, Revised 2013-04-22; Accepted 2013-06-22

ABSTRACT

With the rapid development of Internet, safe covert communications in the network environment become an essential research direction. Steganography is a significant means that secret information is embedded into cover data imperceptibly for transmission, so that information cannot be easily aware by others. Text Steganography is low in redundancy and related to natural language rules these lead to limit manipulation of text, so they are both great challenges to conceal message in text properly and to detect such concealment. This study proposes a novel text steganography method which takes into account the Font Types. This new method depends on the Similarity of English Font Types; we called it (SEFT) technique. It works by replace font by more similar fonts. The secret message was encoded and embedded as similar fonts in capital Letters of cover document. Proposed text steganography method can works in different cover documents of different font types. The size of cover and stego document was increased about 0.766% from original size. The capacity of this method is very high and the secret message was inconspicuous to an adversary.

Keywords: Text Steganography, Similarity of English Font Types

1. INTRODUCTION

Steganography is the ancient art and young science of hidden communication. A broad definition of the subject includes all endeavours to communicate in such a way that the existence of the message cannot be detected. Unlike cryptography, which merely ensures the confidentiality of the message content, steganography adds another layer of secrecy by keeping confidential even the fact that secret communication takes place. The corresponding protection goal is called undetectability (Böhme, 2010).

Earlier information hiding methods merely embed payload (external information) into a cover (e.g., text document, image and audio) and in recent years, specialized data hiding methods are proposed to serve specific purposes. For instance, in steganography, the cover content is carefully manipulated to encode payload while aiming to conceal the very existence of the encoded information (Por *et al*., 2012).

Steganography literally means "covered writing" and is the art of hiding the very existence of a message. A message is the information to be hidden, anything that can be embedded into a bit stream. Together the cover carrier and the embedded message create a stegano-carrier. Hiding information may require a stegano key which is additional secret information, such as a password, required for embedding the information (Khandekar and Dixit, 2012).

Texts are used of a wide range, as numerous text materials are transported on the network every day. However, the studies on text steganography is relatively backward compared to those mainstream hiding methods that use images, audios and videos as cover data, which is due to the lack of redundancy in text (Liu *et al*., 2009)

In this study, a text steganography method used in MS-Word documents is proposed, which depending on

Corresponding Author: Wesam Bhaya, Department of Information Networks, Information Technology Collage, Babylon University, Babil, Iraq

the Similarity of English Font Types. Rest of the study is organized as follows. Section 2 introduces a related works, Section 3 presents materials and methods and Section 4 demonstrated the experimental results. Finally conclusions are provided in Section 5.

1.1. Related Works

Elkamchouchi and Negm (2003) proposed algorithm to apply the principle of walernmrking for hiding English information in Arabic text. Hassan and Shirali-Shahreza (2008) proposed method for hiding information in Persian and Arabic Unicode texts Zhong *et al*. (2007) proposes a steganography technique for hiding data in a kind of PDF texts. Por and Delina (2008) propose a new approach for information hiding using inter-word spacing and inter-paragraph spacing as a hybrid method Por and Delina (2008). Shirali-Shahreza (2008) proposes a Text Steganography by Changing Words Spelling. In this method the US and UK spellings of words substituted in order to hide data in an English text (Shirali-Shahreza, 2008). Khairullah (2009) proposes a new approach for steganography in Microsoft Word documents, by setting any foreground color for invisible characters such as the space or the carriage return is not reflected or viewed in the document. Liu *et al*. (2009) propose algorithm to be used in online chat. Shakir *et al*. (2010) develop a text steganography by using the diacritics-Harakat-of Arabic language as a covered medium to hide the Chinese stroke text. Yang *et al*. (2011) propose a new steganography proposed in MS Excel Document using text rotation technique. Text hiding in mobile phone simple message service using fonts was proposed by Bhaya (2011) which suggests a method of hiding the information (0,1) in cover SMS message by changing the fonts of each character using two fonts of mobile devices. Moraldo (2012) introduces a text steganography method based on Markov chains together with a reference implementation. This method allows for information hiding in texts that are automatically generated following a given Markov model (Moraldo, 2012).

2. MATERIALS AND METHODS

In this section, we explain proposed method in detail with encoding, embedding and extraction procedures. In addition, overview of fonts was presented.

2.1. Fonts Overview

Microsoft Word is popular word processing software which comes with Microsoft Office package. One of the reasons behind its popularity is huge number of text formatting features. One of these features is font format which have advantages of great capacity, good imperceptibility and wide application range (Khairullah, 2009).

A font is a graphic design that is applied to a collection of numbers, symbols and characters. A font describes a certain typeface, together with other qualities such as size, spacing and pitch. Fonts are used to display text on the screen and to print text. Fonts have font styles such as italic, bold and bold italic.a

MS-Word provides a number of standard fonts including Times new Roman, Courier New, Arial and many others which will spice up your document content. Some fonts can be the same size, but look bigger, due to the x-height. The x-height is literally the height of the small letter x in the font family. Different fonts have different x-heights and as a result, some fonts look larger than others, even though they are the same point size. The illustration in **Fig. 1** shows how the font size and x-heights are measured.

Some newer font families, such as Tahoma and Verdana, have been designed with large x-heights. That means different font families that are all the same point size, some look bigger, however, because of their larger x-height (Weinschenk, 2011).

2.2. Proposed Method (SEFT)

This study based on Text documents which are more prevalent and indispensable form of information nowadays and always be used as a cover medium. Most text steganography are based on the format TXT, MS Word, PDF, PPT.

Proposed method introduces a new method for writing hidden messages in text of document file format (which lack of redundancy compared to images or audio) called (Similar English Font Types, SEFT, Technique) use the most similarity types of English fonts in hiding message by changing the font to another.

In general, any type of font has many of types similar to its fronts. This property is the basic of this study.

2.3. Steps of Proposed Method

In this section, we describe the proposed method in detail, which has been implemented in C#.net language. It essentially consists of four main components:

- Create similar font array
- Create code Table
- Embedding process
- Extracting process

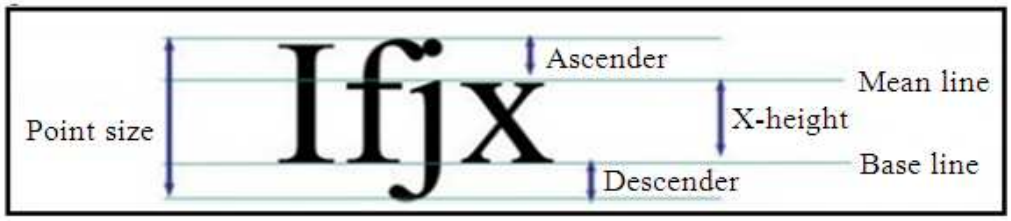

Fig. 1. Font size and x-height measures

2.4. Create Similar Font Array

This is the most important component of the method. Begin by determine the type of document font and then find the more similar types of it. In this study, assumed (15) type of cover document fonts which are more usable and prevalent in text documents (TXT, MS Word, PDF, PPT). **Table 1** explains the cover document fonts and their similars; three similars will be used for each type.

2.5. Create Code Table

The coding of each symbol in secret message represented by three types of fonts, thus, 27 characters (English alphabets with space) can be hidden in 3 letters of cover using 3 different fonts, for example: similar font array of Century font is:

Century = {Century751 BT,CenturyOldStyle,CenturyExpdBT}.

As we will see, if the code of current symbol is (1, 1, 1), then we will use (first similar, first similar, first similar) fonts from similar font array. Also, (1, 2, 2) means (first similar, second similar, second similar). (3, 1, 1) mean (third similar, first similar, first similar) and so on. The begin of message start from first capital letter in document and the end of message represented by code (0, 0, 0), which means the original document font. **Table 2** shows the Code Table used for coding process.

2.6. Embedding Process

In this study, secret message was embedded in Capital letters only of cover document, because the capital letters different in pattern from small English alphabet letters. Embedding process consist of three steps. The first step is determining cover document font to retrieve its similar fonts array. In the second step, scan cover document to find English capital letters, as we saw, need three capitals letters to hide one symbol. Finally, in third step, change the font type of first three capitals letters by similar fonts depending on code. The following procedures explain these processes.

2.7. Embedding Process

- Open cover document, find its type of font
- Scan cover document to find capitals English letters,
- Compute number of capitals English letters to check the capability of embedding
- For each symbol in secret message
- Retrieve its code
- Change font type of three capitals letters by similar font array according to its code

2.8. Extracting Process

Each three capitals letters; determine the code of one hiding symbol. The steps below show the extract process.

2.9. Extracting Process

- Open Stego document
- For each three capitals letters
- Determine the code
- If the code is (0, 0, 0), then the end of secret
- message was reached
- Else, find corresponding secret symbol, using code table

2.10. Explain of Proposed Technique by Example

More details can be found in this section of implementation the software. The corresponding GUI for the proposed SEFT technique was shown in **Fig. 2**.

The following block diagram, **Fig. 3**, explains the GUI operations performed by sender to implement hiding process. The hider chooses cover document and inputs secret message. The system will get font type and check the capability of hiding in selected cover file (by compute the number of capitals letters in cover file with input secrets characters). Finally the system coding the secret letter and hides characters.

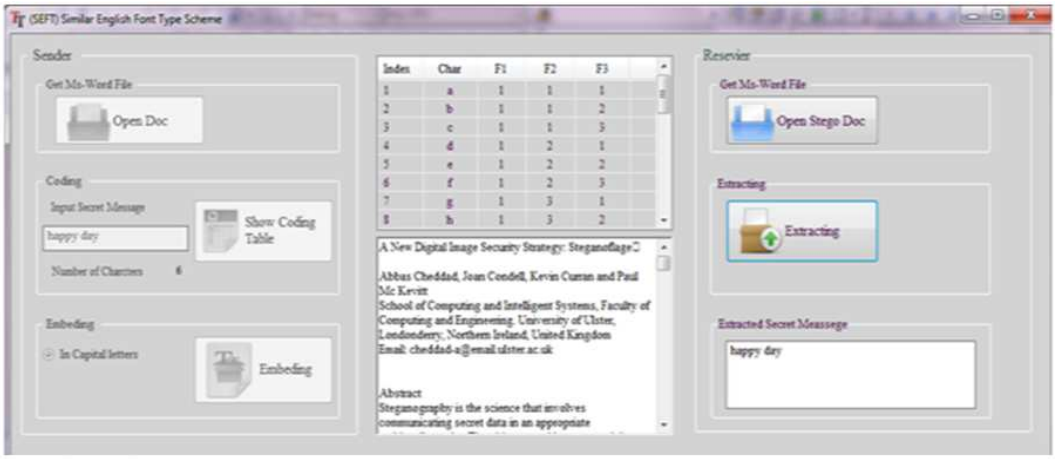

Fig. 2. GUI of SEFT technique

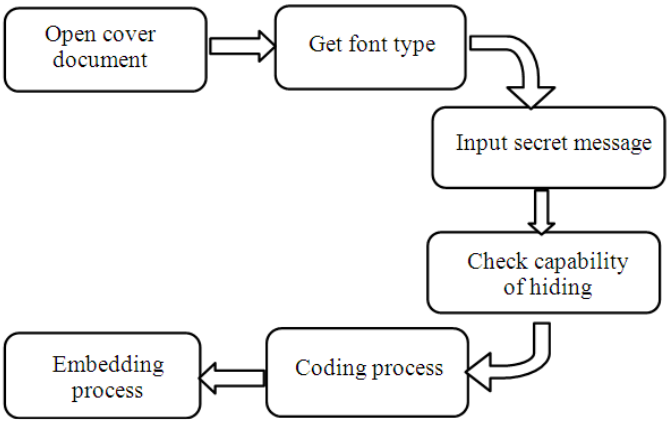

Fig. 3. Hiding operations

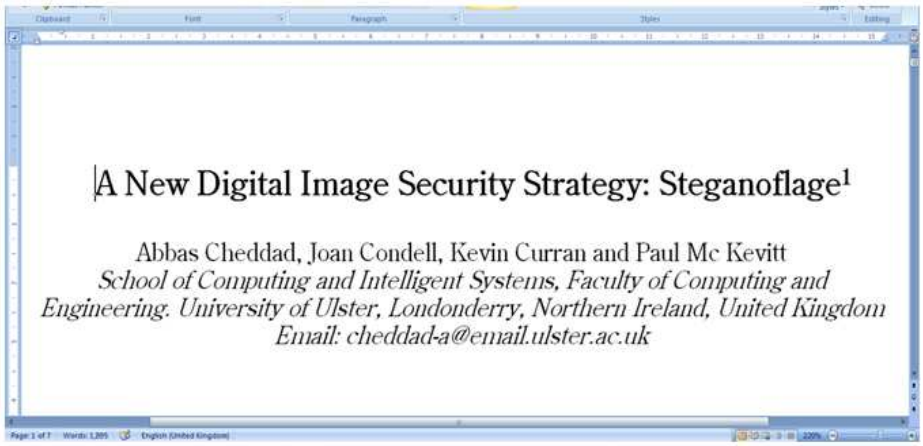

Fig. 4. Cover document

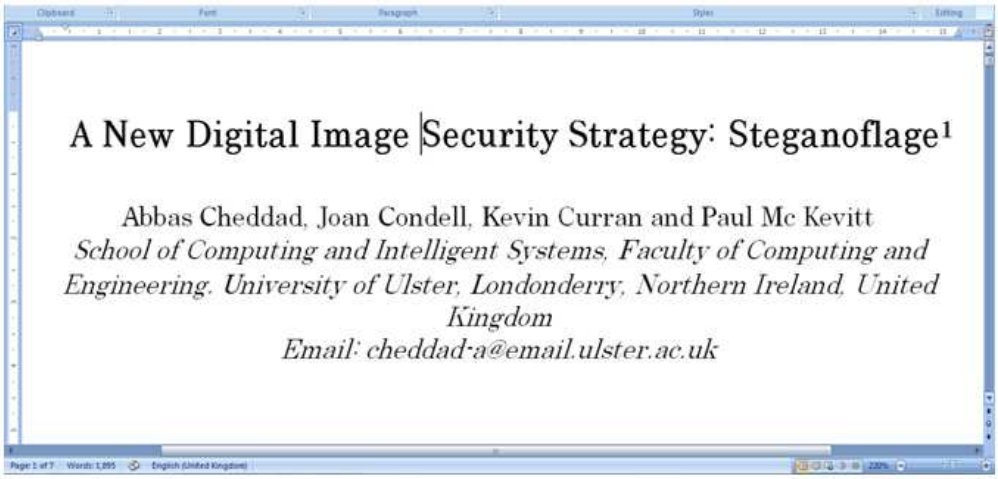

Fig. 5. Stego document

Table 1. Cover document font and their similar

Index	----------Font name----------		----------Similarity----------	
1	Arial	Arial Unicode MS	Geo_Arial	Microsoft Sans Serif
2	Book Antiqua	Palatino Linotype	Antiqua	Caudex
3	Candara	Khmer UI	Ebrima	Microsoft New Tai Lue
4	Century	Century751 BT	CenturyOldStyle	CenturyExpd BT
5	Calibri	Gisha	Leelawadee	Liberation Serif
6	Cambria	Proforma	EideticNeoRegular	Liberation Serif
7	Comic Sans	SF Toontime	Komika Text	SF Arch Rival Extended
8	Times New Roman	Tinos	Liberation Serif	Thorndale
9	Helvetica	Arimo	Geo_Arial	Arial-Relcom
10	Courier New	Courier New CE	TiredOfCourier	TiredOfCourierThin
11	Verdana	Tahoma	MS Reference Sans Serif	Lato
12	Perpetua	ChanticleerRoman	Centaur	CaslonOldFace BT
13	Lucida Sans	Lucida Sans Unicode	Segoe UI	Lucida Sans Typewriter
14	Thorndale	Times New Roman	Liberation Serif	Tinos
15	Franklin Gothic Book	Ebrima	Corbel	Trebuchet MS

Table 2. Code table for (27) alphabets and space

Index	Characters	F1	F2	F3	index	Characters	F1	F2	F3
1	a	1	1	1	16	P	2	3	1
2	b	1	1	2	17	Q	2	3	2
3	c	1	1	3	18	R	2	3	3
4	d	1	2	1	19	S	3	1	1
5	e	1	2	2	20	T	3	1	2
6	f	1	2	3	21	u	3	1	3
7	g	1	3	1	22	v	3	2	1
8	h	1	3	2	23	w	3	2	2
9	I	1	3	3	24	x	3	2	3
10	j	2	1	1	25	y	3	3	1
11	k	2	1	2	26	z	3	3	2
12	l	2	1	3	27	space	3	3	3
13	m	2	2	1					
14	n	2	2	2					
15	o	2	2	3					

Table 3. Experimental results of proposed method

Experiment #	Font name of cover	Capitals letters# in cover	Max.# of characters Can be embedding in cover	Ratio of Stego-Doc. size increasing (%)
1	Arial	537	179	2.531
2	Book Antiqua	1132	377	0.721
3	Candara	1378	459	1.769
4	Century	342	114	0.046
5	Calibri	169	56	1.587
6	Cambria	99	33	1.492
7	Comic Sans	979	326	0.139
8	Times New Roman	430	143	0.176
9	Helvetica	851	283	0.430
10	Courier New	811	270	1.276
11	Verdana	1543	514	0.123
12	Perpetua	2465	821	0.081
13	Lucida Sans	1477	492	0.023
14	Thorndale	89	29	0.606
15	Franklin Gothic Book	3631	1210	0.502
				Avrg. 0.766

As we can see from example, the input in our example is: Happy day. The number of symbols is 6 (with space). According to Code Table of **Table 2**, Symbol (h) coded as:

(1, 3, 2) → (first similar, third similar, second similar) and so on

The cover document font is times new roman, then the symbol (h) coded by first three capitals letters of cover document, by replacing with its similar, **Table 1**. **Figure 4 and 5** represent cover document and stego document respectively.

3. RESULTS

The proposed method of the text steganography method is tested by taking different cover documents of different font types and hiding the same secret message in some of them. We need three characters to hide one character of secret message (one symbol in three capitals letters). That mean, if the cover file contains six characters, we can hiding two characters in it. The results that are got from these experiments can be summarized in the **Table 3**. The size of cover and stego document was compared and shown the average of size for Stego document (after hiding) is increased about 0.766% from original size.

4. DISCUSSION

As it is seen in the **Table 3**, proposed method has good perceptual transparency based on font types similarity, high capacity and robust to digital copy-past operation. This three dimensions are not independent, but should rather be considered as competing goals, which can be balanced when designing a steganographic system. The increasing in stego document size result from using various font types.

5. CONCLUSION

This study proposed a novel method of hiding information in Microsoft Word documents. Microsoft Word documents are very much common in everyday life of today's digital world. The capacity of this method is very high, depending on the number of Capital Letters in cover document. As we show in **Table 3**, some fonts take large size when replace it with their similarity such as *(*Arial) font type and some fonts are not, such as (Lucida Sans, Century) font types.

Because the stego document will not change during compression, copying and paste between computer programs, the data hidden in texts remains intact during these operations.

6. REFERENCES

Bhaya, W.S., 2011. Text hiding in mobile phone simple message service using fonts. J. Comput. Sci., 7: 1626-1628. DOI: 10.3844/jcssp.2011.1626.1628

Bohme, R., 2010. Advanced Statistical Steganalysis. 2nd Edn., Springer, Berlin, ISBN-10: 364214313X, pp: 285.

Elkamchouchi, H. and M. Negm, 2003. Hiding English Information in Extended Arabic Characters (HEMERAC). Proceedings of the 20th National Radio Sience Conference, Mar. 18-20, IEEE Xplore Press, pp: C12-1-8.

Hassan, M. and M. Shirali-Shahreza, 2008. Steganography in persian and arabic unicode texts using pseudo-space and pseudo connection characters. J. Theoretical Applied Inform. Technol., 4: 682-687.

Khairullah, M., 2009. A novel text steganography system using font color of the invisible characters in microsoft word documents. Proceedings of the 2nd International Conference on Computer and Electrical Engineering, Dec. 28-30, IEEE Xplore Press, Dubai, pp: 482-484. DOI: 10.1109/ICCEE.2009.127

Khandekar, S.A. and M.R. Dixit, 2012. Steganography for text messages using image. J. Elect. Commun. Eng., 2: 01-04.

Liu, M., Y. Guo and L. Zhou, 2009. Text steganography based on online chat. Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Sept. 12-14, IEEE Xplore Press, Kyoto, pp: 807-810. DOI: 10.1109/IIH-MSP.2009.318

Moraldo, H., 2012. An Approach for text steganography based on Markov Chains. Proceedings of the 4th Workshop de Seguridad Informatica, (WSI' 12), pp: 26-39.

Por, L.Y. and B. Delina, 2008. Information hiding: A new approach in text steganography. Proceedings of the 7th WSEAS Interenational Conference on Applied Computer and Applied Computational Science, Apr. 6-8, Hangzhou, China, pp: 1-7.

Por, L.Y., K.S. Wong and K.O. Chee, 2012. UniSpaCh: A text-based data hiding method using Unicode space characters. J. Syst. Soft., 85: 1075-1082. DOI: 10.1016/j.jss.2011.12.023

Shakir, A.C., G. Xuemai and J. Min, 2010. Chinese language steganography using the arabic diacritics as a covered media. Int. J. Comput. Applic., 11: 43-46. DOI: 10.5120/1543-2050

Shirali-Shahreza, M., 2008. Text steganography by changing words spelling. Proceedings of the 10th International Conference on Advanced Communication Technology, Feb. 17-20, IEEE Xplore Press, Gangwon-Do, pp: 1912-1913. DOI: 10.1109/ICACT.2008.4494159

Weinschenk, S., 2011. How people read. Graphics.com.

Yang, B., X. Sun, L. Xiang, Z. Ruan and R.Wu, 2011. Steganography in Ms excel document using text-rotation technique. Inform. Technol. J., 10: 889-889.

Zhong, S., X. Cheng and T. Chen, 2007. Data hiding in a kind of PDF texts for secret communication. Int. J. Netw. Sec., 4: 17-26.