New Text Steganography Technique by using Mixed-Case Font

Abdelmgeid Amin Ali

Associate Professor, Dept. of Computer Science
Faculty of Science, Al - Minia University, Egypt

Al - Hussien Seddik Saad

Assistant Lecturer, High Institute for Engineering and Technology (H.I.E.T) Al – Minia, Egypt

ABSTRACT
One of the most important techniques of information hiding is a steganography. Steganography is the art of hiding information within other information or carriers (i.e. text, image, video or audio) in such a way that it is hard or even impossible to tell that it is there. In This paper, a new text steganography method is proposed. The main goal of this method is to use text file as a carrier for the secret message data in such a manner that the resultant text file attracts no attention and hold the same meaning of the original file.

General Terms
Text Steganography, Security, Data hiding in text.

Keywords
Data hiding, Steganography, Stegogramme, Carrier, Secret message.

1. INTRODUCTION
With advancements in digital communication technology and the growth of computer power and storage, the difficulties in ensuring individuals' privacy become increasingly challenging. The degrees to which individuals appreciate privacy differ from one person to another. Various methods have been investigated and developed in information hiding to protect personal privacy[10]. Information hiding is a general term encompassing many subdisciplines. One of the most important subdisciplines is steganography [8] The word steganography is of Greek origin and means "concealed writing" from the Greek words steganos (στεγανός) meaning "covered or protected", and graphein (γράφειν) meaning "writing" [8, 4].

So how basic steganography system is developed? when a steganographic system is developed, it is important to consider what the most appropriate cover work - it is the medium in which the message is embedded and serves to hide the presence of the message, [1]. - should be, and also how the stegogramme - The cover work with the secretly embedded message produced by the encoder - is to reach its recipient. With the Internet offering so much functionality, there are many different ways to send messages to people without anyone knowing they exist. [11].

In terms of development, Steganography is comprised of two algorithms, one for embedding and one for extracting. The embedding process is concerned with hiding a secret message within a cover work, and is the most carefully constructed process of the two. A great deal of attention is paid to ensuring that the secret message goes unnoticed if a third party were to intercept the cover work. The extracting process is traditionally a much simpler process as it is simply an inverse of the embedding process, where the secret message is revealed at the end, [11].

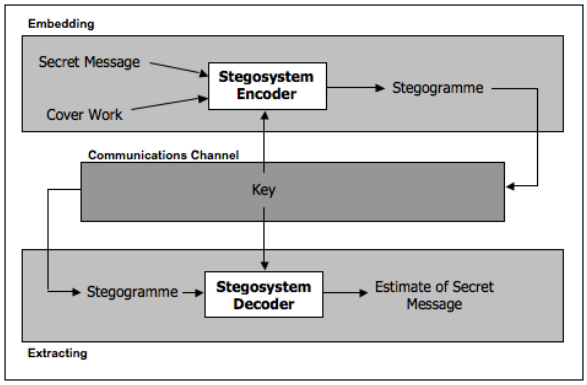

Fig. 1: Basic steganography system[11]

Figure (1) shows one example of how steganography might be used in practice. Two inputs are required for the embedding process:

- Secret message: usually a text file that contains the message you want to transfer.
- Cover work: used to construct a stegogramme that contains a secret message, it may be text, image, video clips or sounds, [3].

The next step is to pass the inputs through the Stego-system encoder, which will be carefully engineered to embed the message within an exact copy of the cover work. The resulting output from the stego-system encoder is the stegogramme, which is designed to be as close to the cover work as possible, except it will contain the secret message. This stegogramme is then sent over some communications channel. [11].

So, steganography is concerned with hiding information in some cover medium, by manipulating properties `of the medium in such a way that the hidden information is not easily detectable by an observer [8]. Also it is one of the information hiding techniques that can be categorized into linguistic steganography and technical steganography. Linguistic steganography defined by Chapman et al. as "the art of using written natural language to conceal secret messages" [12]. A more specific definition by Krista Bennet in explaining linguistic steganography as a medium which required not only the steganographic cover that is composed of natural language text, but the text itself can be either generated to have a cohesive linguistic structure, or the cover text that begin with natural language. On the other hand, technical steganography is explained as a carrier rather than a text.[8]

Text steganography is the most difficult kind of steganography; this is due largely to the relative lack of redundant information in a text file as compared with a picture or a sound file [4, 9].it can be classified in three basic categories: .[8, 12]

1. Format-Based.
2. Random and Statistical Generation.
3. Linguistic Method.

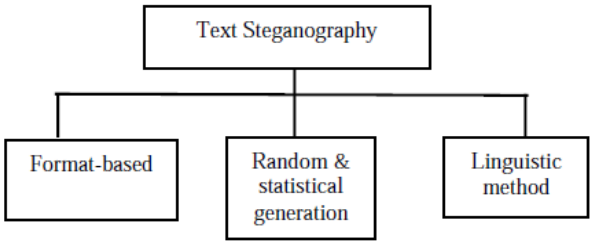

Fig. 2: Three basic categories of text steganography [8].

Format-based methods use physical text formatting of text as a place in which to hide information. Generally, this method modifies existing text in order to hide the steganographic text. Insertion of spaces, deliberate misspellings distributed throughout the text, resizing the fonts are some of the many format-based methods being used in text steganography. However, Bennett has stated that those format-based methods managed to trick most of the human eyes but it cannot trick once computer systems have been used [8].

Random and statistical generation is generating cover text according to the statistical properties. This method is based on character sequences and words sequences. The hiding of information within character sequences is embedding the information to be appeared in random sequence of characters. This sequence must appear to be random to anyone who intercepts the message. A second approach to character generation is to take the statistical properties of word-length and letter frequency in order to create "words" (without lexical value) which will appear to have the same statistical properties as actual words in a given language. The hiding of information within word sequences, the actual dictionary items can be used to encode one or more bits of information per word using a codebook of mappings between lexical items and bit sequences, or words themselves can encode the hidden information [8].

The final category is linguistic method which specifically considers the linguistic properties of generated and modified text, frequently uses linguistic structure as a place for hidden messages. In fact, steganographic data can be hidden within the syntactic structure itself [8].

Finally it can be said that the goal of steganography is to avoid drawing suspicion to the transmission of a hidden message. If suspicion is raised, then this goal is defeated. So, essentially with steganography the actual subject message transmission (be that text, sound or image) is untouched but hidden within another source, [13]. So, steganography is a technique which hides secret information into a cover media or carrier so that it becomes unnoticed and less attractive [5].

This paper is organized as follows. Section I provides an overview for information hiding and text steganography section II discuss some previous work. In section III, the proposed method is explained in details with results. Finally, section IV concludes the paper.

2. PREVIOUS WORK

Following is a list of works has been done on text steganography.

In [6] They proposed a new method to hide info in any letters instead of pointed ones only. They used the pointed letters with extension to hold secret bit 'one' and the un-pointed letters with extension to hold secret bit 'zero'. A very important note is that letter extension doesn't have any affect to the writing content. In fact, this Arabic extension character in electronic typing is considered as a redundant character only for arrangement and format purposes. But the problem in using the extension is that not all letters can be extended with this extension character due to their position in words and Arabic writing nature. The extension can only be added in locations between connected letters of Arabic text; i.e. extensions cannot be placed after letters at end of words or before letter at beginning. So, whenever a letter cannot have an extension or found intentionally without extension it is considered not holding any secret bits.

In [8] Authors proposed a new approach on hiding information in manipulation of white spaces between words and paragraph. The proposed method was able to provide more capacity for hiding more bits of data into a cover-text. The previous embedding scheme was applied in the space which appears between the words. The major drawback of this method was that it requires a great deal of space to encode few bits. But by combining with inter-paragraph in hiding the secret bits can effectively utilizing most of the white spaces in a text document. So, they used inter-word and inter-paragraph spacing for hiding information.

In [7] method proposed depends on one of the characteristics of Arabic language which is the use of Araabs i.e. (Fatah, Kasra, and Damma). Where Fatha is slash like symbol and is written over the character, whereas Kasra is also a slash like symbol but is used below the character and Damma is number nine like symbol which is also placed over the character. But these Araabs are not very commonly used now a days. These Araabs are not only acceptable in Urdu but also rarely used. These Araabs are applicable on every single character of the Arabic language. In general these araabs are noteworthy in Arabic or Urdu text. In this paper, these same characteristics of Arabic or Urdu languages are used for steganography. For this purpose, only fatha is used in reverse order to represent the secret character in the text. For example this fatha in reverse order is named as Reverse fatha and in the text where ever this reverse fatha is used it represents a secret character below it. For applying this method on text first of all make a secret message which is probably of one line hardly 5 words containing 10-15 characters. Now let us select an article of 4 to 5 paragraph or as the number of paragraphs is increased the security is also increased. Then put araabs on the complete text or take an article having araabs already. After this let read the secret message character by character and match it in the article, and they have to put reverse fatha where the secret characters exist sequentially try to use the reverse fatha not on the same line but on different lines. Now the article to the reader is ready to be sent in the form of letter. Finally, it can be said that this is not a practical technique because it has a lot of drawbacks as discussed previous.

In [2] Benefiting from (Gutub and Fattani, 2007) Arabic Text steganography method using letter points and extensions and trying to overcome the low capacity aspect, the authors proposed a technique to hide information in a suitable position inside words instead of pointed letters only. These positions are determined to keep the Arabic text beauty if the text is justified and this allows the message to be hidden without affecting the cover text. They insert the extension letter in the determined position to hold secret bit one and leaving the position empty to hold secret bit zero.

3. PROPOSED METHOD

In this section, the proposed method will be presented which hides the secret message within a cover text file.

Before discussing the proposed method we will explain from where we got this idea, while using the internet and searching for fonts or what so called "cool fonts" that are used for chatting or presentations it has been found that new type of fonts appeared that type capital and small letters at the same time see figures (3,4,5) as an example if you typed the word "hardware" the word will be typed like this "HaRdWaRe" sometimes with different sizes for each letter and another times with same size for whole text. So, we decided to use this new font as a new text steganography method that holds bits of the secret message.

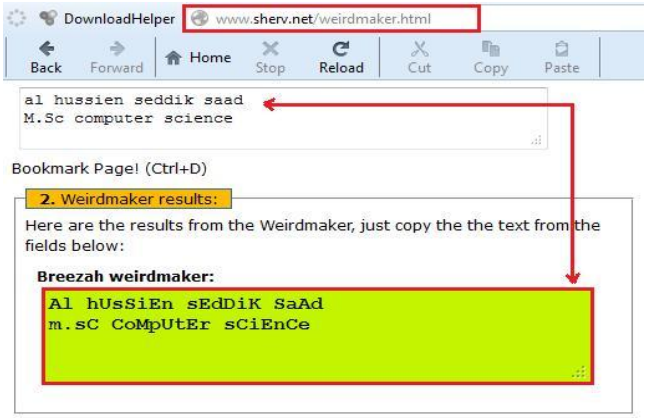

Fig. 3: Font example 1

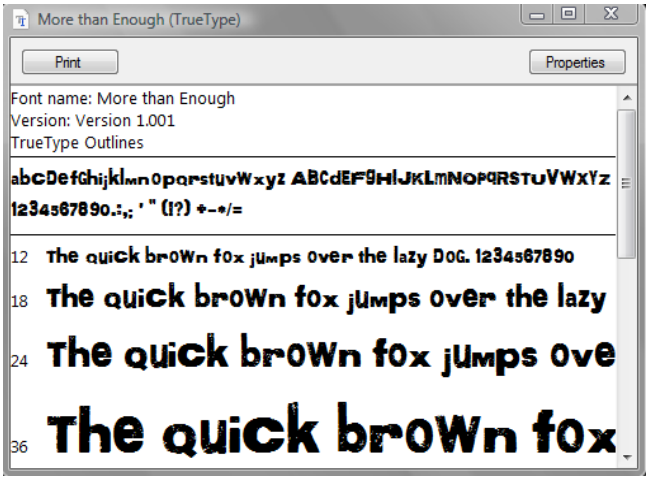

Fig. 4: Font example 2

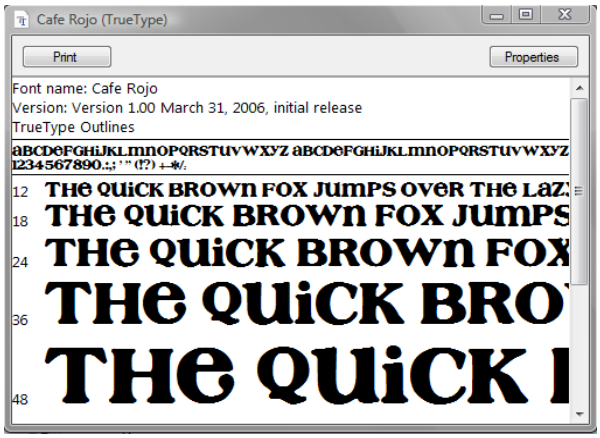

Fig. 5: Font example 3

So, how the secret bits will be hidden? The method will work as in the following algorithm:-

Algorithm: Embedding Using Mixed-Case Font Method
Input : Text File T ; Secret Message M.
Output : Stego Text S.

Steps:-

1) Choose a text file T.
2) Divide T into letters, T = {T_1,T_2 …,T_n}.
3) Get the secret message M.
4) Convert secret message M into stream of bits b.
5) Divide b into bits, b = {b_1,b_2 …, b_n}.
6) Select T_i from T and bi from b.
7) If the b_i is 'one' then change T_i case into capital else change T_i case into small.
8) Repeat steps 6,7 till the whole b is hidden.
9) The resultant file will be the stego text S.

4. RESULTS AND DISCUSSION

In the proposed method, the information can be hidden in English text using the letters as carriers. This approach will insert one character within each 7 letters. So the hiding capacity will be very high compared to other text steganography methods.

The proposed method was tested on some text to compute the capacity of the hiding and to check its advantages. After applying this algorithm the resultant file will looks like the font in Figures (3,4,5) and it will attract no attention as it will be thought as a type of these new cool fonts.

As shown in figure (6,7) for the software of the proposed method, the secret message was "*The Required Password Is 13321154*" and the cover text was "*One of the most important techniques of information hiding is a steganography. Steganography is the art of hiding information within other information or carriers (i.e. text, image, video or audio) in such a way that it is hard or even impossible to tell that it is there. In This paper, a new text steganography method is proposed. The main goal of this method*" the final stego text was "*oNe Of The mOSt ImporTAnt TeChnIques of InForMatIOn hIdInG IS a stEgANOgRaPhY. StEgaNoGRAphY is THe aRt Of HIdiNg infOrmatioN wIthin oTHer inFoRMAtiON oR CArrIErS (I.E. tEXT, iMAgE, VIDeO OR auDio) IN suCh a waY that it Is hArd Or EVEn iMPosSible to tELl thAt iT Is tHEre. IN thIS paPEr, a New*

tEXt stEgaNOgraPhy MEtHoD is PRoPosed. The main goal of this method". And the advantages of the method are :-

1- The stego text will attract no attention because it will look like the "cool fonts" used in chat rooms and presentations, as shown in Figures (6,7) .

2- The hidden text is resistant to enlargement or downsize and these changes do not destroy the hidden information.

3- The text containing hidden phrases is not specific to computer and the hidden information can also be extracted from printed text.

4- In order to recover the information in case of printed text, the text should be scanned using OCR program and then subjected to the relevant program.

5- By this method, a large volume of information can be hidden in text compared to other methods, because the proposed method not using spaces between words or between paragraphs but using the letters themselves.

6- Printing and scanning the stego text didn't destroy the secret data hidden inside it because the proposed method is not dealing with font sizes like other methods.

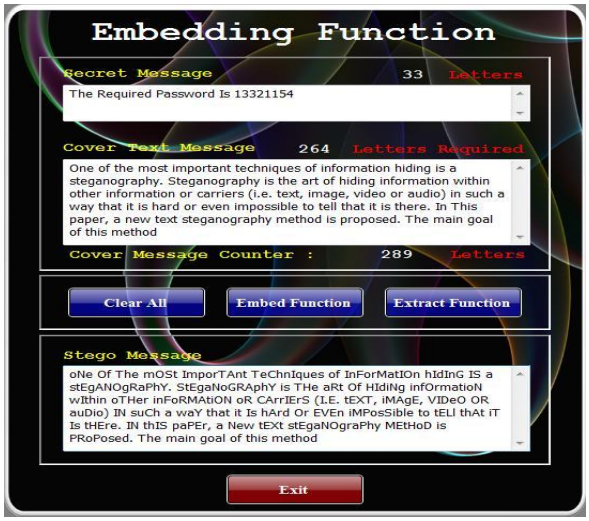

Fig. 6: Embedding Function

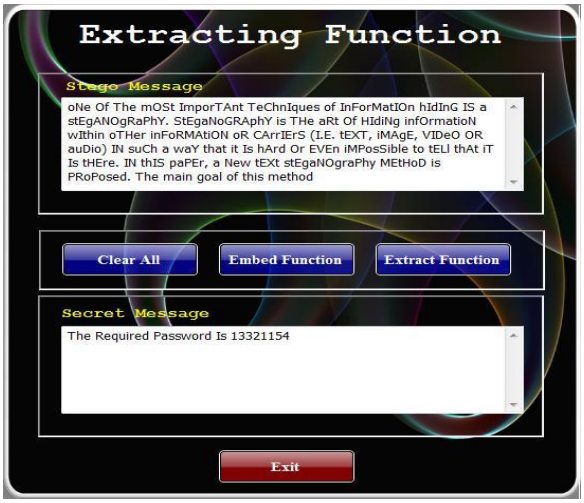

Fig. 7: Extracting Function

5. CONCLUSION

A new approach of text steganography method has been proposed using the letters of the cover file, not spaces between words or between paragraphs, and not using extensions. In the proposed method the letter can be hidden in only 7 letters not 7 words as by using spaces between words. So, its a large amount of data compared to other methods keeping the exact meaning of the text and make it looks like these cool fonts. Finally, we can say that this method achieved the goals of steganography and can be used efficiently.

6. REFERENCES

[1] A. I. Abdul-Sada, Hiding Data Using Lsb-3, J.BASRAH Researches (SCIENCES), VOL. 33., NO.4.(81-88), Dec. 2007.

[2] A. F. AL-Azawi, M. A. Fadhil, Arabic Text Steganography Using Kashida Extensions With Huffman Codes, Journal of Applied Science, 10(5): 436 - 439, ISSN - 1812-5654, 2010

[3] M. A. F. AL-Husainy, Image Steganography By Mapping Pixels To Letters, Journal of Computer Science, 5 (1): 33-38, ISSN 1549-3636, 2009

[4] A. E. Ali, A New Text Steganography Method By Using Non Printing Unicode Characters, Eng. & Tech. Journal, VOL.28, NO.1, 2010

[5] K. C. Chang, C. P. Chang, P. S. Huang, T. M. Tua, A Novel Image Steganographic Method Using Tri-Way Pixel-Value Differencing, Journal of Multimedia, VOL. 3, NO. 2, June 2008

[6] A. A. A. Gutub, M. M. Fattani, A Novel Arabic Text Steganography Method Using Letter Points And Extensions, World Academy of Science, Engineering and Technology 27 2007

[7] J. A. Memon, K. Khowaja, H. Kazi, Evaluation Of Steganography For Urdu /Arabic Text, Journal of Theoretical and Applied Information Technology 2008 JATIT

[8] L. Y. Por, B. Delina, Information Hiding: A New Approach In Text Steganography, 7th WSEAS int. Conf. on Applied Computer & Applied Computational Science (ACACOS '08), Hangzhou, China, April 6-8, 2008

[9] M. H. S. Shahreza, M. S. Shahreza, Arabic/Persian Text Steganography Utilizing Similar Letters With Different Codes The Arabian Journal For Science And Engineering, Volume 35, Number 1b

[10] A. Cheddad, Steganoflage:A New Image Steganography Algorithm, School of Computing & Intelligent Systems, Faculty of Computing & Engineering, University of Ulster, September, 2009.

[11] P. Bateman, H. G. Schaathun, Image Steganography And Steganalysis, Department Of Computing, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey, United Kingdom, 4th August 2008

[12] K. Bennett, Linguistic Steganography: Survey, Analysis, And Robustness Concerns For Hiding Information In Text, Center for Education and Research in Information Assurance and Security, Purdue University, West Lafayette, IN 47907-2086, CERIAS Tech Report 2004-13

[13] J. Watkins, Steganography - Messages Hidden In Bits, Multimedia Systems Coursework, Department of Electronics and Computer Science, University of Southampton, So17 1bj, UK, 15th December 2001