Text Steganography in SMS

Mohammad Shirali-Shahreza
Computer Science Department
Sharif University of Technology
Tehran, Iran
shirali@cs.sharif.edu

M. Hassan Shirali-Shahreza
Computer Engineering Department
Yazd University
Yazd, Iran
hshirali@yazduni.ac.ir

Abstract

One of the services used in mobile phone is the short message service (SMS) which is widely used by the public in all parts of the world especially in Asia and Europe. This service enables people to write and exchange short messages via mobile phone. Due to the limited size of SMS, lack of a proper keyboard on the mobile phone and to improve the speed of typing, new abbreviations have been invented for different words and phrases which has lead to the invention of a new language called SMS-Texting.

One of the main issues in communication is information security and privacy. There are many methods for secret communication and many researchers are working on steganography. In steganography the data is hidden in a cover media such as picture or text.

The present paper offers a new method for secret exchange of information through SMS by using and developing abbreviation text steganography with the use of the invented language of SMS-Texting.

This project has been implemented by J2ME (Java 2 Micro Edition) programming language and tested on a Nokia N71 mobile phone.

Keywords: Text Steganography, SMS (Short Message Service), SMS-Texting, Mobile Phone.

1. Introduction

By development of computer and the expansion of its use in different areas of life and work, the issue of security of information has gained special significance. One of the concerns in the area of information security is the concept of hidden exchange of information. For this purpose, various methods including cryptography, steganography, coding and so on have been used. Steganography is one of the methods which have attracted more attention during the recent years.

In steganography the information is hidden in a cover media so nobody notice the existence of the secret information [1].

Steganography works have been carried out on different media such as pictures [2] and sounds [3].

Text steganography is the most difficult kind of steganography because there is no redundant information in a text file as compared with a picture or a sound file [4].

The structure of text documents is identical with what we observe, while in other types of documents such as in picture, the structure of document is different from what we observe. Therefore, in such documents, we can hide information by introducing changes in the structure of the document without making a notable change in the concerned output.

Contrary to other media such as pictures, sounds and video clips, using text documents has been common since very old times. Even after invention of printing machine, most of the books and documents have contained only texts. This has extended until today and still, using text is preferred over other media, because the texts occupy lesser memory, communicate more information and need less cost for printing as well as some other advantages.

As the use of text and hidden communication goes back to antiquity, we have witnessed to steganography of information in texts since past. For This has been done by classic poets of Iran as well.

Today, the computer systems have facilitated hiding information in texts. The range of using hiding information in text has also developed from hiding information in electronic texts and documents to hide information in web pages.

In the next section we will discuss some text steganography methods.

0-7695-3038-9/07 $25.00 © 2007 IEEE
DOI 10.1109/ICCIT.2007.100

On the other hand mobile phone has been widely welcomed by people. In less than two decades, the mobile phone has turned into a necessary device for people and now one out of every six individuals throughout the world has a mobile phone.

With the expanding use of mobile phones and the development of mobile telecommunications, telecommunication companies as well as companies manufacturing mobile phones decided to add additional features to their mobile phones in order to attract more customers. One of the services that were provided on the mobile phone was the SMS.

The SMS (Short Message Service) is the transfer and exchange of short text messages between mobile phones. The SMS is defined based on GSM digital mobile phones. According to the GSM03.40 standard [5], the length of the exchanged message is 160 characters at most, which are saved in 140 bytes depending to how information is saved according to the standards. These messages may be a combination of digits and letters or be saved in non-text binary form. Using the same binary messages, one can also send pictures as well. The pictures, however, are two-color and have a low quality.

SMS messages are exchanged indirectly and through a component known as the SMSC [6]. SMS messages have the following advantages:

- Communication is possible when the network is busy
- We can exchange SMS messages while making telephone calls
- Sending offline SMS messages
- Providing various services such as e-commerce

Considering the vast use of mobile phones and the exchange of a lot of SMS through mobile phones, hiding information in SMS can be a good choice for establishing secret communications. Therefore, the present paper provides a new method for steganography in SMS.

In our proposed method, the word abbreviations method– which will be discussed in the next section -is used and by development and customizing this method for mobile phone and SMS with the use of SMS language, the new invented language which is called SMS-Texting, information are hidden in SMS.

We will describe this method in section 3. As we know, only one other work has been reported in the area of hiding information in SMS which we will indicate in the next section.

Section 4 is devoted to the study of advantages and disadvantages of our suggested algorithm. A final conclusion ends the paper.

2. Related Works

In this section first we review previous works on the text steganography including abbreviation text steganography method. After that we explain the only reported work on hiding information in SMS.

A few works have been done on hiding information in texts. Following is the list of some different text steganography methods which are reported thus far.

2.1. Line Shifting [7]

In this method, the lines of the text are vertically shifted to some degree (for example, each line shifts 1/300 inch up or down) and information are hidden by creating a unique shape of the text. This method is proper for printed texts.

However, in this method, the distances can be observed by using special instruments of distance assessment and necessary changes can be introduced to destroy the hidden information. Also if the text is retyped or if character recognition programs (OCR) are used, the hidden information would get destroyed.

2.2. Word Shifting [8]

In this method, by shifting words horizontally and by changing distance between words, information are hidden in the text. This method is acceptable for texts where the distance between words is varying. This method can be identified less, because change of distance between words to fill a line is quite common.

But if somebody was aware of the algorithm of distances, he can compare the present text with the algorithm and extract the hidden information by using the difference. The text image can be also closely studied to identify the changed distances. Although this method is very time consuming, there is a high probability of finding information hidden in the text. The same as in the method described under 2.2, retyping of the text or using OCR programs destroys the hidden information.

2.3. Semantic Methods [9]

In this method, they use the synonym of words for certain words thereby hiding information in the text. A major advantage of this method is the protection of information in case of retyping or using OCR programs (contrary to methods listed under 2.1 and 2.2). However, this method may alter the meaning of the text.

2.4. Feature Coding [10]

In this method, some of the features of the text are altered. For example, the end part of some characters such as h, d, b or so on, are elongated or shortened a little thereby hiding information in the text. In this method, a large volume of information can be hidden in the text without making the reader aware of the existence of such information in the text.

By placing characters in a fixed shape, the information is lost. Retyping the text or using OCR program (as in methods 2.1 and 2.2) destroys the hidden information.

2.5. Abbreviation [4]

Another method for hiding information is the use of abbreviations.

In this method, very little information can be hidden in the text. For example, only a few bits can be hidden in a file of several kilobytes.

We expand this method and use it in our steganography method. More details are explained in next section.

2.6. Open Spaces [11]

In this method, hiding information is done through adding extra white-spaces in the text. These white-spaces can be placed at the end of each line, at the end of each paragraph or between the words. This method can be implemented on any arbitrary text and does not raise attention of the reader.

However, the volume of information hidden under this method is very little. Also, some text editor programs automatically delete extra white-spaces and thus destroy the hidden information.

2.7. Persian/Arabic Text Steganography [12]

In this method data is hidden in Persian and Arabic texts by using a special characteristic of these languages.

Considering the existence of too many points in Persian and Arabic phrases, in this approach by vertical displacement of the point, we hide information in the texts.

Although this method doesn't attract attention and can hide a large volume of information in text, it can be only applied to Persian and Arabic and similar languages and for example can't be used for hiding data in English texts, because only two English characters "i" and "j" have dot.

2.8. Stealth Steganography in SMS [13]

As we know, this method is the only reported work on hiding data in SMS. In this method, the information is hidden in the "SMS picture message."

This is a summary of the algorithm used for steganography in the pictures of SMS messages in mobile phones: After converting the picture into the B&W color and a suitable format for the mobile phone, the picture is divided into 3×3 blocks. Information is encoded by the password. The possibility of steganography in each block of the picture is considered. If the result is positive, one bit of information is hidden in the picture by maximally changing one block cell.

Extracting algorithm is the reverse of hiding algorithm. After extracting hidden information from the SMS picture message, the information is removed from the picture by using Morphological methods and the picture message is saved without any data on the recipient's mobile phone (Figure 1).

3. Our Suggested Algorithm

The purpose of this project is to hide information in SMS messages by abbreviations text steganography and with the use of the SMS-Texting language.

First, the SMS-Texting language is described. Then we will examine the abbreviations text steganography method. Finally we will describe our suggested algorithm.

3.1. SMS-Texting

SMS-Texting is a new language used by the youth (and some elderly people) all over the world for sending SMS, internet chatting or sending email. In this method, abbreviations are widely used. For example instead of "Thank You", they just say "10Q" [14]. This language has been coined due to various causes such as the limited size of SMS, lack of proper keyboard on the mobile phones and also to increase the speed in writing SMS. Table I shows a number of such abbreviations commonly in use in SMS.

A) Original picture	
B) Stegano picture	
C) Stealth picture	

Figure 1. Hiding "Iran" message in an SMS picture message [13]

Table 1- List of some SMS acronyms [14]

Acronym	Translation	Explanation
2l8	Too late	The time is too late, missed opportunity
ASAP	As Soon As Possible	Immediately
C	See	Do you understand? OR the verb 'to see'
CM	Call Me	Asking someone to telephone
F2F	Face to face	In person
NC	No Comment	I can't say what I think
R	are	The verb 'to be'
SRY	Sorry	An apology
T+	Think positive	You need to be positive about a situation
ZZZZ	Sleeping	I'm tired, bored or annoyed

3.2. Abbreviation Text Steganography

In the abbreviation text steganography, a list of words which have abbreviated form is prepared. Then, the incoming text is being searched for words and phrases which exist in the above list and which have abbreviated form as well or the phrases and words which have been abbreviated in the text. Considering the data intended to hide in the text, if the aim were to hide bit 0, the full form of the word is used in the text and if the aim were to hide bit 1, the abbreviated form is used in the text. By using full or abbreviated form of words which have abbreviated form in the text, the intended data is hidden in the text. At the time of extracting data from the text, the words which have abbreviated form are identified. If the word is present in full, it shows bit 0 and if the abbreviated from is used it shows bit 1. By concatenating extracted bits, the hidden data reveals.

3.3. Our Method

As mentioned above, the SMS-Texting language is a combination of abbreviated words used in SMS. These words can be also used in abbreviation text steganography method. For this purpose, the words and phrases that have abbreviated forms are identified in the SMS. These words may be ordinary words, such as University which has the known abbreviation of Univ. or may be a word from the collection of words of SMS-Texting, for example, the word "you" with its abbreviated form of "u". As described above, by using full or abbreviated form of words or phrases, the information are hidden in the text. Extraction of information is done by reverse operations.

In this method, not only using SMS words attracts no attention but also one has an increased choice of words, because in addition to ordinary abbreviations, the abbreviated phrases common in SMS are also used.

For implementing this method, J2ME (Java 2 Micro Edition) programming language is used. This language is a version of Java language specifically developed for small devices such as Pocket PCs, PDAs (Personal Digital Assistant) and mobile phones.

This project is composed of two programs:

1- Steganography program which has the duty of hiding information in the SMS

2- Extractor program which has the duty of extracting information from the SMS.

Both above programs are written by J2ME and are implemented on mobile phone.

A list of abbreviated words and phrases of SMS-Texting were prepared from the [14]. This list had fields with two values: one full word and the other its abbreviated form. If a word had more than one abbreviation, the abbreviations were separated by comma (,).

These two lists are then merged and used as a single list for two programs, the steganography and the extractor programs.

In preparing list of SMS-Texting words, words coined by some people for special purposes can also be used. However, this technique makes the SMS more private and damages its readability by the public, so we don’t use it.

The steganography program searches the SMS for the words existing in the list according to the algorithm described. If the number of words found were less than the length of array of zero and one bits which made from the data we want to hide, the program announces that it can not hide the data in the given SMS and that the size of information given is big. Otherwise, it acts according to the algorithm and uses the full word in the text for hiding bit 0 and the abbreviated form for hiding bit 1.

Thus, the information is hidden in the SMS. At the end, the SMS is sent to the consignee whose number is received from the user and on a special port.

After receiving the SMS on the recipient's mobile phone, the extractor program identifies the words by the use of the list of words with abbreviated forms and stores the zero or one value in an array based on the full or abbreviated form of the words.

Now by changing this array from zero and one bits to its plain equivalent, the hidden information is extracted. At the end, this information is stored on the user's mobile phone.

This project has been implemented on a Nokia N71 mobile phone.

By comparing a number of information which is hidden in the SMS and the extracted information from that SMS, we observed that the information was the same and that the system has functioned accurately.

4. Advantages and Disadvantages

In this section some advantages and disadvantages of our method are mentioned.

4.1. Advantages

- This method does not attract any attention because most of the people use abbreviated form in sending SMS. So there are more choices for abbreviations words.
- The cost of sending SMS is low.
- Implementing this method needs little processing. Therefore, it can be easily implemented on mobile phones especially on low cost models as well.
- As indicated in section 3 , only a list of words that have abbreviated form is needed and no complete dictionary is needed. Therefore, the program occupies a small memory and this is an advantage in mobile phones where the storage space is limited.
- Every day, millions of SMS are exchanged throughout the world. Therefore, hiding information in SMS attracts little attention and the risk of identifying SMS containing the hidden information is very low.
- This program has been implemented by Java language. Most of the mobile phones support Java. Therefore, it can be implemented on many types of the mobile phones.

4.2. Disadvantages

- If somebody has our algorithm, he can easily extract the hidden information from the text.
- The volume of SMS is low and limited. Therefore, a small amount of information can be hidden. Of course by the use of concatenate SMS, some SMS can be sent together and thereby a larger amount of information can be hidden in a number of SMS.
- In general, a small amount of information can be hidden in a given text in the abbreviation text steganography method, because the number of words with abbreviated form is also limited.

5. Conclusion

In this paper a new method for steganography in SMS messages is proposed using abbreviation text steganography method and by help of SMS-Texting language.

This method can be used on other devices such as Pocket PC and PDAs. Also this method can be implemented on desktop PCs using SMS gateway for sending and receiving SMS messages.

Getting the idea of this method, the abbreviation text steganography method can be developed and privatized in other areas as well. For example, the abbreviated equivalents of engineering terminology can be used in engineering texts.

In addition to English, SMS can be sent in other languages such as Arabic, Greek and so on. Therefore, the abbreviation text steganography method in these languages can also be used for hiding information in SMS.

A small amount of information can be hidden in this method. Therefore, the size of input data can be decreased by compression before hidden the data.

Cryptography of the intended data can also add the security of this method.

However, both of these operations need further computations which can not therefore be implemented on any type of mobile phone and if used, the number of mobile phones capable of executing this program will be decreased.

10. References

[1] J.C. Judge, "Steganography: Past, Present, Future", *SANS white paper*, November 30, 2001, http://www.sans.org/rr/papers/index.php?id=552, last visited: 23 August 2007.

[2] M. Shirali Shahreza, “An Improved Method for Steganography on Mobile Phone”, *WSEAS Transactions on Systems*, vol. 4, issue 7, July 2005, pp. 955-957.

[3] K. Gopalan, "Audio steganography using bit modification", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03)*, vol. 2, 6-10 April 2003, pp. 421-424.

[4] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", *IBM Systems Journal*, vol. 35, Issues 3&4, 1996, pp. 313-336.

[5] Nokia Mobile Phones Ltd., "Sending Content over SMS to Nokia Phones", *Version 1.0 Forum Nokia*, May 2001 http://www.forum.nokia.com.

[6] European Telecommunications Standards Institute (ETSI), GSM 03.40 v7.4.0, Digital cellular telecommunications system (Phase 2+), Technical realization of the Short Message Service (SMS), *ETSI 2000*, http://www.etsi.org.

[7] S.H. Low, N.F. Maxemchuk, J.T. Brassil, and L. O'Gorman, "Document marking and identification using both line and word shifting", *Proceedings of the Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '95)*, vol.2, 2-6 April 1995, pp. 853 - 860.

[8] Y. Kim, K. Moon, and I. Oh, "A Text Watermarking Algorithm based on Word Classification and Inter-word Space Statistics", *Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, 2003, pp. 775–779.

[9] M. Niimi, S. Minewaki, H. Noda, and E. Kawaguchi, "A Framework of Text-based Steganography Using SD-Form Semantics Model", *Proceedings of the Pacific Rim Workshop on Digital Steganography 2003*, Kyushu Institute of Technology, Kitakyushu, Japan, 3-4 July, 2003.

[10] K. Rabah, "Steganography-The Art of Hiding Data", *Information Technology Journal*, vol. 3, Issue 3, 2004, pp. 245-269.

[11] D. Huang and H. Yan, "Interword Distance Changes Represented by Sine Waves for Watermarking Text Images", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, December 2001, pp. 1237-1245

[12] M.H. Shirali-Shahreza and M. Shirali-Shahreza, "A New Approach to Persian/Arabic Text Steganography", *Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2006)*, Honolulu, HI, USA, 10-12 July, 2006, pp. 310-315.

[13] M. Shirali-Shahreza, "Stealth Steganography in SMS", *Proceedings of the third IEEE and IFIP International Conference on Wireless and Optical Communications Networks (WOCN 2006)*, 11-13 April, 2006.

[14] K. Beare, "SMS - Texting", *English as 2nd Language*, http://esl.about.com/, last visited: 23 August 2007.