International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2015): 6.391

A Comparison of Different Approaches in Text Steganography

Tewhasom Aregay Weldu

Lecturer, HOD, Department of Computer Science, Adigrat University, Ethiopia

Abstract: ***The important of reducing a chance of the information being detected during the transmission is being an issue now days in the digital age, Steganography is the art and science of data hiding the existence of the communication by concealing a secret message inside another unsuspicious message. Steganography is often being used together with cryptography and offers an acceptable amount of privacy and security over the communication channel.This paper presents an overview of text and a brief history of steganography along with various existing text-based steganography techniques. Highlighting some of the problems inherent in text steganography as well as issues with existing solutions. The first approach hides a message in a wordlist where ASCII value of embedded character determines length and starting letter of a word. The second approach conceals a message, without degrading cover, by using start and end letter of words of the cover. The third approach is proposed in information hiding using inter-word spacing and inter-paragraph spacing as a hybrid method. Our method offers dynamic generated stego-text with six options of maximum capacity according to the length of the secret message. This paper also analyzed the significant drawbacks of each existing method and how our new approach could be recommended as a solution.***

Keywords: Information hiding, steganography, text steganography

1. Introduction

Steganography is a technique of hiding information in digital media. In contrast to cryptography, it is not to keep others from knowing the hidden information but it is to keep others from thinking that the information even exists.

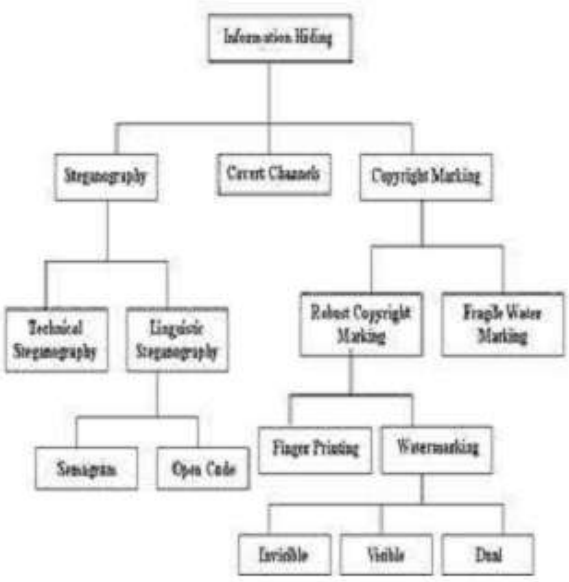


Figure 1: Information hiding techniques

Cryptography and Steganography are ways of secure data transfer over the Internet. Cryptography scrambles a message to conceal its contents; steganography conceals the existence of a message. It is not enough to simply encipher the traffic, as criminals detect, and react to, the presence of encrypted communications.

The encoding algorithm receives three inputs, the secret data to be embedded, the cover data, and an optional steganographic key. The algorithm then produces a stego cover that can be stored and/or transmitted. The decoding algorithm receives the stegocover and the (optional) stego-key, and extracts the secret data. In some algorithms, the decoder cannot actually extract the data and can only answer the question, "Are these data really embedded in the file being examined?" This makes sense in cases where the hidden data are a watermark, originally placed in the cover to prove ownership or simply for pride of ownership.

2. Hiding and Extracting the data

The encoding algorithm receives three inputs, the secret data to be embedded, the cover data, and an optional steganographic key. The algorithm then produces a stego cover that can be stored and/or transmitted. The decoding algorithm receives the stegocover and the (optional) stego-key, and extracts the secret data. In some algorithms, the decoder cannot actually extract the data and can only answer the question, "Are these data really embedded in the file being examined?" This makes sense in cases where the hidden data are a watermark, originally placed in the cover to prove ownership or simply for pride of ownership.

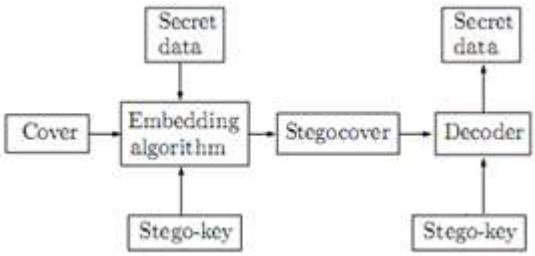


Figure 2: Data hiding and extracting

In a digital world, Steganography and cryptography are both intended to protect information from unwanted parties. Both Steganography and Cryptography are excellent means by which to accomplish this but neither technology alone is perfect and both can be broken. It is for this reason that most experts would suggest using both to add multiple layers of security.

3. Steganography Vs Cryptography

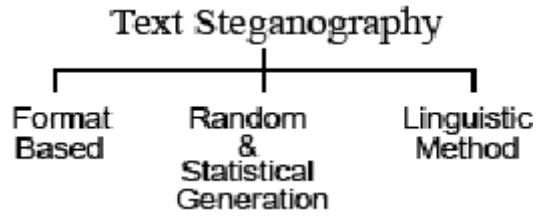
The term Steganography means, "cover writing" whereas cryptography means "secret writing". Cryptography is the study of methods of sending

Table 1: comparison of Advantage and disadvantage

Steganography	Cryptography
Unknown message passing	Known message passing
Little known technology	Common technology
Technology still being developed for certain formats	Most algorithms known to government
Carrier format	Strong algorithm are currently resistant to brute force attack Large expensive computing power required for cracking Technology increase reduces strength

4. Text Steganography

Steganography can beclassified into image, text, audio and video steganography depending on the cover media used to embed secret data. Text steganography can involve anything from changing the formatting of an existing text, to changing words within a text, to generating random character sequences or using context-free grammars to generate readable texts [4]. Text steganography is believed to be the trickiest due to deficiency of redundant information which is present in image, audio or a video file. 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 [3]. Unperceivable changes can be made to an image or an audio file, but, in text files, even an additional letter or punctuation can be marked by a casual reader [5]. Storing text file require less memory and its faster as well as easier communication makes it preferable to other types of steganographic methods [6]. Text steganography can be broadly classified into three types: Format based Random and Statistical generation, Linguistic methods.



Format-based methods [6]: This method uses the physical formatting of text as a space in which to hide information. Insertion of spaces or non-displayed characters, careful errors tinny throughout the text and resizing of fonts are some of the many format-based methods used in text steganography. Some of these methods, such as deliberate misspellings and space insertion, might fool human readers who ignore occasional misspellings, but can often be easily detected by a computer.

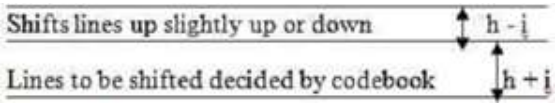
Random and statistical generation method [6]: Random and statistical generation methods are used to generate cover-text automatically according to the statistical properties of language. These methods use example grammars to produce cover-text in a certain natural language. A probabilistic context-free grammar (PCFG) is a commonly used language model where each transformation rule of a context- free grammar has a probability associated with it [6].

Linguistic methods [6]: Linguistic steganography specifically considers the linguistic properties of generated and modified text, and in many cases, uses linguistic structure as the space in which messages are hidden [2].

5. Techniques of the Text Steganography

Hiding information in plain text can be done in many different ways.

5.1 Format based method

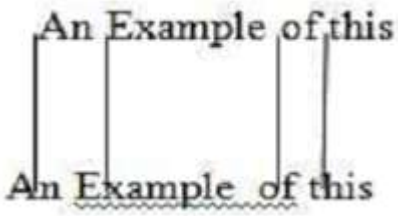


5.1.1 Line Shift

In this method, secret message is hidden by vertically shifting the text lines to some degree [7]. A line marked has two unmarked control lines one on either side of it for detecting the direction of movement of the marked line . To hide bit 0, a line is shifted up and to hide bit 1, the line is shifted down [8]. Determination of whether the line has been shifted up or down is done by measuring the distance of the centroid of marked line and its control lines [8]. If the text is retyped or if a character recognition program (OCR) is used, the hidden information would get destroyed. Also, the distances can be observed by using special instruments of distance assessment [7].

5.1.2. Word Shift

In this method, secret message is hidden by shifting the words horizontally, i.e. left or right to represent bit 0 or 1 respectively [9]. Words shift are detected using correlation method that treats a profile as a waveform and decides whether it originated from a waveform whose middle block has been shifted left or right [10]. This method can be identified less, because change of distance between words to fill a line is quitecommon. But if someone knows the algorithm of distances, he can compare the stego text with the algorithm and obtain the hidden content by using the difference. Also, retyping or using OCR programs destroys the hidden information [10].



Feature coding: In feature coding 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. Retying the text or using OCR program destroys the hidden information [17].

5.2 Random and statistical generation method:

Random and statistical generation is generating cover text according to the statistical properties. This method is based on character sequences and words sequences.

5.2.1. Word Mapping

This technique encrypts a secret message using genetic operator crossover and then embeds the resulting cipher text, taking two bits at a time, in a cover file by inserting blank spaces between words of even or odd length using a certain mapping technique. The embedding positions are saved in another file and transmitted to the receiver along with the stego object.

5.2.2. MS Word Document

In this technique, text segments in a document are degenerated, mimicking to be the work of an author with inferior writing skills, with secret message being embedded in the choice of degenerations which are then revised with changes being tracked. Data embedding is disguised such that the stego document appears to be the product of collaborative writing [3].

5.3 Linguistic methods

Linguistic method is a combination of syntax and semantics methods. Syntactic steganalysis is to ensure that structures are syntactically correct. In Semantic Method you can assign the value to synonyms and data can be encoded into actual words of text.

5.3.1 Syntactic Method: This technique uses punctuation marks such as full stop (.), comma (,), etc. to hide bits 0 and 1. But problem with this method is that it requires identification of correct places to insert punctuation marks [10, 11]. Therefore, care should be taken in using this method as readers can notice improper use of the punctuations [9].

5.3.2 Semantic method: This method uses the synonym of certain words thereby hiding information in the text. The synonym substitution may represent a single or multiple bit combination for the secret information. However, this method may alter the meaning of the text [11].

6. Inter-word Spacing and Inter- paragraph Spacing Approach

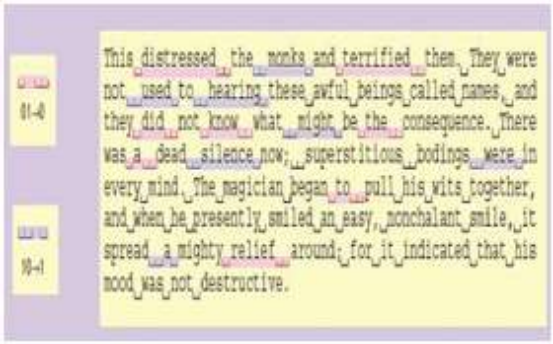


Figure 3: Inter word spacing

In inter-word spacing method, the secret message is embedded using the spaces between words. The inter-word spacing method utilises a single space between words to represent "0" bit and two spaces to represent "1" bit. In Manchester – encoding method is used to determine the inter-word spaces when concealing secret message bits to encoded like "01" and "10" is define as "1" bit and "0" bit respectively as "00" and "11" are not to be used for concealing any data in fig:3

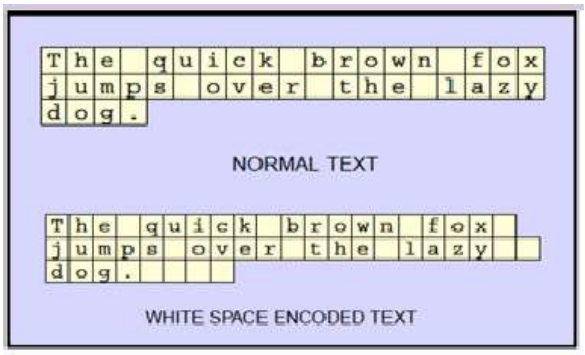


Figure 4: End-of – line spacing

In end-of-line spacing method, the secret message is hidden at the end of each line by inserting extra spaces. The end –of –line spacing allows two spaces to conceal 1-bitof secret message or four spaces to conceal 2-bits of secret message or eight spaces to conceal 3-bits of secret message and etc., fig : 4

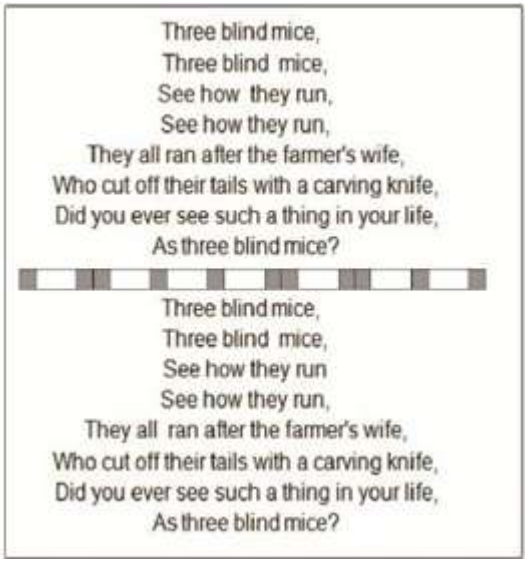


Figure 5: Inter-paragraph spacing

The inter-paragraph spacing method conceals the secret message by inserting space between two lines or an empty line between paragraphs in fig:5

7. Conclusion

Steganography although conceals the existence of a message but is not completely secure. It is not meant to supersede cryptography but to supplement it. This paper presents three approaches of text steganography.

Text Steganography methods, each method has respective capability to hide data in text. By using line shifting method, we can hide huge amount of data, but line shifting method only capable for printed text because in this method, other than printed textcharacter reorganization program(OCR) is used and hidden information get destroyed. Word shifting method is quite useful to hide data. In this method key term is algorithm made & used for word shifting. If this algorithm found by someone else than also security destroyed. Syntactic method used in wars to send very important information and hide very small amount of data. Like, we can use (.) and (,) in a poem and hide data in (0) as (.) and (1) as (,)

The future work should be focused towards optimizing the robustness of the decoding algorithm. This is because the hidden data will be destroyed once the spaces are deleted by some word processing software. Besides that, it is important to improve the capacity of the embedded scheme by taking other compression method into consideration.

References

[1] Nielsprovos and peter honeyman, University of Michigan, Hide and Seek: An Introduction to Steganography,IEEE Security & Privacy.
[2] N.F.Johnson. and S. Jajodia. Steganography: seeing the unseen. IEEE Computer, 16:26– 34, 1998.
[3] S. H. Low, N. F. Maxemchuk, J. T. Brassil, and L. O. Gorman, "Document marking and identificationusing both line and word shifting," *INFOCOM'95 Proceedings of the Fourteenth Annual Joint Conf. of the IEEE Computer and Communication Societies*, 1995, pp. 853-860.
[4] M. S. Shahreza, and M. H. S. Shahreza, "Text steganography in SMS," 2007 *Int. Conf. on Convergence Information Technology*, 2007, pp. 2260-2265.
[5] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol.35, pp. 313-336, 1996.
[6] I. Banerjee, S. Bhattacharyya, and G. Sanyal, "Novel text steganography through special code generation," *Int. Conf. on Systemics, Cybernetics and Informatics*, 2011, pp. 298-303.
[7] M. H. S. Shahreza, and M. S. Shahreza, "A new synonym text steganography," *Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing*, 2006, pp. 1524-1526.
[8] J. Cummins, P. Diskin, S. Lau, and R. Parlett, "Steganography and digital watermarking," School of Computer Science, 2004, pp.1-24.
[9] L. Y. Por, and B. Delina, "Information hiding- a new approach in text steganography," *7WSEAS Int. Conf. on Applied Computer and Applied Computational Science*, 2008, pp. 689-695.
[10] L. Y. Por, T. F. Ang, and B. Delina, "WhiteSteg- a new scheme in information hiding using text steganography," *WSEAS Transactions on Computers*, vol.7, no.6, pp. 735-745, 2008.
[11] K. Rabah, "Steganography-the art of hiding data," *Information Technology Journal*, vol.3, pp. 245-269, 2004.
[12] S. Bhattacharyya, I. Banerjee, and G. Sanyal, "A novel approach of secure text based steganography model using word mapping method," *Int. Journal of Computer and Information Engineering*, vol.4, pp. 96-103, 2010.
[13] B. Pfiztzmann, "Information Hiding Terminology." pp.347{350,ISBN 3-54061996-8, results of an informal plenary meeting and additional proposals, 1996.
[14] William Stallings, *Cryptography and Network Security: Principles and Practice 5/e.*, India, Prentice Hall, 2011.