Journal of Engineering and Applied Sciences 14 (12): 4004-4009, 2019
ISSN: 1816-949X
© Medwell Journals, 2019

# A Survey on the Use of Text Steganography Practical Methods

Sahrul Alam Sukma, Tito Waluyo Purboyo and Roswan Latuconsina
Department of Computer Engineering, Faculty of Electrical Engineering, Telkom University,
Bandung, Indonesia

**Abstract:** Digital media is very popular now a days. Digital media is an absolute necessity because of the technological advances currently not escape the role of digital media. Digital media is widely used because it is easy to use and easily duplicated, so that, it can be accessed at any time if needed. However, from the convenience there are always people who abuse it. For example, is the emergence of an intruder to take confidential information. This shows that the vulnerability of digital media to attack. That requires a technique that can protect a document from crime, one using steganography. Steganography is currently widely used to protect and transmit secret messages. This study discusses the use of steganography in a variety of methods. However, there are still some steganography methods that lack attention. Surely this is a challenge for researchers to continue to innovate develop steganography.

**Key words:** Digital media, steganography, duplicated, vulnerability, emergence, convenience

## INTRODUCTION

The rapid development of technology to make the means of information systems also experienced growth. Media information is no longer just use the print media but now is the moment in which human needs fixed on digital media. Digital media has always been very important for today's modern man. Without digital media, technological developments would not be so soon. But if digital media is not equipped with the security, the digital media would be highly susceptible to misuse of data or information for the user can freely duplicate and modify the information. Information hiding has many types. Among them are convert channel, steganography, anonymity and copyright marking. In addition, cryptography is included in the information hiding. Much research has been done to do a combination of steganography and cryptography. Steganography and cryptography have different principles but stick to the same purpose that conceals secret messages. Steganography is created, so that, the inserted message is not detected by the human sense system. While cryptography has a principle that the message cannot be read but still known by the human senses. Figure 1 is information hiding classification.

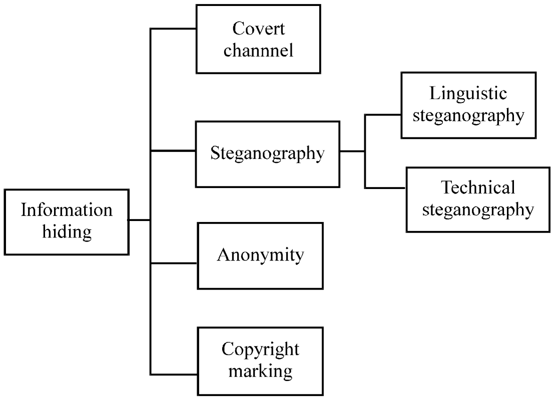

Fig. 1: Classification information hiding

**Literature review:** As it is known that steganography is growing. The implementation is even more varied. Rajeev and others using text steganography to hide messages in an email address. They use a combination of methods LZW+BWT+MTF to increase the capacity of hiding. The function of the BWT+MTF for improve data correlation and LZW function is to increase the capacity of hiding. The workings of this insertion is to insert random characters before the '@' in addition to the number of characters in the email are also influential in the insertion process. Hiding capacity is the main parameter used. The results of this new method is that this method has a better performance than other methods in terms of the capacity of hiding (Kumar *et al*., 2014).

Rajeev *et al.* conducted a study back on steganography text in the email. This time they used a method Huffman. They optimize the use of the number of characters in the email address. To make optimal character, they add a new character to indicate

**Corresponding Author:** Sahrul Alam Sukma, Department of Computer Engineering, Faculty of Electrical Engineering, Telkom University, Bandung, Indonesia

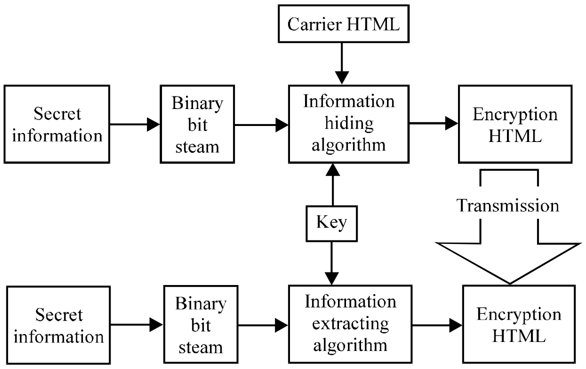

Fig. 2: Webpage information hiding model

confidential data placed before the symbol '@'. It is proven to increase the capacity of concealment. The parameters used to measure the performance of this method is the capacity parameter concealment. This method was compared with another popular method in the case of concealment. The result is evident that by using this method because the performance is more like the optimal character (Kumar *et al*., 2016).

New research by Shi *et al.* (2016) on text steganography based search implemented on the internet. In the study explained that the text is divided into two, namely text structured and unstructured text. Examples of structured text is HTML, DOCX and PDF. While the examples of unstructured text is a dialogue, novels and articles. In a previous study conducted (Huang *et al*., 2008), the concealment of a structured text on a web page that is based on attributes of permutations. The bits that have been converted and then inserted the message as usual. Additionally, also did something similar, namely the insertion of a message on a web page using a dictionary of attributes. The first step is choose dictionary first, then, the conversion, then, insert with a secret message. Figure 2 describes webpage information hiding model.

The latter who hide confidential information on a web page by using Tag and CSS attribute substitution. This method is used to change the color or font size, so that, the reader does not realize that there is a secret message.

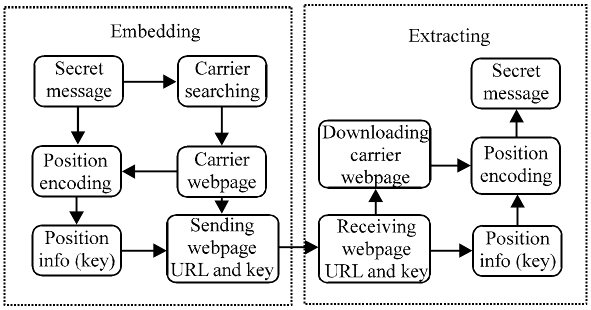

Fig. 3: Steganography framework based on webpages

Table 1: Example hiding information mappings

| Secret | Tag settings |
|---|---|
| 0000 | <font size = 4 style= "margin: 10px;"> |
| 0001 | <span style = "font-size:1.13em; margin:10px;"> |
| 0010 | <span style = "font-size:113%; margin: 10px;"> |
| 0011 | <span style = "font-se:18.08px; margin:10px;"> |
| 0100 | <font size = 4 style = "margin: 10px 10px;"> |
| 0101 | <span style = "font-size:1.13em; margin:10px 10px;"> |
| 0110 | <span style = "font-size:113%; margin:10px 10px;"> |
| 0111 | <span style = "font-size:18.08px; margin:10px 10px;"> |
| 1000 | <font size = 4 style = "margin: 10px 10px 10px;"> |
| 1001 | <span style = "font-size:1.13em: margin:10px 10px 10px;"> |
| 1010 | <span style = "font-size:113%; margin:10px 10px 10px;"> |
| 1011 | <span style = "font-size:18.08px; margin:10px 10px 10px;"> |
| 1100 | <font size = 4 style = "margin: 10px 10px 10px 10px:"› |
| 1101 | <span style = "font-size:1.13em; margin:10px 10px 10px 10px;"> |
| 1110 | <span style = "font-size:113%; margin:10px 10px 10px 10px;"> |
| 1111 | <span style = "font-size:18.08px; margin:10px 10px 10px 10px;"> |

Table 1 is an example data embedding for hiding information with CSS. Back to research conducted (Sharma *et al*., 2016) they use unigram and probability distribution obtained by counting characters in a web page. The test results showed that the value of the capacity of concealment of 10.32%. In the visible, the URL has been inserted the message looks like a URL in general, so, it can be said this method has a pretty good concealment. This method is used to change the color or font size, so that, the reader does not realize that there is a secret message.

The prerequisite of the proposed model is that the web exists and can be found by search engines like Google. Further research is expected to the researchers to develop an efficient position to increase concealment capacity.

Figure 3 describes steganography framework (Sharma *et al*., 2016). Little research has been conducted on the linguistic-based text steganography synonymous substitution (Chang and Clark, 2010). In fact, synonymous substitution is a major transformation of linguistic steganography. On that basis they approach the synonymous substitution with the help of Google N-gram 1T to check synonyms and Lexical SemEval to evaluate the method. They developed words into a node in the graph are connected by each end and a specified bit of each word is determined by the algorithm vertex. This method produces a unique bit without reducing the synonyms large that maintaining the capacity of concealment and remain reasonable.

Ammar *et al.* explain that steganography is usually hidden by changing the form of a file or inserted through other files. Free steganography to conceal the message

Table 2: Summary of text steganography practical use method

| Researcher’s name | Methods | Practical use |
|---|---|---|
| Rajeev *et al.* in 2016 | Combination BWT + LZW + MTF | E-mail |
| Shangwei and Huang in 2016 | Hiding information based on attributes substitution | Webpage |
| Huajun *et al.* in 2008 | Hiding information based on attributes permutation | Webpage |
| Ren and Wang in 2012 | Hiding information based on css attribute setting | Webpage |
| Chen and Liao (2014) | Hiding information based on tag dictionary | Webpage |
| Chang and Clark (2010) | Synonym substitution | Collaboration with Google N-gram 1T and SemEval Lexical |
| Ammar *et al.* in 2014 | Text steganography | Applied to the machine |
| Kundur and Ahsan (2003) | Internet steganography | Internet protocol |

largely implemented in software is not fast enough and in real time. Study of Ammar *et al.* present a steganography technique is implemented in hardware. The result steganography reached 11.27 Gbps speeds. Testing is done by hiding the data in a text file (Table 2).

The next practice is done which is concealment of data using IP. Steganography is the internet is exploitation elements and IP to gather confidential information. In his study on the internet mention that the steganography received less attention. In fact, the volume of traffic on the internet has a large enough bandwidth for confidential communications. There are two approaches made by Deepa and Kamran namely the approach of packet header manipulation and sorting approach uses packet sequence number field in IPSec. Scenario testing conducted to provide facilities for exploration.

## MATERIALS AND METHODS

**Basic theory:** This basic theory will focus on the history of steganography, steganography, steganography text and practical use of text steganography.

**History of steganography:** Steganography was first discovered by Johannes Trithemius (1462-1516) who was a German abbess wrote a book titled “Steganography: Arts through hidden posts that require restoration of the human mind”. This research is considered as a method to communicate with spirits (Judge, 2001). When it is used as a warning invasion steganography which was often painted timber under the pedestal candle. Casual people thought that it was empty timber without writing (Raggo, 2001). The Greeks and ancient Romans used to use the slave to insert secrets through the shave of his hair. Then the Romans also recognize steganography by using transparent ink to write a secret message. In addition, the Germans use steganography with microdots method. Microdots are very small photos and stored on envelopes. This method is often not realized by people. Then look in the mirror of recent history, steganography method was used during World War 2 by the Nazi Army. Steganography comes from the Greek meaning “Covered Writing” (Agarwal, 2013). Currently, steganography experiencing growth. By relying on technology, steganography venturing into the digital world. Figure 4 explains the steganography classification.

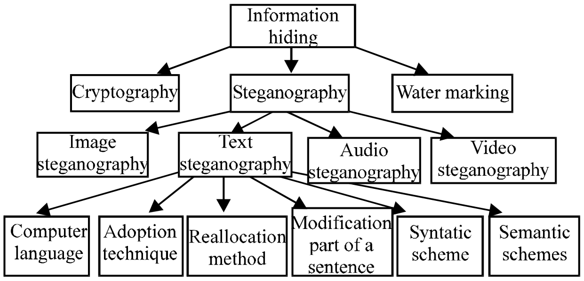

Fig. 4: Steganography classification (Koley and Mandal, 2016)

## RESULTS AND DISCUSSION

**Steganography:** Steganography is a technique to hide secret information so that the information is not known by the third party or unwanted party. Besides steganography, cryptography is also an information hiding technique but different purposes. Cryptography intended that messages can not be read while steganography intended that the message can not be seen by the human senses (Uddin *et al.*, 2014). Steganography is the concealment of information that is drilled into a cover (Johri *et al.*, 2015). As a result of the development technology, cover not only text but also images, audio and video. In general, steganography is described by Fig. 5. Conducted a survey of steganography media are as follows (Johri *et al.*, 2016).

**Text steganography:** Concealment of information using text media used by changing the format that is inside a text document. Steganography text is divided into 2, namely soft copy and hard copy. In soft copy steganography, the message is inserted by changing the number of spaces after the punctuation. Usually the use of tabs and space becomes the most effective thing in hiding information. While the example of hard copy is the method of line shift and word shift.

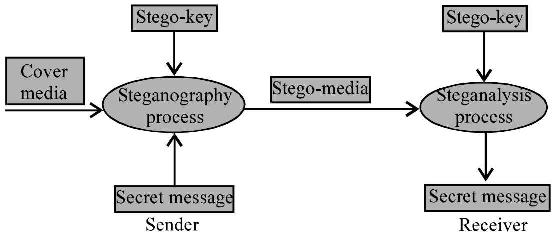

Fig. 5: Steganography process (Al Sadi, 2015)

**Image steganography:** Digital image is a combination of low and high frequency content. Insertion of messages using pictures is closely associated with the pixel. Image Steganography is the most widely used media of research, because the challenge of image steganography is easily disseminated through the web or forum. If the image file is formatted to a different format, the hidden information will be lost.

**Audio steganography:** Inserting a message in audio steganography is more complicated than in the message in the picture. There are several methods used in audio steganography include LSB coding, coding parity and echo data hiding. Care must be taken to create an audio steganography because it is easily damaged.

**Video steganography:** Concealment of the video is a combination of text, audio and images. Concealment process through a video message is to perform the separation of the video frame first. After masking and filtering and then select the area to be inserted message. This method is the most difficult to attack. Video steganography is very similar to steganography on the image but on video steganography, messages are inserted through the frame pieces. The video must be designed in such a way that the result does not look like it has been modified. Steganography video can use the method applied to steganography image.

In steganography, there are 3 important terms that are always used. Among others are:

**Cover object:** Cover object is used as a media message insertion. Can be text, audio, images and video.

**Secret data/message:** Is a secret message to be conveyed to the recipient. This message is planted on a cover object.

**Stego object:** Stego object is an output of a cover object that has been inserted a secret message.

**Stego key:** This is the secret key used to extract the message back. This key can only be known by the sender and recipient of the message.

**Steganalysis:** Steganalysis is a process to detect the existence of secret messages in a file. According study (Munir, 2015), there are several criteria that must be met, so that, steganography can be spelled out well. Among others are:

**Imperceptible:** That is a steganography must be made in such a way that the secret message can not be known visually or audio.

**Fidelity:** The quality of the object that the message has inserted must be the same as the object before the message is inserted.

**Recovery:** Messages that have been inserted should be retrievable.

**Capacity:** As much as possible the size of the inserted message must be large.

**Text steganography:** Text steganography is a technique most difficult due to the size of a text file does not have a large redundancy when compared to media audio, image and video (Sharma *et al.*, 2016).

Based on the study of Por *et al.* (2008) there are three basic categories of text steganography method are as follows.

**Format based:** Text-based format using physical modifications, for example, enter a space and change the font size.

**Random and statistical generation:** The workings of this type is to embed a message based on a sequence of words and characters. Concealment of information through a sequence of characters to be done at random until the reader does not realize that there is a message in these characters. In the concealment of information through a sequence of words, we can use a dictionary that can be used to create one or several bits of code.

**Linguistic steganography:** This mode uses Natural Language Program (NLP) for inserting a message. This method of paying particular attention to the linguistic nature of a text is generated and then be modified.

**Technique of steganography text:** Of the three basic methods above, then, we will discuss briefly the

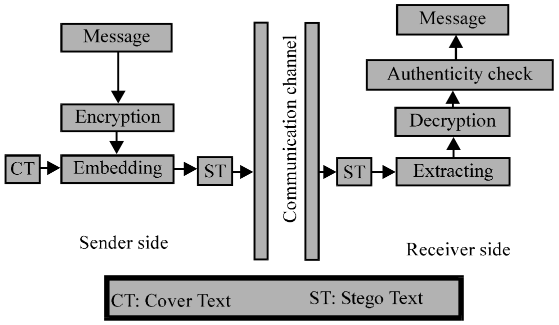

Fig. 6: Steganography text process (Saraswathi and Kingslin, 2014)

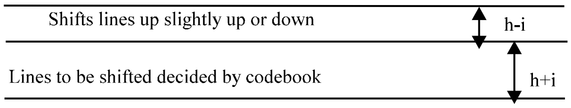
Shifts lines up slightly up or down
h-i
Lines to be shifted decided by codebook
h+i

Fig. 7: Line-shift method (Saraswathi and Kingslin, 2014)

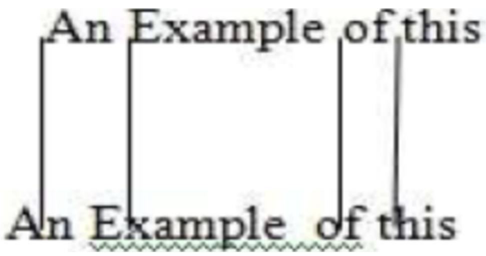

Fig. 8: Word-shift method (Saraswathi and Kingslin, 2014)

steganography techniques. There are several techniques that can be used to protect confidential messages that have different characteristics. Figure 2 is an explanation of the text steganography scheme.

Review on the study of Shirali and Shahreza (2008) about steganography techniques that have been used today. Here is the explanation.

**Line shift and word shift:** Line shift is a concealment technique a message by sliding rows vertically. While the shift is concealment of the message word by sliding horizontal lines. Both methods are suitable for use in printed documents. But if the document is altered using Programs Character Recognition (OCR), the stored information will be destroyed. Figure 6 illustrates the line shift method and Fig. 7 explains the word shift (Fig. 8).

**Sintatic method:** This technique inserts a message into punctuation dot (.) Or comma (,). Users who use this technique must put in place the appropriate punctuation, so that, the unsuspecting reader.

Table 3: Semantic method example (Sharma *et al*., 2016)

| Word | Synonym |
|---|---|
| Lazy | Idle |
| Hard | Difficult |
| Unhappy | Sad |

Table 4: Text Abbreviation method example (Sharma *et al*., 2016)

| Acronym | Word |
|---|---|
| ID | Identification |
| DOB | Date of birth |
| ASAP | As soon as possible |

**Semantic method:** The workings of this technique is to hide a secret message into a word synonymous. The advantage of this technique is resistant to OCR program. However, this technique could change the meaning of the posts (Table 3).

**Text abbreviation:** This technique of hiding information on the abbreviations. With this technique is not a lot of messages that can be inserted (Table 4).

## CONCLUSION

The conclusion from all that has been discussed is steganography experiencing rapid growth as a result of progress technology. Lots of people use steganography to hide messages and other needs like data security. Various practice of steganography has been done. But of all the practices that have been made there are still a lot of development steganography less attention. Surely this is a challenge for the researchers to continue to develop steganography.

## REFERENCES

Agarwal, M., 2013. Text steganographic approaches: A comparison. Intl. J. Network Secur. Appl., 5: 91-106.

Al Sadi, G., 2015. Image steganography approach. Intl. J. Comput. Sci. Mob. Comput., 4: 166-169.

Chang, C.Y. and S. Clark, 2010. Practical linguistic steganography using contextual synonym substitution and vertex colour coding. Proceedings of the 2010 International Conference on Empirical Methods in Natural Language Processing (EMNLP '10), October 09-11, 2010, ACM, Cambridge, Massachusetts, USA., pp: 1194-1203.

Huang, H.J., S.H. Zhong and X.M. Sun, 2008. An algorithm of webpage information hiding based on attributes permutation. Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, August 15-17, 2008, Harbin, China, pp: 257-260.

Johri, P. and A. Kumar, 2015. Review paper on text and audio steganography using GA. Proceedings of the International Conference on Computing, Communication and Automation (ICCCA), May 15-16, 2015, IEEE, Noida, India, ISBN:978-1-4799-8889-1, pp: 190-192.

Johri, P., A. Mishra, S. Das and A. Kumar, 2016. Survey on steganography methods (text, image, audio, video, protocol and network steganography). Proceedings of the 3rd International IEEE Conference on Computing for Sustainable Global Development (INDIACom), March 16-18, 2016, IEEE, New Delhi, India, ISBN:978-1-4673-9417-8, pp: 2906-2909.

Judge, J.C., 2001. Steganography: Past, Present, Future. SANS Institute Publication, North Bethesda, Montgomery,.

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.

Kumar, R., A. Malik, S. Singh and S. Chand, 2016. A high capacity Email based text steganography scheme using Huffman compression. Proceedings of the 3rd International Conference on Signal Processing and Integrated Networks (SPIN), February 11-12, 2016, IEEE, Noida, India, ISBN:978-1-4673-9198-6, pp: 53-56.

Kumar, R., S. Chand and S. Singh, 2014. An Email based high capacity text steganography scheme using combinatorial compression. Proceedings of the 5th International Conference on the Next Generation Information Technology Summit (Confluence), September 25-26, 2014, IEEE, Noida, India, ISBN:978-1-4799-4237-4, pp: 336-339.

Munir, R., 2015. Steganography. STEI-Intstitut Teknolog Bandung, Bandung, Indonesia,.

Por, L.Y., T.F. Ang and B. Delina, 2008. Whitesteg: A new scheme in information hiding using text steganography. WSEAS. Trans. Comput., 7: 735-745.

Raggo, M.T., 2001. Steganography, steganalysis and cryptanalysis. VeriSign, Virginia, USA.

Saraswathi, V. and M.S. Kingslin, 2014. Different approaches to text steganography: A comparison. Intl. J. Emerging Res. Manage. Technol., 3: 124-217.

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.

Shi, S., Y. Qi and Y. Huang, 2016. An approach to text steganography based on search in internet. Proceedings of the International Symposium on Computer (ICS), December 15-17, 2016, IEEE, Chiayi, Taiwan, ISBN:978-1-5090-3439-0, pp: 227-232.

Shirali, S.M.H. and M.S. Shahreza, 2008. A new synonym text steganography. Proceedings of the 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'08), August 15-17, 2008, IEEE, Yazd, Iran, ISBN:978-0-7695-3278-3, pp: 1524-1526.

Uddin, M.P., M. Saha, S.J. Ferdousi, M.I. Afjal and M.A. Marjan, 2014. Developing an efficient solution to information hiding through text steganography along with cryptography. Proceedings of the 9th International Forum on Strategic Technology (IFOST), October 21-23, 2014, IEEE, Cox's Bazar, Bangladesh, ISBN:978-1-4799-6060-6, pp: 14-17.