2014 1st International Congress on Computer, Electronics, Electrical, and Communication Engineering (ICCEECE2014)
IPCSIT vol. 59 (2014) © (2014) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2014.V59.3

An Intelligent Approach for Secure Data Transmission using Text Steganography

Mr. Hitesh Singh [1] and Mr. Anirudra Diwakar [1+]

[1] HMRITM, Delhi

Abstract. Secure Data Transmission refers to the transfer of data using secure or insecure channels. Security mechanisms are used to prevent unauthorized access of data, either by making the transmission channel secure or by securing the data using cryptographic techniques making it suitable for transmission on insecure channels. This paper presents a novel approach for secure data transmission over an insecure channel using text steganography.

Keywords: secure data transmission, transmission on insecure channels, security mechanism, text steganography, rich text format, protection.

1. Introduction

Security of data is of prime importance nowadays because of growing user base of the Internet. There is a need to protect your data from unauthorized access and ensure that it remains confidential and prone to snooping. Security can be ensured either by protecting your data using sophisticated encryption techniques (cryptography) or by securing the channel through which the data is to be sent, or both. Securing the transmission channel is a costly affair and prevents "Man in the Middle" attacks. Securing the channel however is not feasible in many situations. On the other hand, cryptographic and steganography [1] [2] techniques are easier to implement and the necessity for a secure channel is removed. This paper presents an innovative method for secure data transmission by generating obfuscated data and using steganography [1] [2] as an additional security measure.

Steganography is an ancient method of hiding messages in a medium such as text, audio [3][4], video, images [1] [2] wherein only the sender and the receiver know about the presence of a hidden message. A third person shall not be able to detect whether there was a hidden message embedded somewhere in the document. This feature is also an advantage of Steganography [1] [2] over Cryptography. Cryptography involves generating cipher codes, which although are difficult to reverse engineer but easily attract unwanted attention of a hacker or malicious user. Steganography on the other hand avoids this unwanted attention by preserving the integrity of original document. The document appears same before and after the steganography algorithm is applied.

Steganography means forming covert channels. The broad filed of steganography has the following two directions,

- Protection against detection
- Protection against removal

Protection against detection deals with hiding of messages in such a way that they are undetectable to a human eye. The original message can only be retrieved using sophisticated methods. The process of extracting the message from covered text is known as Steganalysis [1].

Protection against removal employs fingerprinting and watermarking techniques. Fingerprinting mechanisms allow identification of source of a document by embedding a unique identifier. Extracting the identifier reveals the source. Watermarking techniques deal with marking objects or pages of text with special and unique marks which may or may not be noticeable. Special watermark detection techniques must be employed if the watermark is invisible to human eye.

[+] Corresponding author. Tel.: + (91)9958353874
E-mail address: anirudradiwakar@gmail.com

Text Steganography is one of the most complex forms of steganography [5], as finding the vulnerable positions to hide data are less. Text steganography mechanisms for protection against detection are described in the next section.

2. Previous methods

2.1 Format Based Text Steganography

Format based text steganography exploits properties of text formatting and is broadly classified into four categories [5], as follows

- Line Shift Encoding [6] [7] [8], shift the line of an alphabet vertically up or down to hide a message. A shifted down line may represent binary 0, and an up-shifted line may represent binary 1.

- Word Shift Encoding [6] [7], shifts the whole word horizontally to the left or right and hides message exploiting the property of variable white spaces in justified text alignment.
- Feature Encoding [6] [7], exploits the features of text such as colors, font, font size, height of characters. Messages are hidden by changing one of these properties, which may seem indiscernible to a third party and can only be detected using sophisticated Steganalysis procedures.
- White Space Method [6], hides data by manipulating white spaces in a document. White spaces exist in between words, sentences, lines and paragraphs. Some extra spaces can be added to the start and end of a word, line, sentence or paragraph without arousing any suspicion as it preserves the integrity of the original document.

2.2 Linguistic Text Steganography

Linguistic Text Steganography exploits languages and their properties to hide messages. Linguistic means language, so it is also known as Language based text steganography. Semantic Text Steganography and Syntax Based Text Steganography are two major types of Linguistic Text Steganography.

- Semantic Method [6] of Text Steganography exploits multiple meaning of a word, ie synonyms. A word can have multiple meaning, and data can be hidden using primary and secondary meanings for same word. For example, the words large and big can be used as primary and secondary. When decoding the text, primary word will be read as binary 1 and secondary word will be read as binary 0, or vice-versa, thus retrieving the hidden message. A word dictionary may be used to search for synonyms for keywords in the document. There might be multiple for a given word, which allows hiding of more than one bit at a time, hence increasing the capacity.
- Syntax Method [6] of Text Steganography exploits syntax of the languages. For example, a word might have two different spellings in two languages, such as “Color, Colour”, “Fulfill, Fulfil”, “Criticize, Criticise”. The choice remains arbitrary and can used to hide data by alternating the words for binary 0 and 1.

3. Proposed algorithm

Let us assume there are two entities, say sender and receiver. The sender wants to send a message to the receiver. This algorithm generates an on-site key for the sender, which in turn used as input to color text steganography algorithm. A stego object, which hides the key from unwanted entities is then transmitted to the receiver using an insecure channel. The receiver performs Steganalysis to retrieve the key, which in turn is used as input to the reverse algorithm and output of final message is received.

3.1 On-site Key Generation

The on-site key generation mechanism generates a unique key for a specific message. The generator document must be present with both the Sender and Receiver in order to successfully send and receive messages from one another. Let us assume a binary message of 0101 is to be sent. The generator document is taken as a text file having the content “This is to inform you that xyz”.

A mapping based on “absolute height of lowercase English alphabets” is used in key generation. Alphabets with less absolute height are assigned as binary 0, and alphabets with larger absolute height are assigned binary 1, as shown in Table 1.

Table 1 Mapping of Alphabets

Value	Alphabets Mapped
Binary 0	a, c, e, i, m, n, o, r, s, u, v, w, x, z
Binary 1	b, d, f, g, h, j, k, l, p, q, t, y

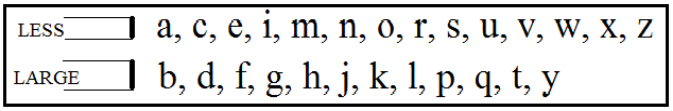

Fig. 1 Comparison of Absolute Heights of English Alphabets

Fig. 1 compares absolute height of all English alphabets. The lowercase alphabets “a, c, e, i, m, n, o, r, s, u, v, w, x, z” are alphabets having comparatively lesser length than the alphabets “b, d, f, g, h, j, k, l, p, q, t, y”.

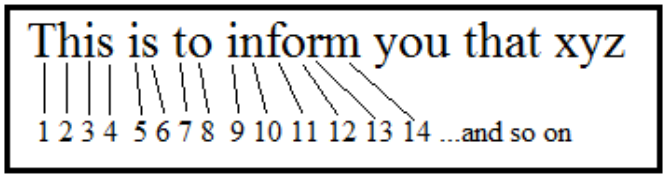

Fig. 2 Counter Values of Generator Document

Counter Values are generated using “Generator Document” by assigning increasing integer values (starting from 1) to each alphabet, as shown in Fig. 2. White spaces are ignored.

Key for message 0110 is generated by searching for an alphabet mapped with least counter value not previously used. The mapping to be used is defined in Table 1. 0 – maps to ‘i’ at counter value 3, 1 – maps to ‘t’ at counter value 7, 0 – maps to ‘o’ at counter value 8, 1 – maps to ‘f’ at counter value 11. Therefore the key generated is concatenation of above counter values 3, 7, 8, 11.

3.2 Text Steganography of key

The key generated in previous step must be hidden in a cover document before transmitting it to the receiver. Color text steganography is used here. When two identical alphabets are colored with successive or similar RGB values, it is visually impossible to differentiate between the two, as shown in Fig. 3.

Fig. 3 Indiscernible Changes

Fig. 3 shows English alphabet A colored with RGB values 000 and 111, which are visually very similar to one another. This property is used to hide data in a text file. Text with RGB value (0,0,0) (hereby named as C0) may be assigned as binary 0, and text with RGB value (1,1,1) (hereby named as C1) may be assigned as binary 1. Thus a message 0110 can be hidden as four alphabets having colors (C0,C1,C1,C0). This maintains visual integrity of text file as well as allows detection of message using specially programmed software. The capacity of this method is however limited due to the use of only two colors during the message encoding process. The storage capacity is directly proportional to number of alphabets present in input file and the number of colors used for hiding text. Thus by increasing number of available colors to 4, namely RGB (0,0,0) , RGB (1,1,1), RGB (2,2,2) and RGB (3,3,3), the capacity can be doubled as two bits of message are parsed directly. This method is called 2-bit encoding.

In another variation, 16 colors, RGB(0,0,0) to RGB(15,15,15) were used to quadruple the storage capacity without disturbing visual integrity of the document. This method is called 4-bit encoding.

Thus, the key can be successfully hidden by converting 3, 7, 8 and 11 to their respective binary values and using color text steganography. It won't attract unwanted attention when sent through an insecure channel.

3.3 Steganalysis at Reciever Side

The process of retrieving message from a Stego Object is known as Steganalysis [1]. Since 4-bit encoding with Color Text Steganography was used by the Sender, the reverse process must be used by the receiver.

Thus a binary representation of key can be retrieved using the Mapping Scheme given in Table 2. Conversion of retrieved binary to ASCII yields 3, 7, 8, 11, which is our key

Table 2 RGB to Binary Mapping

Detected (R,G,B) Values	Retrieved Binary Value
0, 0, 0	0000
1, 1, 1	0001
2, 2, 2	0010
3, 3, 3	0011
4, 4, 4	0100
5, 5, 5	0101
6, 6, 6	0110
7, 7, 7	0111
8, 8, 8	1000
9, 9, 9	1001
10, 10, 10	1010
11, 11, 11	1011
12, 12, 12	1100
13, 13, 13	1101
14, 14, 14	1110
15, 15, 15	1111

3.4 Retrieval of original message

When the key is successfully retrieved by the receiver, the original message sent by the sender can be generated again using the Generator Document and Counter values using Mapping provided in Table 1 and Figure 2.

The key 3, 7, 8, 11 can be reversed as follows. 3 maps to third counter value, which represents lowercase English alphabet 'i' in the Generator Document. Since alphabet 'i' can be mapped to binary 0 using Table 1, the first bit of our message is 0. Similarly, 7 maps to alphabet 't' having value binary 1, 8 maps to alphabet 'o' having binary value 0, and 11 maps to alphabet 'f' having binary value 1. Thus original message 0101 is received by the Receiver successfully.

4. Results

The implemented algorithm defines a novel method of Secure Data Transmission using Text Steganography. The results of implemented algorithm using 4-bit encoding Text Steganography are shown in Table 3.

Table 3 Implementation Results

Generator Document	Input message	Execution Time	Data Embedded %
100 characters	30 bits	10 ms	100
220 characters	208 bits	16 ms	100
1100 characters	1000 bits	50 ms	100
4000 characters	3700 bits	80 ms	100

5. Conclusion

This paper presents innovative ideas in the field of Secure Data Transmission using Text Steganography. No mechanism is perfectly secure, however necessary steps are taken in order to reduce the risk of

compromising data to unauthorized personnel. Novel Text Steganography mechanisms were used to create Covered Writing, which obfuscates a malicious user into thinking there is no hidden data in the transmitted document. Encoding Techniques were used to increase the capacity and effective execution time of given algorithm.

6. References

[1] T. Moerland, "Steganography and Steganalysis", May 15, 2003.

[2] J.C Judge,"Steganography:past,present,future" *Sans white paper*, November 30,2001

[3] N Provos and P.Honeyman "hide and seek: an introduction to steganography", *IEEE security and privacy*, p.p, 32-44 May/June 2003

[4] T. Morkel, J.H.P. Eloff , M.S. Olivier "AN OVERVIEW OF IMAGE STEGANOGRAPHY".

[5] Richard Popa, "An Analysis of Steganographic Techniques", *The Politehnica University of Timisoara, Faculty of Automatics and Computers*, Department of computer science and Software Engineering,1998.

[6] Mrs. Kalavathi.Alla,Dr. R. Siva Ram Prasad, "A Novel Hindi Text Steganography Using Letter Diatrics and it's Compound Words", *IJCSNS International Journel of Computer Science and National Security*, Vol.8 No.12, December 2008.

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

[8] S. H. Low N. F. Maxemchuk J. T. Brassil 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 Communication Societies.*

Hitesh Singh is an Assistant Professor at Department of Computer Science Engineering at HMR Institute of Technology and Management (HMRITM). He was born in Delhi and completed his Bachelor's in Technology in 2007, Master's in Technology in 2009 from C-DAC Noida, and Executive MBA in 2013 from Indian Institute of Technology, Mumbai. His current research interests include Steganography and Telecommunication.

Anirudra Diwakar is pursuing his Bachelor's in Technology in the field of Computer Science Engineering from HMR Institute of Technology and Management (HMRITM). He was born in Delhi and trained in platforms such as Java, Salesforce and was awarded Academic Excellence Award in 2013 by HMRITM. His curent research interests include Network Security and Steganography.