

Steganography on Text using Word-Shift Coding and Centroid Methods

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

Key words: Steganography, word-shift, centroid, inserted, enhancement

Corresponding Author:

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

Page No.: 3095-3100 Volume: 15, Issue 16, 2020 ISSN: 1816-949x Journal of Engineering and Applied Sciences Copy Right: Medwell Publications

INTRODUCTION

The development of technology at this time digital media cause into something familiar, 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 which 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 we want to reopen it in future. This convenience will eventually be able to be exploited by certain parties negatively without regard to the copyright of the documents. In such circumstances techniques are needed to maintain the confidentiality of digital text documents, growing, so, security enhancement will also be more important. One of the factors is the level of security in hiding data/information. It is known that steganography method is used in to hide an information in digital media. Digital media that is used in steganography are Text, Picture, Audio and Video. In this study, we discuss one of the techniques of steganography on the text media. This technique is a Word-shift technique where the secret message is converted into a series of bits embedding process into the document file by shifting spaces between words. The process of shifting to the right between words occurs when the input bit is "1". If the input bit is "0" then between the words will shift to the left. After the secret message is inserted then it will be extracted from the document in the printing and scanning process. In the process of extracting this secret message, we use the centroid method to detect the center of the image and then get the secret message that has been inserted.

Abstract: Information and communication technology is

one of which is by using steganography. Steganography has a definition of art and science of concealment of that information or message by using certain media such as text, sound and images to make information into another form^[1]. In this study will conduct research on inserting messages in text media steganography using word-shift technique and as a method of extracting the message using centroid method.

BASIC THEORY

Steganography itself is a technique to hide messages by way of sender and receiver who knows the message hidden in it. The need for steganography in extending in the world war, the use of steganography in the end increased because cryptography is no longer used which became the difference between cryptography and steganography in terms of code cipher generated as a cryptographic algorithm and code cipher can be read by others because it can attract attention to others. On the other hand, steganography is a technique for generating cipher codes, hiding information by means of hidden secret messages that seem plain and unattractive. using steganographic secret messages will be hidden inside the cover text, stego media is a transmission line where the stego object will be transmitted^[1, 2]. According Siti Rohayah, to insert both text messages, images, sound or video in need of steamy input digital files that will be inserted in the message, the digital file is a message and key^[3].

According Por *et al.*^[4] in this journal, this is a medium that is needed to make a concealment message on steganography are:

Text: In steganography on text media, the usual technique used is the NLP technique so that text on documents that have been inserted secret messages will not be suspicious.

Image: Steganography media that is often used is a picture format, because the image format is a file format that is often used and exchanged in the internet in addition steganography algorithm for image media.

Sound: on the sound media is often select because the usual sound format is used and the size of the sound format is relatively large, so it can accommodate a large number of secret messages.

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 (Fig. 1). This is criteria according Rinaldi^[5] to be consider 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.

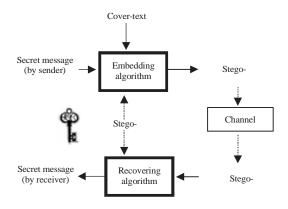


Fig. 1: Mechanism os text steganography

The mechanism of steganography text begins by hiding the secret message into the cover document and applying the embedding algorithm to generate the Stego key. Then, Stego Key will sent through the communication channel to the recipient, then to get the message that has been inserted by the sender of the message, the recipient must be the recipient must use the recovery algorithm also Stego Key to be able to extract hidden messages that have been hidden. Stego Key can also be used to control the process on the concealment of messages as to 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 information insertion capacity in comparison with other media. 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. other media the embedding process is done easily without any significant change in the output in question^[6]. There are 3 types of steganography in the text according to Sharma et al.^[6].

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

Random and statistical generation methods: This method based in the order of characters and word order. Concealment of information in random order of characters, this sequence should appear random to anyone

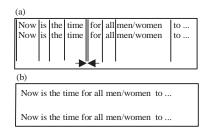


Fig. 2(a, b): Word shift method

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 use 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 methods, consider the linguistic nature generated in the modified text, in most cases found in this method, the linguistic structure as a space that can be used to hide messages in the word shift.

Word shifting: Acoording Por *et al.*^[4] this is a suitable method for the type of text in print, horizontal bit shifts and have shortcomings if the OCR (character recognition program) 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^[7]. Using this method, secret messages are hidden by horizontally shifting words, i.e., left or right and representing bits 0 and 1, respectively. Hiding an information in the form of text can be done by using the appropriate method, each text document has a range of words that vary. This method is judged to be lacking, due to changes in spaces between words to fill in enough lines. if anyone knows the distance algorithm that can be used to compare Stego Text with Cover text and also extract it to get information hidden inside the text. When OCR is applied then the hidden information/messages will be destroyed^[1].

CENTROID METHODS

Centroid has a central definition of a large number of projected pixel profiles. The profile in this centroid is the projection of 2-dimensional array into 1-dimensional array. According to Singh *et al.*^[2] this method computes each profile and calculates the distance between centroids on 2 successive rows. Centroid method can be formulated:

$$c_{i} = \frac{\sum_{bi}^{ei} yh(y)}{\sum_{bi}^{ei} h(y)}, i = 1, 2, ...$$

Where:

c_i = Centroid on line i y = The length of the array h(y) = Profile on the length of the y-array

In the word-shift technique, the centroid will search for the bit value in words in each line that has been inserted a secret message. Centroid on this word-shift technique will search for the middle value of each line, according to the detection of centroid method in that row which has been shifted to the right containing bit 1, if on the left-handed row contains bit 0 in accordance with the steganography technique used . Collection of bits that get converted into actual message. In image the process of feeding digitization will produce an array of length L and width W. The array element is:

$$f(x,y)$$
 where $x = 0, 1, 2, ..., W$, where $y = 0, 1, ..., L$

Function f(x, y) represents each pixel of an image at position (x, y). If f(x, y) is 1 then the position (x, y) is black and vice versa. Below is an example of an array of digitized results^[8]:

fx0	fx1	fx2	fx3	fx4	fx5	
0	0	0	0	1	1	fx0
0	1	1	0	0	0	fx1
1	0	0	0	1	1	fx2
0	0	1	0	1	1	fx3
0	1	0	0	0	0	fx4
0	0	1	1	0	0	fx5

Profile is a projection of 2 dimensional array into 1 dimension with the formula as:

$$v(x) = \sum_{y=t}^{b} f(x, y), x = 0, 1, ..., W$$

Calculate the profile value in column x1 or v(1):

- v(1) = f(1, 0) + f(1, 1) + f(1, 2) + f(1, 3) + f(1, 4) + f(1, 5)
- v(1) = 0 + 1 + 0 + 0 + 1 + 0
- v(1) = 2

Experiment using word shift and centroid for manual extraction: In this chapter, we will discuss about the experiment of insertion of secret messages using the word shift technique with the test parameters (Table 1 and 2).

J. Eng. Applied Sci., 15 (16): 3095-3100, 2020

Table 1: Parameter	r	
Message	Size (bytes)	No. of line
Hai	5	22
Kamu	6	22
Kannu	0	22

Table 2: Ascii to binary conversion						
Message	Binary code					
Hai	0100 1000 0110 0001 0110 1001					
Kamu	0100 1011 0110 0001 0110 1101 0111 0101					

Cover text	->	Embedding	┢	Recover

Fig. 3: Experiment scenario

3.3: EXperiment scenario
3.3: Experiment scenario
3.5: Experiment scenario
3.5: Experiment scenario
3.6: Application of the second scenario of the scenario Sukarno was popularly referred to as bung ("older brother"), and he painted himself as a man of the people carrying the aspirations of Indonesia and one who dared take on the West. He nstigated a number of large, ideologically driven infrastructure and monuments celebrating indonesia's identity, which were citicised as substitutes for a real development in a deteriorating

Fig. 4: Original document

The following is a test scenario in this experimental paper, starting from converting the message to a series of bits 0 and 1, then the message is inserted using steganographic word shift technique on the document with the number 22 lines, then the message will be printed, after the printed document will enter into the scanning phase of the document and extraction on the image format. bmp by using centroid method in review of the image profile (Fig. 3 and 4).

Embedding: First prepare the documents that will be inserted a secret message. Then the message will be inserted each character in the conversion into binary code as in Table 2. After insertion with the secret message "Hi" The new text document as in Fig. 5 and 6.

And under his increasingly authoritarian rule. Indonesia moved on a course of stormy nationalism. Sukamo was populary referred to as *burg* ("older brother"), and he painted hinself as a man of the people carrying the aspirations of Indonesia and one who dared take on the West. He instgated a number of large, ideologically driven infrastructure and monuments celebrating indonesia's biently, which were englicised as substitutes for a real development in a deteriorating

Fig. 5: Document after inserted a secret message "Hai"

The history of Indonesia bas, been, shaped by its geographic position, its natural resources, a series of human migratons and contacts, wars and conquests, as well as by trade, economics and politics. Indonesia is an archipelagic country of 17,000 to 18,000 islands (8,844 named and 622 permanently inhabited) stretching along the equator in South East Asia. The country's strategic seal-ane position fostered inter-island and international trade, trade has since fundamentally shaped indonesian history. The area of Indonesia is populated by popules of various migrations, creating a civersity of outures, ethnolities, and languages. The archipelago's indones and climate significant fundamentally and trade, and the formation of states. The boundaries of the state of indonesia and his tools, populary known as the "Java Man", suggest the indonesian archipelago was inhabited by least 1.5 million years ago. Austronesian for million may and the state of indone excits and this tools, populary known as the "Java Man", suggest the indonesian archipelago was inhabited by least 1.5 million years ago. Austronesian for the majority of the modern population, archive contravy, been form the majority of the modern and Buddhat influences with 1. The aniotultural Buddhat Salendra and Hindu Matanan dynasties subsequently thrived and declined in initinal gauge. The last significant non-Musim kingdon, the Hindu Matanati kingdom, florithed from the late 13° century, and the outroe as stretched over much of Indonesia and microse stretched over much of Indonesia and microse structure and the structure in the structure selecting or themoson and the state of the dimension in donesia dates to the Sine overlad and minidonesian tinde structure selecting or the modern and table structure is a structure in the structure selecting or the mode part is also overlad and minidonesian tinde selecting the dominant religion in Java and Sumatar to the late trady selecting the structure selecting or the mode selecting the structure selecting or the substruc

Fig. 6: Document after inserted a secret message "Kamu"

Recover: After the process of inserting a secret message successfully done, the next process is the extraction of

the face of international

Fig. 7: Detection, The "f" Characters Shift Left

coup in 1965 le d to

Fig. 8: The "d" Characters Shift Right

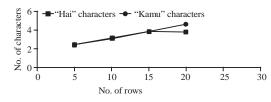


Fig. 9: Chart of message capacity

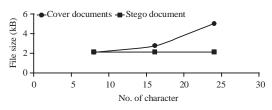


Fig. 10: Chart of file size

messages using the method Centroids, at this stage review the image profile that has been done scanning. Here is the calculation of centroid scanning done starting from the line X, (X+1), ..., (X+W). Image matrix with Secret message "Hai":

fx0	fx1	fx2	fx3	fx4	fx5	
0	1	0	1	1	0	fx0
1	0	0	1	0	0	fx1
0	0	1	0	0	0	fx2
0	1	1	1	0	1	fx3
1	1	0	0	0	0	fx4
0	1	0	0	1	0	fx5

Using the profile formula to calculate the profile on the image using centroid, the results (Fig. 7-10):

- v(1): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(1): 0+1+0+0+1+0 = 2
- v(2): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(2): 1+0+0+1+1+1 = 4
- v(3): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)

- v(3): 0+0+1+1+0+0=2
- v(4): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(4): 1+1+0+1+0+0 = 3
- v(5): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(5): 1+0+0+0+1=2
- v(6): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(6): 0+0+0+1+0+0 = 1

Centroid calculation:

$$\frac{(2+4+2+3+2+1)*6}{(2+4+2+3+2+1)+6} = 14*6/14+6 = 4, 2$$

Image matrix with Secret message "Kamu"

fx0	fx1	fx2	fx3	fx4	fx5	
1	0	0	0	0	0	fx0
0	0	1	0	0	1	fx1
0	1	0	0	0	1	fx2
1	0	1	1	0	1	fx3
1	0	0	0	0	1	fx4
0	1	1	0	0	0	fx5

Using the profile formula to calculate the profile on the image using centroid, the results is :

- v(1): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(1): 1+0+0+1+1+0 = 3
- v(2): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(2): 0+0+1+0+0+1 = 2
- v(3): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(3): 0+1+0+1+0+1 = 3
- v(4): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(4): 0+0+0+1+0+0 = 1
- v(5): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(5): 0+0+0+0+0+0=0
- v(6): f1(1, 0)+f2(1, 1)+f2(1, 2)+f3(1, 3)+f4(1, 4)+f5(1, 5)
- v(6): 0+1+1+1+1=4

Centroid calculation:

 $\frac{(4+0+1+3+2+3)*6}{(4+0+1+3+2+3)+6} = 19*6/19+6 = 4,56$

CONCLUSION

From the experiments that have been done in this paper can be concluded, the number of characters that can be inserted in the word-shift technique is directly proportional to the number of lines in a text document, the capacity of messages that can be inserted messages are also more. The centroid extraction method can be used to extract secret messages in a word-shift technique.

REFERENCES

- Saraswathi, V. and M.S. Kingslin, 2014. Different approaches to text steganography: A comparison. Intl. J. Emerging Res. Manage. Technol., 3: 124-217.
- 02. 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.
- Rohayah, S., G.W. Sasmito and O. Somantri, 2015. [Steganography application for message insertion (In Indonesian)]. J. Inf. Ahmad Dahlan, Vol. 9 No. 1. 10.26555/jifo.v9i1.a2038

- 04. 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.
- 05. Rinaldi, M., 2004. [Steganography and watermarking]. Institut Teknologi Bandung, Bandung, Indonesia. (In Indonesian)
- 06. 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, pp: 130-133.
- Andiniarti, I., 2009. [Steganographic implementation in text media using line-shift coding method and centroid method]. M.Sc. Thesis, IPB University, Bogor, Indonesia. (In Indonesian)
- 07. 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, pp: 689-695.