Volume 6, Issue 1, January 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A New Steganographic Method using Word Length

Monika Agarwal
Department of Information Technology
Rajkumar Goel Institute of Technology, Ghaziabad,
India

Abstract— Steganography is the technique of masking a message inside an innocuous media object without creating hunch to others. In this paper, we present a new approach to text steganography. The approach uses length of the words to conceal a secret message inside a text file. The resulting stego file is exactly the same as the cover file. We also present an empirical comparison of the proposed approach with some of the existing approaches and show that our approach yields better results.

Keywords— Cover, Cryptography, Information hiding, Steganography, Text steganography

I. INTRODUCTION

Steganography is the skill of disguising a message inside a media object without raising any suspicion to others [1]. Enciphering the data is not enough, as criminals detect, and react to encrypted communications [2]. But when steganography is used, even if an attacker gets the stego file, he cannot suspect the communication as it is carried out in a hidden way. Steganography comes from the Greek words and means "covered writing" [3], [4], [5]. Steganography gained importance because the US and British government banned the use of cryptographic communications including mailing of chess games, clippings from newspapers, etc [6].

Instead of physical objects, modern day steganography deals with electronic media [7]. Text Steganography uses character based text to hide the secret information [8]. Text Steganography is preferred because text files require less memory and can be communicated easily [9].

This paper presents a new and simple approach in text steganography by hiding each bit of message in a word of cover file using length of that word. As this is done without altering the cover, the resulting stego file is exactly the same as the cover.

The rest of the paper is organized as follows: Section II keys out some of the popular approaches of text steganography. Section III describes the proposed approach. Section IV traces the results when the proposed approach is compared with the existing approaches. Section V discusses the merits and demerits of the proposed approach and other related issues. Section VI casts the conclusion.

II. RELATED WORK

This section highlights some of the existing approaches of text steganography.

1. Line Shift

In this method, the secret bits are hidden by vertically shifting the text lines to some degree [10], [11]. For hiding bit 0, a line is shifted up and to hide bit 1, the line is shifted down [12], [13].

2. Word Shift

This method shifts the words horizontally, i.e. left or right to represent the bit 0 or 1 respectively [13].

3. Syntactic Method

Punctuation marks such as full stop (.), comma (,), etc. are used in this method, at proper places for hiding bits 0 and 1 [10], [11], [14].

4. White Steg

This technique uses white spaces for hiding the secret message. For example, one space after a word represents the bit 0 and two spaces after a word represents the bit 1 [3], [5], [14].

5. Spam Texts

XML and HTML files hide the bits using white space in tags [13].

6. SMS-Texting

SMS-Texting hides bit 0 using full form of a word and the bit 1 is hidden by using abbreviated form of that word [15].

7. Feature Coding

In feature coding, the secret message is hidden by altering one or more features of the text [13]. OCR program or re-typing can destroy the hidden information [6], [16].

8. SSCE (Secret Steganographic Code for Embedding)

This technique scrambles a message using SSCE table and then embed the scrambled text in a cover by inserting articles a or an with the non specific nouns in English language [8].

9. Word Mapping Method

This technique encrypts the secret message using genetic operator crossover and then embed the cipher text in a cover by inserting blank spaces between words of even or odd length [9].

10. MS Word documents

This technique degenerates the text segments of a word document and the secret message is embedded in the choice of degenerations [17].

11. Cricket Match Scorecard

In this method, data is hidden in a cricket match scorecard by pre-appending a meaningless zero before a number to represent bit 1 and leaving the number as it is to represent bit 0 [18].

12. CSS (Cascading Style Sheets)

This technique encrypts a message using RSA public key cryptosystem and then the resulting cipher text is embedded through a Cascading Style Sheet (CSS) by inserting white space after a semicolon [19].

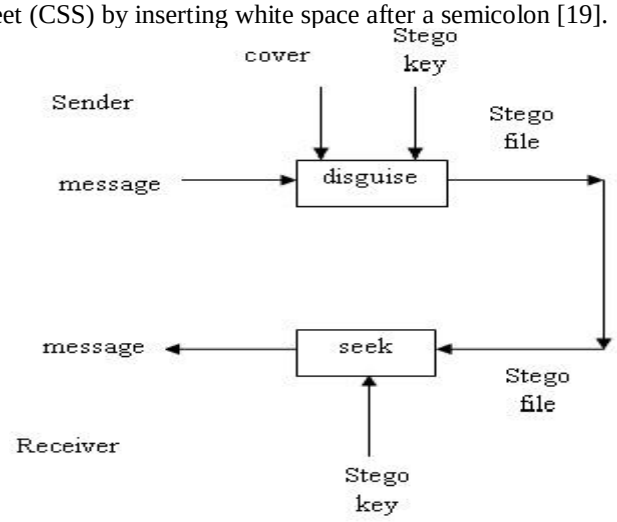

Fig. 1 The basic model of text steganography

III. PROPOSED APPROACH

The proposed method masks a secret message in an innocuous cover. Cover can be drawn from any source like newspaper, book, magazine, etc. The method is applicable not only to English language but to any other language. The method works by hiding each bit of message in a word of cover using word length.

Fig. 1 shows the basic framework of text steganography. The framework consists of two functions: Disguise function, which conceals message using a stego key in a cover, and Seek function, which extracts the concealed information from the received stego file using the same stego key.

A. Algorithm for Hiding

1) Select a cover file.
2) Convert the message to a bit stream.
3) Read a bit of the message.
4) Read a word from the cover file.
5) Calculate length (l) of the word.
6) d = 2*l
7) If the bit to be hidden is one, write the value of l in the stego key.
8) If the bit to be hidden is zero, write the value of d in the stego key.
9) Write a space in the stego key.
10) Write the word in the stego file.
11) Repeat steps 3 to10 till the end of the bit stream..
12) Send the stego file to the receiver.
13) Encrypt the stego key file and send it to the receiver.

B. Algorithm for Extraction

1) Read a number (n) from the stego key.
2) Read a word from the stego file.
3) l= length of the word.
4) If n = l, then bit b = 1.
5) Else if n = 2*l, then bit b=0.
6) Write b in the message file.
7) Repeat the above steps till the end of the stego key.
8) Convert the extracted bit stream into characters to get the message.

C. Example

Consider a message "An" to be concealed in a cover file. Fig. 2 shows the cover, Fig. 3 shows the stego file, Fig. 4 shows the stego key file and Fig. 5 shows the extracted bits which when converted to characters yields the message "An".

The house itself looked empty. The doors and windows were locked. The front verandah bare. Unfurnished. But the skyblue Plymouth with chrome tailfins was still parked outside, and inside, Baby Kochamma was still alive. She was Rahel's baby grandaunt, her grandfather's younger sister. Her name was really Navomi, Navomi Ipe, but everybody called her Baby. She became Baby Kochamma when she was old enough to be an aunt.

Fig. 1 Cover

The house itself looked empty. The doors and windows were locked. The front verandah bare. Unfurnished. But the skyblue Plymouth with chrome tailfins was still parked outside, and inside, Baby Kochamma was still alive. She was Rahel's baby grandaunt, her grandfather's younger sister. Her name was really Navomi, Navomi Ipe, but everybody called her Baby. She became Baby Kochamma when she was old enough to be an aunt.

Fig. 3 Stego file

6 5 12 12 10 6 10 3
14 4 6 6 5 8 4 22

Fig. 4 Stego key

0100000101101110

Fig. 5 Extracted bits

IV. EXPERIMENTAL RESULTS

This section presents an emperical comparison, based on capacity ratio, of the proposed approach with the other text steganographic approaches. The capacity ratio is computed by dividing the amount of hidden bytes with the size of the cover text in bytes [20], as

Capacity ratio = (amount of hidden bytes) / (size of the cover)

Assuming one character takes one byte in memory, we have calculated the percentage capacity which is capacity ratio multiplied by 100. The samples of embedded data used are:

1) Ego (3 byte)
2) Minute (6 byte)
3) Hello World! (12 byte)
4) Failure is never final ! (24 byte)
5) Smile is an inexpensive way to improve your looks. (50 byte)
6) Its not the load that breaks you down, its the way you carry it. (63 byte)
7) Don't find hundred reasons why you can't do a thing, but just find one reason why you can and do it. (100 byte)
8) Tide recedes and leaves behind bright sea shells on sand
   Sun sets but its warmth lingers on land
   Music stops and its echoes on in sweet refrains
   For every joy that passes, something beautiful remains (202 byte)
9) Steganography is not a new area. It dates back to 5th century BC. Harpagus used hare to send his message by killing it and hiding the message inside its belly. A person disguised as hunter carried the hare to the destination. Another incident was of King Darius of Susa. Histiaeus was assigned the duty of shaving the head of his most trusted slave. (349 byte)
10) Steganography is not a new area. It dates back to 5th century BC. Harpagus used hare to send his message by killing it and hiding the message inside its belly. A person disguised as hunter carried the hare to the destination.

Another incident was of King Darius of Susa. Histiaeus, prisoner of Darius, was assigned the duty of shaving the head of his most trusted slave and then the message was tattooed on his shaved scalp. After some time, when the hairs of the slave grew back, his head was shaved again. (508 byte)

V. DISCUSSION

There are three main issues to be considered when studying steganographic systems: capacity, security, and robustness. Capacity is defined as the ability of cover to store secret information. Security refers to the ability of an attacker to suspect hidden data easily. Robustness is the ability of protecting the unseen data from alteration [20].

Table 1 shows the percentage capacity obtained when our method is applied to the aforementioned ten experimental samples. In Table 2, the comparison of the proposed approach is done with some of the popular approaches of text Steganography on the basis of average percentage capacity. It can be seen that the average percentage capacity of the proposed method is greater than the other approaches.

As the stego file looks exactly the same as cover, an attacker cannot figure out any difference in the two files. Thus, stego file is secure. Also, there are no extra spaces or tabs or misspelled words. So, opening it with a word processor program will not draw suspicion regarding the existence of concealed content.

The stego file can withstand OCR techniques and rewriting does not lead to the loss of the hidden data. Hence, the proposed approach is robust.

TABLE I PERCENTAGE CAPACITY OF THE PROPOSED APPROACH OVER THE TEN SAMPLES

I	II	III	IV	V	VI	VII	VIII	IX	X
2.30	2.15	2.21	2.21	2.29	2.26	2.28	2.26	2.28	2.25

TABLE III AVERAGE PERCENTAGE CAPACITY OF THE APPROACHES

Proposed Approach	White Steg	SMS Texting	Feature Coding	Word Mapping	Spam Text	Word Shift
2.25	1.874	1.71	1.479	1.464	1.164	1.03

VI. CONCLUSION

This paper presents a novel approach to text steganography. The proposed method masks a secret message in an innocuous cover. Cover can be drawn from any source like newspaper, book, magazine, etc. An advantage of the method is that it is applicable not only to English language but to any language. The method works by hiding a bit of message in a word of cover file using length of that word. The proposed approach has capacity better than the other approaches and is robust. Also, stego file is natural looking meaningful piece of text written in any language, does not contain any extra white spaces or misspelled words, and is exactly the same as cover. So it does not draw suspicion and hence is secure.

REFERENCES

[1] S. Changder, D. Ghosh, and N. C. Debnath, "Linguistic approach for text steganography through Indian text," *2nd Int. Conf. on Computer Technology and Development*, 2010, p. 318-322.

[2] R. J. Anderson, and F. A. P. Petitcolas, "On the limits of steganography," *IEEE J. of Selected Areas in Communication*, vol. 16, pp. 474-481, May 1998.

[3] L. Y. Por, and B. Delina, "Information hiding-a new approach in text steganography," *7th WSEAS Int. Conf. on Applied Computer and Applied Computational Science*, 2008, p. 689-695.

[4] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding- a survey," in *Proc. IEEE*, 1999, p. 1062-1078.

[5] L. Y. Por, T. F. Ang, and B. Delina, "White steg-a new scheme in information hiding using text steganography," *WSEAS Transactions on Computers*, vol. 7, pp. 735-745, June 2008.

[6] K. Rabah, "Steganography-the art of hiding data," *Information Technology Journal*, vol. 3, pp. 245-269, 2004.

[7] K. Benett, "Linguistic steganography- survey, analysis and robustness concerns for hiding information in text," Purdue University, West Lafayette, IN, CERIAS Tech. Rep. 2004-13, 2004.

[8] I. Banerjee, S. Bhattacharyya and G. Sanyal, "Novel text steganography through special code generation," *Int. Conf. on Systemics, Cybernetics and Informatics*, 2011, p. 298-303.

[9] S. Bhattacharyya, I. Banerjee, and G. Sanyal, "A novel approach of secure text based steganography model using word mapping method," *Int. J. of Computer and Information Engineering*, vol. 4, pp. 96-103, 2010.

[10] M. H. S. Shahreza, and M. S. Shahreza, "A new synonym text steganography," *Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing*, 2008, pp. 1524-1526.

[11] M. H. S. Shahreza, and M. S. Shahreza, "A new approach to Persian/Arabic text steganography," *5th IEEE/ACIS Int. Conf. on Computer and Information Science and 1st IEEE/ACIS Int. Workshop on Component-Based Software Engineering, Software Architecture and Reuse*, 2006, p. 310-315.

[12] J. T. Brassil, S. H. Low, N. F. Maxemchuk, and L. O'Gorman, "Document marking and identification using both line and word shifting," in *Proc. INFOCOM '95 of the 14th Annual Joint Conf. of the IEEE Computer and Communication Societies*, 1995, p. 853-860.

[13] J. Cummins, P. Diskin, S. Lau, and R. Parlett, "Steganography and Digital Watermarking," School of Computer Science, Univ. of Birmingham, pp. 1-24, 2004.

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

[15] M. S. Shahreza, and M. H. S. Shahreza,"Text steganography in SMS," *Int. Conf. on Convergence Information Technology*, 2007, p. 2260-2265.

[16] J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O'Gorman, "Electronic marking and identification techniques to discourage document copying," *IEEE J. on Selected Areas in Communication*, vol. 13, pp. 1495-1504, Oct 1995.

[17] T. -Y. Liu, and W. -H. Tsai, "A new steganographic method for data hiding in Microsoft Word documents by a change tracking technique," *IEEE Transactions on Information Forensics and Security*, vol. 2, pp. 24-30, Mar 2007.

[18] M. Khairullah,"A novel text steganography system in cricket match scorecard," *Int. J. of Computer Applications*, vol. 21, pp. 43-47, May 2011.

[19] H. Kabetta, B. Y. Dwiandiyanta, and Suyoto, "Information hiding in CSS: a secure scheme text steganography using public key cryptosystem," *Int. J. on Cryptography and Information Security*, vol. 1, pp. 13-22, Dec 2011.

[20] F. A. Haidari, A. Gutub, K. A. Kahsah, and J. Hamodi, "Improving security and capacity for Arabic text steganography using Kashida extensions," *IEEE/ACS Int. Conf. On Computer Systems and Applications*, 2009, p. 396-399.

