Journal of Engineering and Applied Sciences 15 (2): 369-372, 2020
ISSN: 1816-949X

A Study of Text Steganography Methods

Fitra Chairil Akbar, Tito Waluyo Purboyo and Roswan Latuconsina
Department of Computer Engineering, Faculty of Electrical Engineering,
Telkom University, Bandung, Indonesia

Abstract: Based on the development of information and communication technology, then the security improvement will be more important. One of them is the level of security in inserting data or information. Increased security to insert data and information can be resolved using steganography technique. Steganography itself is a technique to hide messages into a digital medium. The digital media used in this steganography is text, picture, sound and video. In the text media, the methods and techniques used there are various. In this study, we will discuss the methods and techniques as well as comparing of each methods.

Key words: Steganography, data hiding, steganography text in text, media, information, security

INTRODUCTION

The development of technology at this time has led digital media into a common something. One of the developments in digital media is a digital text document. This digital document can facilitate the opening and browsing of contents in the history of the document was previously difficult to do on paper-based documents. The use of this digital document can also facilitate the delivery of internet media and will provide ease in performing the duplication of documents and facilitate the storage, if one day want to reopen. These conveniences will eventually be able to be exploit by certain parties negatively without regard to the copyright of the document. In such circumstances a technique is required to maintain the confidentiality of a digital text document, one of which is by using steganography techniques, steganography technique itself is the art and science of communicating by hiding the existence of the information itself by using certain media such as text, sound and images to make the information into another form.

Basic theory: This basic theory section will focus discusses the history of steganography, steganography on text, steganography methods.

HISTORY OF STEGANOGRAPHY

Such as cryptography, the use of steganography actually can used centuries ago even before the term steganography itself appears. The word steganography (steganography) is derived from the Greece, i.e., steganos means hidden or veiled and graph, meaning to write, so that, more or less means "write posts that are hidden or veiled". This technique include the myriad methods of communication to hide secret messages. This method including ink that doesn't look, microdots, setting words, digital signatures, hidden line and spectrum wide communication. Here is an example of the use of steganography in the past.

Steganography has known by the nation of Greece. Herodatus, ruler of Greece, sending secret messages by using the head of a slave or as media soldiers. In this case, the slave's hair is remove and secret messages written on scalp slave. When the slave's hair grows, the slaves sent the secret message behind her hair.

The Romans know steganography using an ink looks (invisible ink) to write a message. The ink made from a mixture of fruit juice, milk and vinegar. If the ink used to write then writing did not appear. Writing on paper can read by means of heating the paper.

The method used by the people of ancient Greece is by using a candle as the media hides their message. The message written on a sheet, the sheet and will be covered with wax to hide messages that you have written. The receiving party will then remove the wax from the sheet to see the message conveyed by the sender

Insert message: To insert messages either in text messages, images, sound and video input required in the form of digital files to be inserted messages, messages to be inserted message and key. That there are several examples of secret message insertion media used in steganography techniques, among others are (Gupta and Jain, 2015).

Corresponding Author: Fitra Chairil Akbar, Department of Computer Engineering, Faculty of Electrical Engineering, Telkom University, Bandung, Indonesia

Text: In the steganography algorithm that uses text as its insertion medium is usually used techniques Natural Language Processing (NLP), so that, the text that has been inserted secret messages will not be suspicious to the person who saw it.

Picture: The most commonly used image format because this format is one file format that is often exchange in the internet world. Another reason is the abundance of available steganography algorithms for the image container media.

Sound: Voice formats are often select because usually files with this format are relatively large. So can accommodate a large number of secret messages as well.

Video: The format of the video is indeed a format with a relatively large file size but is rarely used because of its size is too large, thus, reducing the practicality and the lack of algorithms that support this format. Steganography serves to hide the existence of messages and can be consider as a complement of cryptography that aims to hide the message content. In contrast to cryptography in steganography messages are hidden in such a way that the other party cant know of any secret message. Secret messages are not change to 'weird' characters like cryptography. The message is only hide into a media in the form of images, text, music or other digital media and looks like a normal message. To facilitate the process of concealment of text messages into the image, then designed a steganography application for message insertion. The application designed with three processes: taking images, adding messages to the image (encode image) and displaying extract messages in the image. This is criteria to be considered in data masking are:

Fidelity: The image quality of the container has not changed much. After the addition of secret data, the image of the steganography still looks good. Observers do not know that in the image there is secret data.

Robustness: The hidden data must withstand the manipulations performed on the container image (such as contrast modifications, sharpening, compression, rotation, image magnification, cropping, encryption, etc.). If the image is done image processing operations, then the hidden data is not damaged.

Recovery: The hidden data must be recoverable. Since, the purpose of steganography is hiding data, then at any time the confidential data within the container image must be retrievable for further use.

STEGANOGRAPHY TEXT

The steganographic mechanism begins with a secret message that will be hidden in the cover text by applying the embedding algorithm to generate the stego key, then the stego key will be sent to the communication channel to be sent to the recipient, to recover the secrets sent by the sender, the recipient must use the recovery algorithm which are given parameters by the stego key to extract messages. The stego key is also used to control the hiding process, so that, it can limit the detection/recovery of data sent to those who know it. Steganography in the text is the most difficult type of steganography because text files do not have a large scale information insertion capacity in comparison with other media. According Sharma *et al*. (2016) the structure of the document in the text media remains the same throughout the document, the data embedding on the text media in the note is the structure, if the data structure at the time of embedding changed, then all the meaning in the file will also change in other media the embedding process is done easily without any significant change in the output in question. There are 3 types of steganography in the text a (Por and Delina, 2008) are:

Format based methods: This method uses text as a place to hide information. Generally, this method modifies existing text to hide steganography text. Insertion of spaces between words or ends of sentences, deliberate misspellings and resizing of fonts throughout the text are some of the many methods used in the steganography of this text, in this method can not be seen by the human senses but will be quickly detected by the computer system.

Random and statistical generation methods: This method is based on the order of characters and word order. Concealment of information in random order of characters, this sequence should appear random to anyone who intercepts his message. The second approach to character generation is to take the statistical properties of the word frequency and frequency to make "words" seem to have the same statistical properties as actual words in a particular language. Hiding information in word order, actual dictionary items can be used to encode one or several bits of information per word using mapping logs between lexical items and bit sequences or words themselves can encode hidden information.

Linguistic methods: In this method consider the linguistic nature generated in the modified text, in many cases using linguistic structures as the space where the message is hidden in the word shift (Fig. 1).

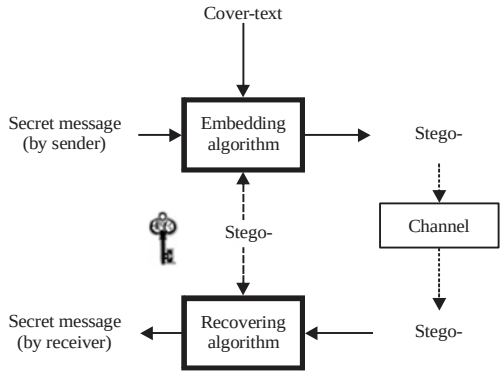

Fig. 1: Mechanism of text steganography

STEGANOGRAPHY METHODS

Steganography in the text can be classified according to the interests and purposes, in this study describes various types of steganography in terms of advantages and disadvantages (Koley and Mandal, 2016; Mandal *et al*., 2014; Por and Delina, 2008).

Line shifting: It is a suitable method for the type of text in print, vertical bit shift and has shortcomings if OCR (character recognition program) is enabled then hidden information or messages will be lost (Fig. 2).

Word shifting: It is a suitable method for the type of text in print, horizontal bit shifts and have shortcomings if the Character Recognition Program (OCR) is enabled then hidden information or messages will be lost and methods associated with the shift of words so easy in getting the hidden data (Por and Delina, 2008) (Fig. 3).

Syntactic method: A steganography method that is easy to hide information has the disadvantage that if there is a shrewd reader can easily find hidden messages. In this method of hiding information by changing punctuation such as point (.) and comma (,) in the appropriate position, the possibility of detection of hidden messages by the attacker is minimal because the attacker will not be interested if the syntactical method is used correctly (Koley and Mandal, 2016; Singh *et al*., 2014).

Semantic based hiding: Methods that can not be detected by retyping/using OCR programs have a weakness, if there is a shrewd reader will find easily synonyms/antonyms that exist in the document (Sharma *et al*., 2016) (Table 1).

Abbreviation based hiding: According to Sharma *et al*. (2016) in this method the messages are inserted in hidden abbreviations. Secret messages can be hidden in a few kilobytes file (2016) (Table 2).

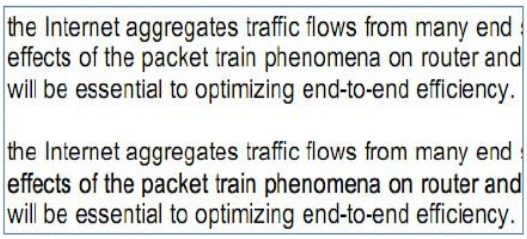

Fig. 2: Line shift method

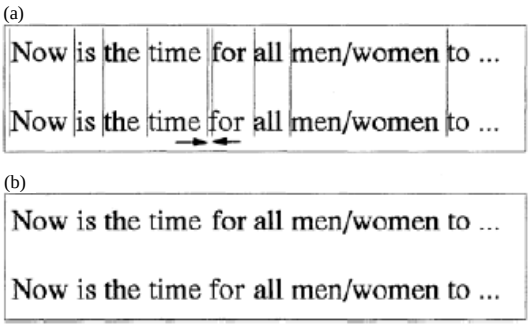

Fig. 3: Word shift method

Table 1: Semantic method

Words	Synonyms
Lazy	Idle
Hard	Difficult
Unhappy	Sad

Table 2: Abbreviation method

Acronym	Words
ID	Identification
DOB	Date of birth
ASAP	As soon as possible

Table 3: Word spelling method

American English	British English
Airplane	Aeroplane
Fiscal	Financial
Unalike	Unlike

Hiding data using white spaces: Methods that work to hide data in white space because white space is usually ignored and also usually predominantly white text editor, the deficiency in this method is if this method is applied in large text will be very space consuming, so can be suspected that in the white row there is hidden information.

Hiding data in paragraph: It is a method that works to hide a message using the beginning and end letters.

Change of spelling: It is a method used to embed secret messages in text files, secret messages are inserted by representing the same but different words in spelling (Table 3).

Table 4: Advantages and disadvantages of steganography text methods

Methods	Advantages	Disadvantages
Line-shift	Is a suitable method for the type of text in print	Have less message insertion capacity on the document. When the OCR Program runs the information on the document becomes corrupted and lost
Word-shift	Is a suitable method for the type of text in print has a larger information insertion capacity than the line-shift method	The worse endurance level compared to the line shift method, if there is noise in the document, then the extraction process in this method is not good and there is an undetectable secret message on the extraction process
Syntactic method	Steganography methods are easy to hide information, information is saved by changing punctuation positions such as dots (.) and commas (,)	If there is smart reader, then the information inserted in the message will soon be found
Semantic method	In this steganography method, hidden information can not be detected by the the OCR program	Smart readers can find hidden messages through synonyms or antonyms residing in documents
Abbreviation	In this steganography method there are abbreviations as a secret message	Can only insert secret messages with small size
Hiding data using white spaces	Hides secret messages in white rungs in paragraphs because white space is usually ignored by readers	If this method is applied to the document by consuming a large white space then this method can be suspected by the reader
Hiding data in paragraph	This method works by hiding secret messages using the first letter and the final letter in each sentence	If data or secret messages will be hidden in a paragraph many, eating will be a challenge to insert the secret message
Feature coding	Insert secret information in words by changing the font shape, font size and color of the text	A smart reader will easily find out hidden messages hidden in words because there are words with prominent differences such as font size different from others, fonts are different from others and colors different from others

Feature coding: Is a steganography technique that deals with feature changes in the text, by modifying the message that produces the cover text. features that are intended here are the text, text color, text font. when the feature on the text is changed, only the sender and the recipient can detect the message (Koley and Mandal, 2016; Singh *et al*., 2014).

Advantages and disadvantages: Table 4 shows the advantages and disadvantages of steganography text methods.

CONCLUSION

All methods of steganography in text media, each method has the characteristics that are reviewed from the advantages and disadvantages in hiding information in the text as well as the reliability of each method steganography refers to the implementation of the method is suitable or not suitable in solving problems that arise in daily life-day. In this study is expected to help overcome the problem of concealment of confidential information by using steganography methods and techniques in the media of the text, so that, problems can be resolved.

REFERENCES

Gupta, S. and R. Jain, 2015. An innovative method of Text Steganography. Proceedings of the 2015 3rd International Conference on Image Information Processing (ICIIP'15), December 21-24, 2015, IEEE, Waknaghat, India, pp: 60-64.

Koley, S. and K.K. Mandal, 2016. A novel approach of secret message passing through text steganography. Proceedings of the International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), October 3-5, 2016, IEEE, Paralakhemundi, India, ISBN: 978-1-5090-4621-8, pp: 1164-1169.

Mandal, K.K., A. Jana and V. Agarwal, 2014. A new approach of text Steganography based on mathematical model of number system. Proceedings of the 2014 International Conference on Circuits, Power and Computing Technologies (ICCPCT'14), March 20-21, 2014, IEEE, Nagercoil, India, pp: 1737-1741.

Por, L.Y. and B. Delina, 2008. Information hiding: A new approach in text steganography. Proceeding of the 7th WSEAS International Conference on Mathematics and Computers in Science and Engineering, April 6-8, 2008, World Scientific and Engineering Academy and Society, Hangzhou, China, ISBN:978-960-6766-49-7, pp: 689-695.

Sharma, S., A. Gupta, M.C. Trivedi and V.K. Yadav, 2016. Analysis of different text steganography techniques: A survey. Proceedings of the 2016 2nd International Conference on Computational Intelligence and Communication Technology (CICT), February 12-13, 2016, IEEE, Ghaziabad, India, ISBN:978-1-5090-0210-8, pp: 130-133.

Singh, H., A. Diwakar and S. Upadhyaya, 2014. A novel approach to text steganography. Proceedings of the 2014 1st International Congress on Computer, Electronics, Electrical and Communication Engineering (ICCEECE'14), March 17-18, 2014, Chennai, India, pp: 7-12.