

Graph Steganography Based on Multimedia Cover To Improve Security and Capacity

Ilham Firman Ashari
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
ilhamfirman39@gmail.com

Rinaldi Munir
School Of Electrical Engineering and Informatics
Bandung Institute Of Technology
Bandung, Indonesia
rinaldi.munir@itb.ac.id

Abstract— Information is important asset, so it takes effort to maintain the confidentiality, integrity, and availability of information. Cryptography and steganography methods can be combined to improve information security. Steganography is divided into two types, Noisy Steganography and Noiseless Steganography. Noisy steganography has some disadvantages that can cause noise and the process of concealment that requires a container, while Noiseless Steganography (Nostega) will not produce noise and not require a container. One of the nostega paradigms is Graph Steganography (Graphstega). Graphstega is a technique that inserts messages as plotted data on a graph. This paper proposes a method that produces a graphstega encoding method that reduce the plot ratio gap, enhances data security by encrypting data, improve reality data with combination of excel cover, increasing message capacity without overlapping plots using (data splitting techniques, setting graphstega image resolution, setting font size on the graph plot), and implementing graphstega on text cover, digital image, and image hardcopy. The result of proposed graphstega method shows that it can improve data security, the value of the plot looks more realistic, reduce the plot ratio gap, increase message capacity, and can be implemented in text cover, image cover, and image hardcopy so the message distribution increases.

Keywords—information security, steganography, noiseless graphsteganography, graph steganography, cryptography

I. INTRODUCTION

Information is a very important asset, so it takes effort to maintain data confidentiality, integrity, and availability of information [1]. Information protection can be done in several ways that are cryptography and steganography methods [2]. Cryptography is a study of mathematical techniques related to information security aspects such as confidentiality, data integrity, and data authentication [3]. Steganography is a technique of securing files, messages, images, audio, or video by inserting messages on other media cover, such as pictures, audio, or video [4]. Steganography is divided into two types that are noisy Steganography and noiseless steganography [5].

Noisy steganography is a steganography technique that aims to hide the existence of messages by altering the bit of message and cover, so it will raise suspicion and not secure [5]. Nostega offers a technique that is able to hide messages without cover and hide delivery. Nostega can legitimize the interaction among parties communicating through an

approved domain or content so that the other party will not expect the interaction between the sender and the receiver [5]. One of the Noiseless Steganography paradigms is Graph Steganography (Graphstega). Graph Steganography is a steganography method that inserts messages as plotted data in a graph. Graphstega camouflages both message and its transmittal. Graphstega is resilient to contemporary attacks, such as traffic analysis attack and comparison attacks [6].

Graphstega represents a message by using a particular encoding scheme and plotting the values on the graph, this makes the plot value sometimes unrealistic. For example, showing a graph that has a value between 100 and 2000 makes not possible to do. One of the problems that can be raised about the use of cover chart is the size limit of messages that can be hidden. If the message is long, it will make the value of each plot on the graph will overlap each other and make the graph look less realistic. The technique of plotting messages into the graph is a matter of concern, since the value of each plot represents the word or character of the message, therefore it is necessary to note the security aspects and data reality. Data reality is related to the value gap ratio between each plot on the graph. Ratio gap is the difference value between each plot resulted from the encoding method.

In this paper, we proposed a new approach encoding method that enhances data security by encrypting data plot, increasing message capacity without cause overlapping plots in graphstega by using (data splitting techniques, setting graphstega image resolution, setting font size on plots), and implements graphstega on some cover types such as text cover, image cover, and hardcopy image, so it is expected to improve the message distribution. Ratio gap makes graphstega image look less realistic, to make the plot of the graph more real than combination technique using excel cover.

The ratio gap in graphstega can be reduced by using encoding method that are syllabification and bit divider factor. The syllabification method is a syllable splitting method based on predefined rules, each word will be splitted by syllable and plotted on graphstega. The bit divider factor is the ally factor of the values that can divide the binary value of the message, each binary word will be converted to a decimal by a dividing factor and plotted on the graph. The purpose of using two encoding methods in this research are to compare the number of plots generated from each method, so it will know optimal encoding method in reducing file size. In this research, we used two encryption algorithms that are AES and RC4. Its purpose is to compare the effect of encryption algorithms used on each encoding method to produce more efficient file cover sizes and reduce the number of plots in graph cover.

II. RELATED WORKS

In Nostega, the steganographic goal is achieved by determining a suitable domain that is capable of generating an innocent appearing steganographic cover in which a message is intrinsically embedded in the form of innocent data compatible with the chosen domain [5]. In addition, Nostega establishes a covert channel by employing a selected domain to establish communication between the sender and receiver. Graphsteganography (Graphstega) is one of the paradigms of Noiseless Steganography (Nostega). The advantage of Graphstega is that the message plotting on the graph does not produce noise so it is anti-distortion. Graph type can consist of several forms such as bar, pie, scatter, etc. Graphstega can convert data content to text cover, image cover, and audio cover. Graphstega is a public noiseless steganography model that does not require the confidentiality of the techniques or methods used.

RC4 is one of a kind stream cipher cryptography. The RC4 encryption is developed by Ronald Rivest in 1984. RC4 operates with byte orientation [2]. In 2001, NIST (National Institute of Standards and Technology) published AES (Advanced Encryption Standard) as a document and information processing standard. AES operates with block byte operation [2]. Encryption is used to secure data on the graph, so that the same character in the message will not have the same plot value.

The research related to graphstega has been proposed by Akhter [7]. The method that proposed by Akhter has limited space or capacity for embedding message because element in sudoku puzzle is only 81 cells (9x9). The study [7] proposes a Huffman coding method to embed secret messages by prefix code from every word in message and plotting the prefix code value of each word to graphstega. To improve data security then used scale value [7]. Graphstega results will be converted to text cover or excel cover. The method that proposed by [6] and [7] is vulnerable to overlap if many characters are inserted, media cover has limitations because it only uses excel cover, and security issues because the same word and character always has the same plot value. Comparison of the previous method with the proposed method can be seen in the table I.

TABLE I. COMPARISON GRAPH STEGANOGRAPHY METHODS

| Parameters | Graph Steganography Methods | |
|--------------------------------------|--|--|
| | Graphstega Method By [7] | Proposes Method |
| Overlapping Plot Values on The Graph | Yes (if the message is inserted a lot or small resolution of image, it will occur overlapping on the graph. it can be inferred the number of characters is limited) | No (Using Split Method, the number of plots displayed on the graph can be adjusted) |
| Security on The Graph | Scale Value (the same words and characters will always have the same plot value) | Encryption (words and the same characters will have different plot values) |
| Data Reality | the plot values that showed in the graph are the result of the encoding method, there can be a large difference ratio depending on the prefix code of each word generated) | Using Combination Excel Cover and using encoding method that are syllabification and bit divider factor, so the gap ratio on the plot can be solved. |

| Parameters | Graph Steganography Methods | |
|-------------|-----------------------------|---|
| | Graphstega Method By [7] | Proposes Method |
| Graph Media | Text (Excel Cover) | Text (Excel Cover), Digital Image, and Hardcopy Image |
| Graph Type | Bar Chart | Line Chart, Scatter Plot, and Bar Chart |

III. PROPOSED METHODS

In this section, it will be described the proposed graphstega method which consists of encoding and encryption, then decoding and decryption. In this paper, we proposed two encoding methods, that are the syllabification and the bit dividing factor. Encryption and encoding methods are techniques to hide messages by encrypting every word or character and plotting result on the cover chart. Decoding and decryption method is a technique of extracting values from each plot on graphstega cover and decrypting values on each plot value into a message.

A. Proposed Encryption and Encoding Approach

Users need to input parameters ie message, encryption type, key encryption, and graph type (bar, scatter, line chart) to perform encryption and encoding process. Parameter combination excel cover is optional. All parameters entered before the encryption and encoding process will be verified. The syllabification and bit divider encoding algorithm can be seen in figure 1.

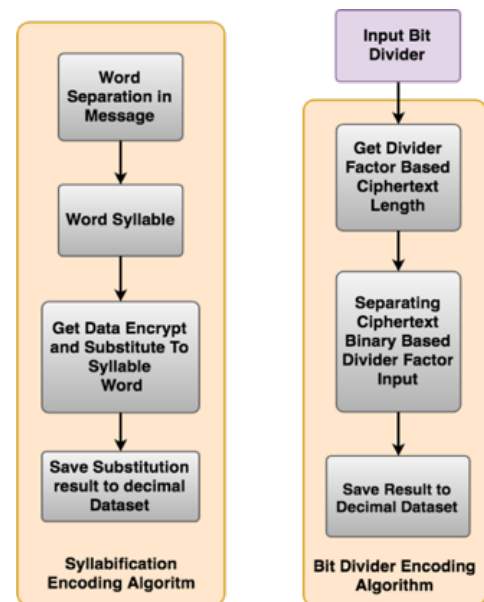


Fig. 1. Syllabification and Bit Divider Encoding Algorithm

Encoding by a syllabification algorithm, words will be separated from messages based on spaces. Each word will be separated by syllable word algorithm which was adopted from Frank Liang [8]. And then do the substitution of ciphertext byte into syllable byte. The result of the substitution byte will be converted into a decimal value and stored in the dataset variable. Encoding by using bit divider algorithm, ciphertext result will be converted into binary string. Then we will look for bit divider factor that can divide binary ciphertext length. Then do a binary string separation based on the input of dividing factor and convert it to decimal value.

The conversion results will be stored in the dataset variable. The dataset is the variable used to fill each plot on the graphste cover. Any input such as (dataset, chart title, legend chart, chart value plot, horizontal title chart, vertical title chart, and dataset) will be in the process of realizing the graph. In addition, graphstega can be split based on plot input, which is adjusted based on the number of datasets, so it generates some images and excel. An overview of the proposed encryption and encoding method can be seen in Figure 3.

B. Proposed Decoding and Decryption Approach

An overview of the proposed decoding and decryption method can be seen in Figure 4. In this research, we proposed two graphstega covers, namely excel cover and image cover. Here is the decoding and decryption process.

- If the users choose the image cover, then cover should be cropped first to take up areas that represent only plot values. After that, it will be extraction process using OCR (Optical Character Recognition) engine, before decoding process done, it will do pre-processing to filter extraction result from OCR so just take every number not character. Users need the column index table that represents OCR extraction message result for message decoding. If the user chooses excel cover, then the user inputs a column index on excel representing the message.
- Each cover type is divided into two types, namely split excel and split image. Users can select extract by using graphstega split with cover type (excel cover and image cover). The user must input every correct split cover to perform the decoding process.

Before the decoding and decryption processes are done, user need to input parameter that are decryption type algorithm, decryption key, and column index based on cover graphstega to perform decoding and decryption process. The syllabification and bit divider decoding algorithm can be seen in figure 2.

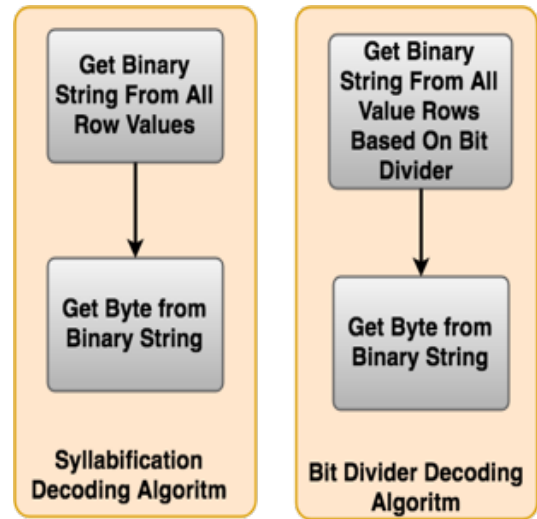


Fig. 2. Syllabification and Bit Divider Decoding Algorithm

Decoding using syllabification, every binary string will be taken at each plot cover value. Decoding using bit divider, binary string of each cover plot will be taken based on the bit divider factor. The binary string that obtained will be converted to byte. After the byte is obtained it will do the derpytion process. If the decryption process succeeds, the original message will be obtained.

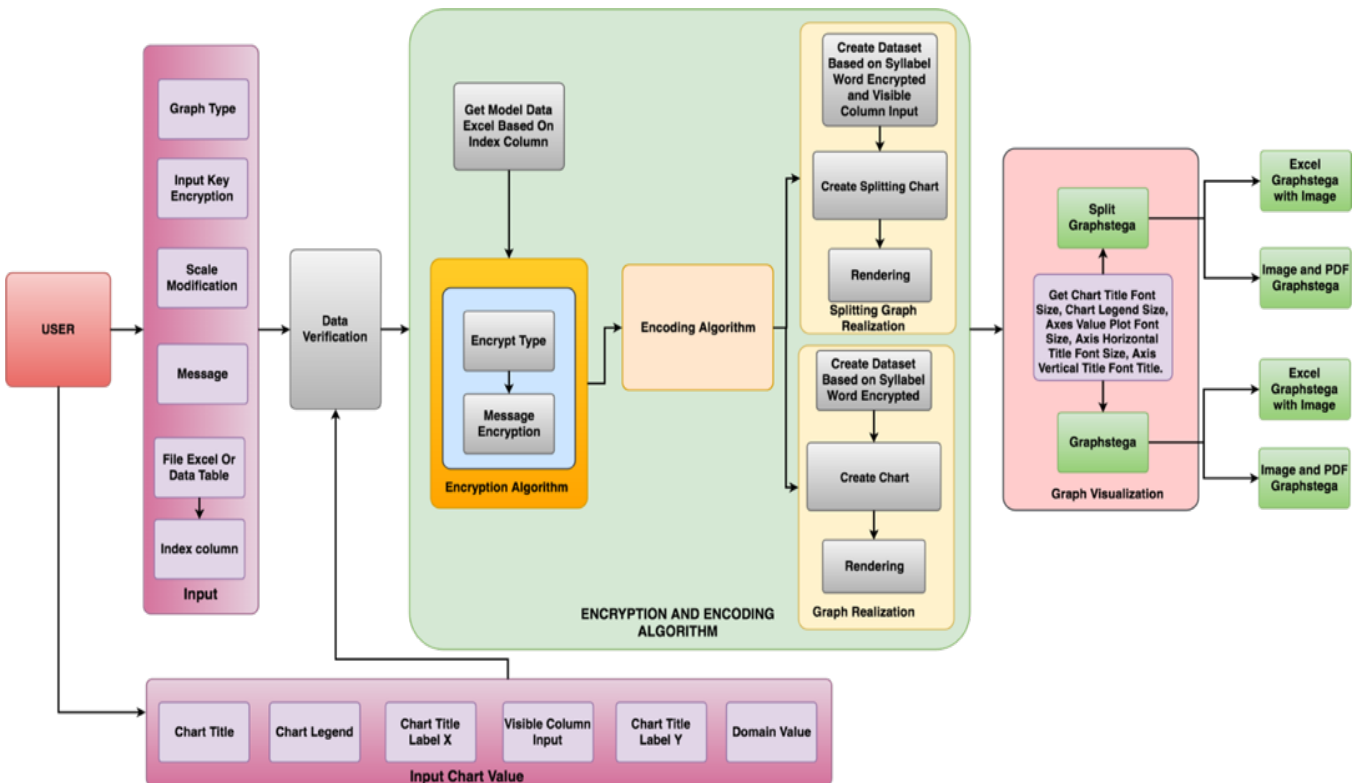


Fig. 3. Syllabification and Bit Divider Encoding Algorithm

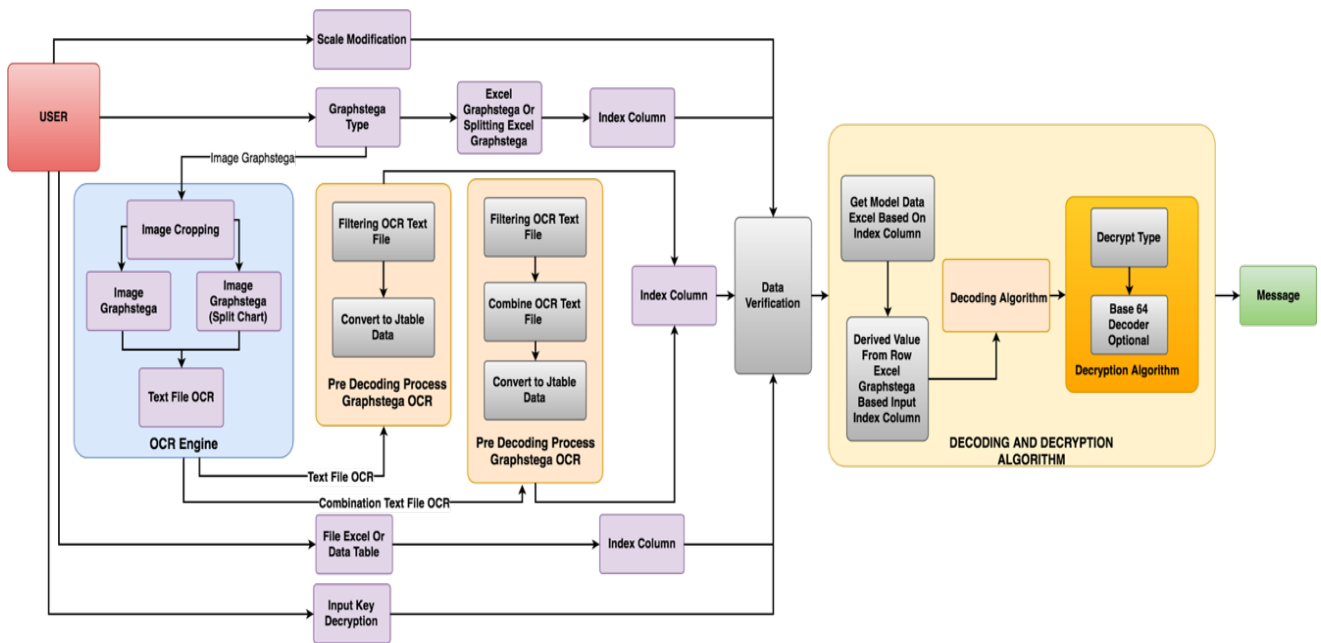


Fig. 4. Syllabification and Bit Divider Decoding Algorithm

The structure of Indonesian pattern for syllabification algorithm can be seen in table II.

TABLE II. INDONESIAN PATTERN (HYPHENATION.ORG)

| Pattern | Detail and Example in Indonesian word |
|---|--|
| al e1 il ol ul | Syllabification is done after vowel words |
| 2b1d 2b1j 2b1k 2b1n 2b1s 2b1t 2c1k 2c1n 2d1k 2d1n 2d1p 2f1d 2f1k 2f1n 2f1t 2g1g 2g1k 2g1n 2h1k 2h1l 2h1m 2h1n 2h1w 2j1k 2j1n 2k1b 2k1k 2k1m 2k1n 2k1r 2k1s 2k1t 2l1b 2l1f 2l1g 2l1h 2l1k 2l1m 2l1n 2l1s 2l1t 2l1q 2m1b 2m1k 2m1l 2m1m 2m1n 2m1p 2m1r 2m1s 2n1c 2n1d 2n1f 2n1j 2n1k 2n1n 2n1p 2n1s 2n1t 2n1v 2p1k 2p1n 2p1p 2p1r 2p1t 2r1b 2r1c 2r1f 2r1g 2r1h 2r1j 2r1k 2r1l 2r1m 2r1n 2r1p 2r1r 2r1s 2r1t 2r1w 2r1y 2s1b 2s1k 2s1l 2s1m 2s1n 2s1p 2s1r 2s1s 2s1t 2s1w 2t1k 2t1l 2t1n 2t1t 2w1t | Syllabification is done between two consonants consecutively |
| 2ng1g 2ng1h 2ng1k 2ng1n 2ng1s | Three consonants, Syllabification is done in the third consonant. |
| .be2r3 .te2r3 .me2ng3 .pe2r3 | Word prefix (ber-, ter-, meng-, per-) |
| 2b1kan. 2c1kan. 2d1kan. 2f1kan. 2g1kan. 2h1kan. 2j1kan. 2l1kan. 2m1kan. 2ng1kan. 2n1kan. 2p1kan. 2r1kan. 2s1kan. 2t1kan. 2v1kan. 2z1kan. | Syllabification is done before suffix (-kan). E.g (tuntaskan) |
| 2n1lah. 1lah. | Syllabification is done before -lah |
| 2ng. 2ny. | There is no syllabification -ng and -ny at the end of the word [9] |
| 2b1an. 2c1an. 2d1an. 2f1an. 2g1an. 2h1an. 2j1an. 2k1an. 2l1an. 2m1an. 2ng1an. 2n1an. 2p1an. 2r1an. 2s1an. 2t1an. 2v1an. 2z1an. 3an. | Syllabification of suffix (-an) is done after letter (b, c, d, f, g, h, j, k, l, m, n, p, r, s, t, v, dan z) |
| .ta3ng4an. .le3ng4an. .ja3ng4an. .ma3ng4an. .pa3ng4an. .ri3ng4an. .de3ng4an. | E.g : ta3ng4an = tangan .le3ng4an. = lengan .ja3ng4an. = jangan .ma3ng4an. = mangan .pa3ng4an. = pangan .ri3ng4an. = ringan .de3ng4an. = dengan |

The characters on the pattern can be divided into three kinds, namely:

- Non-numeric is all letters that will be represented as characters to be evaluated for beheading word.
- Point (.) represents the boundary of a pattern in a word, for example the initial or final limit of a word, so that it will not be beheaded before (.) and after (.).
- Numeric is the number that represents a scoring system to determine the point of beheading. This numerical character is divided into two, that are odd and even.

Illustration of calculation to get the syllable of “pendapat” word in Indonesian can be seen in table III. The process of separating words into syllables is done if character mark point is odd.

TABLE III. ILLUSTRATION TO GET THE SYLLABLE OF THE WORD

| Character Trie Structure | Charachter | | | | | | |
|--------------------------|------------|----|----|----|----|----|-----|
| | P | | | | | | |
| | e1 | | | | | | |
| | e2 | n1 | d0 | | | | |
| | | | d0 | a1 | | | |
| | | | | | p0 | a | t |
| Word pattern structure | p | e2 | n1 | d0 | a1 | p0 | a t |
| Character Mark Point | 0 | 2 | 1 | 0 | 1 | 0 | 0 |

At the beginning of the word, there is no syllabification process. In the second letter there is a pattern trie, namely e1, which means there is a beheading afterwards. Then compared again with another patterns from letter e, namely 2n1d. The pattern trie taken is the largest value after the letter e, which is the 2n1d pattern. The next step, check the last letter of the pattern trie that is d, the letter d does not have a trie pattern so it is filled with 0. Then proceed to the letter a and there is a pattern, namely a1, after the letter a, the letter p. The letter

p does not have a trie pattern so it is filled with 0. The letters a and t are ignored because the limit of beheading in Indonesian is two letters before the end of the word.

In the Indonesian pattern, the initial character separation is done on one character at the beginning of the word and there is no need to check the trie pattern. The end of character separation is done on two characters before the end of a word. Subpattern level division is based on the values contained in the pattern. The highest pattern value will be chosen as the character mark point. The result of word syllable in table III is "pen - da - pat". Flowchart word syllable can be seen in figure 5.

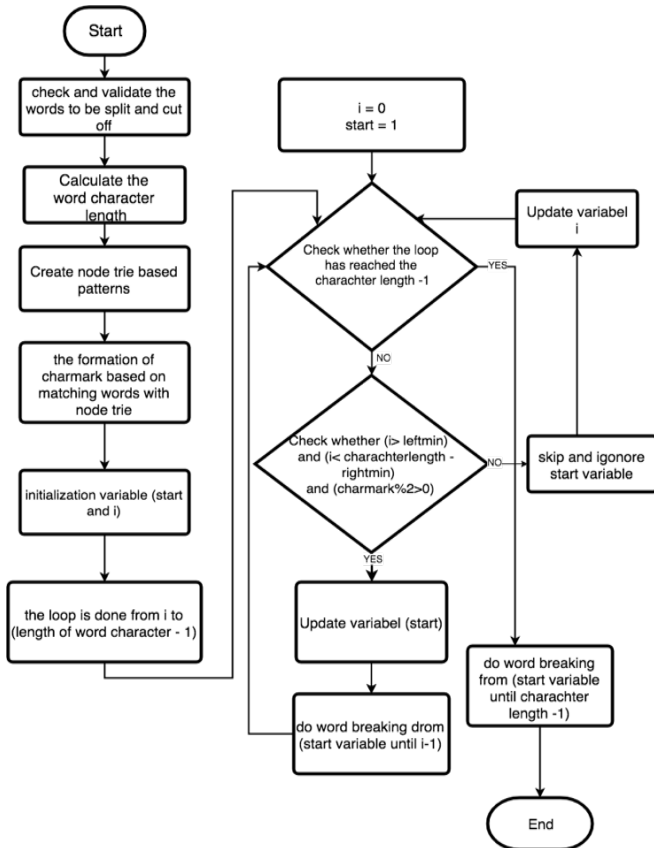


Fig. 5. Flowchart Word Syllable

The following table IV is the decimal value of the plot generated from the encoding process using silabification algorithm and encryption using RC4 algorithm.

TABLE IV. DECIMAL VALUE USING SYLLABIFICATION AND RC4

| Syllable | Encryption Byte | Binary Encryption | Decimal Plot Value on Graph |
|----------|-----------------|-------------------|-----------------------------|
| p | -116 | 10001100 | 9218018 |
| e | -89 | 10100111 | |
| n | -30 | 11100010 | |
| d | -3 | 11111101 | 64855 |
| a | 87 | 01010111 | 16650703 |
| p | -2 | 11111110 | |
| a | 17 | 00010001 | |
| t | -49 | 11001111 | |

The decimal value of the plot in the graph is 9218018, 64855, and 16650703.

Suppose there is a word that is "pendapat", the calculation table to get divider factor from binary text "pendapat" can be seen in table V below.

TABLE V. THE CALCULATION TO GET BIT DIVIDER FACTOR OF "PENDAPAT" WORD

| Length Binary Cipher Text (a) | Limit = sqrt (a) | looping until i = limit increment if (odd = i+1) else (i+2) | if (a % i = 0) factor = i | if (a % i = 0 and i != (a / i) and a/i < 64) factor = a / i | Divider Factor |
|-------------------------------|------------------|---|---------------------------|---|----------------|
| 128 | 11 | 1 | 128 % 1 = 0 | 128 / 1 = 128 | 1 |
| | | 2 | 128 % 2 = 0 | 128 / 2 = 64 | 2 |
| | | 3 | 128 % 3 = 2 | null | - |
| | | 4 | 128 % 4 = 0 | 128 / 4 = 32 | 4,32 |
| | | 5 | 128 % 5 = 3 | null | - |
| | | 6 | 128 % 6 = 2 | null | - |
| | | 7 | 128 % 7 = 2 | null | - |
| | | 8 | 128 % 8 = 0 | 128 / 8 = 16 | 8,16 |
| | | 9 | 128 % 9 = 2 | null | - |
| | | 10 | 128 % 10 = 8 | - | - |
| | | 11 | 128 % 11 = 7 | null | - |

Based on table V the divider factor values is 1, 2, 4, 8, 16, and 32. Divider factor values are values that qualifies:

$$a \% i = 0 \text{ or } (a \% i = 0 \text{ and } a \neq \left(\frac{a}{i}\right) \text{ and } \left(\frac{a}{i} < 64\right)) \quad (1)$$

Each condition in formula (1) must be fulfilled, if one of the conditions is not fulfilled then it is not included in bit divider factor. The binary length of ciphertext (a) is done looping process (i) by calculating operation with modulo to limit = sqrt (a). Factoring is limited to $\frac{a}{i} < 64$ because the use of data types with values over 64 bits causes the decimal value to be too large. In the table VI below can be found the decimal value of the plot generated using bit divider factor 16 and AES.

TABLE VI. BINARY AND DECIMAL PLOT VALUE USING BIT DIVIDER FACTOR (16) AND AES

| Ciphertext Binary | Bit Divider | Decimal Grouping Based on Bit Divider Factor (16) |
|-------------------|-------------|---|
| 1001110000110011 | 16 | 39987 |
| 0110001010111000 | | 25272 |
| 1100011110110000 | | 51120 |
| 1001110010011000 | | 40088 |
| 0111000000111100 | | 28732 |
| 0000111011110100 | | 3828 |
| 1010110101010111 | | 44375 |
| 0110111000101011 | | 28203 |

The ciphertext binary length is 128 so if divided by 16 bits it will produce 8 plot values (39987, 25272, 51120, 40088, 28732, 3828, 44375, 28203).

IV. EXPERIMENT RESULTS AND EVALUATION

In this section, we test the performance of the proposed graphstega method in some tests.

A. Testing Encoding and Encryption Algorithms on Data Security and Computation Time.

Let the message be “**pesan tersurat dan pesan tersirat**” in Indonesian language and the syllable word is **pe-san ter-su-rat dan pe-san ter-si-rat**. The Graphstega result that proposed by [7] can be seen in figure 6. The graph looks less realistic because the difference in ratios between each plot value is so great and the same word or character have the same decimal plot values. The minimum plot value is 1 digit number and the maximum value is 8 digit number. The difference of these digits can be larger, depending on the decimal value of the prefix code generated. In figure 6, space characters always have the same decimal value.

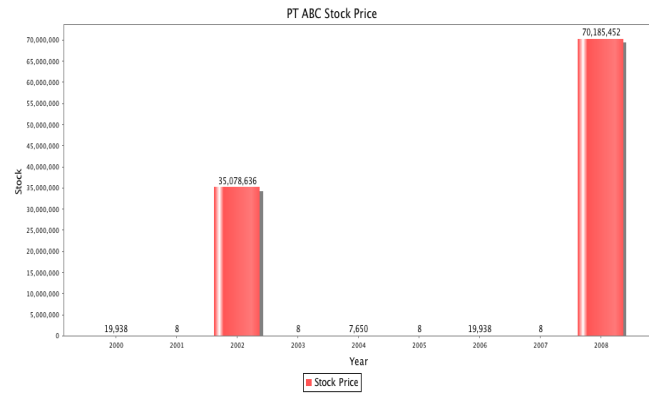


Fig. 6. Graphstega Using Huffman Encoding Proposed by [7]

On the proposed method, the messages are encrypted using RC4 and AES. The first experiment was done using scatter plot. The decimal result from syllabification using RC4 encryption can be seen in tables VII.

TABLE VII. VALUE OF DECIMAL SYLLABIFICATION ENCODING AND ENCRYPTION RC4 USING SCATTER PLOT

| Decimal Value Syllabification Encoding and RC4 Encryption [word, decimal value] (Number of Plot = 15) | Time | File Size |
|---|--------|-----------|
| [“pe”, 54315], [“san”, 16519438], [“”, 198], [“ter”, 6748029], [“su”, 50618], [“rat”, 3893647], [“”, 158], [“dan”, 8349601], [“”, 230], [“pe”, 58915], [“san”, 6589788], [“”, 116], [“ter”, 15477583], [“si”, 60897], [“rat”, 14509607] | 9.5 ms | 11.15 kb |

Graphstega using RC4 encryption and syllabification can be seen in figure 7.

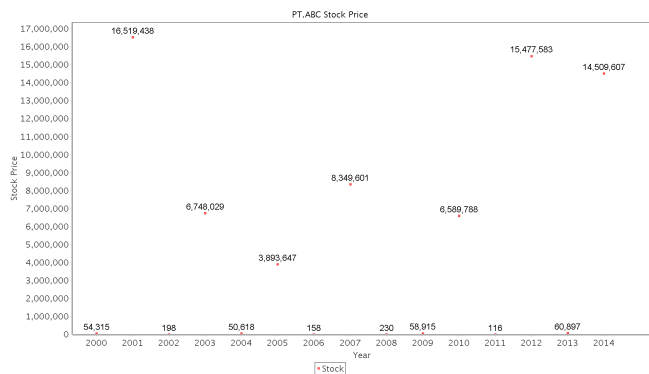


Fig. 7. Graphstega using RC4 and Syllabification

The decimal plot result from bit divider using RC4 encryption can be seen in tables VIII.

TABLE VIII. VALUE OF DECIMAL BIT DIVIDER ENCODING AND ENCRYPTION RC4 USING SCATTER PLOT

| Decimal Value Bit Divider Encoding and RC4 Encryption (24 bit) (maximum divider bit = 42) (the number of plot = 11) | Time | File Size |
|---|---------|-----------|
| [13904892, 1117894, 6748029, 12958267, 6918046, 8349601, 15132195, 6589788, 7662635, 5238241, 14509607] | 7.03 ms | 8.5 kb |

Graphstega using RC4 encryption and bit divider factor can be seen in figure 8.

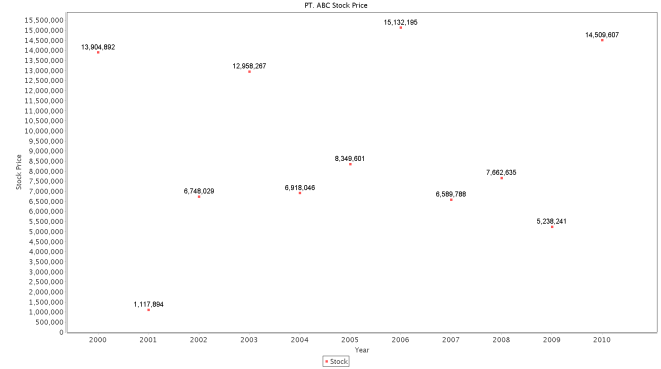


Fig. 8. Graphstega Using RC4 and Bit Divider Factor

The second experiment was done using another chart type, that is bar chart. The decimal result from syllabification using AES encryption can be seen in tables IX.

TABLE IX. VALUE OF DECIMAL SYLLABIFICATION ENCODING AND ENCRYPTION AES USING BAR CHART

| Decimal Value Syllabification and AES Encryption (remain 15 bytes, divider 3 byte) (number of Plot = 20) | Time | File Size |
|--|----------|-----------|
| [“pe”, 54896], [“san”, 5816312], [“”, 62], [“ter”, 5061433], [“su”, 18776], [“rat”, 10521619], [“”, 51], [“dan”, 14710421], [“”, 142], [“pe”, 55718], [“san”, 5082667], [“”, 233], [“ter”, 4516517], [“si”, 4778], [“rat”, 15328572], [3 byte, 6326858], [3 byte, 13301134], [3 byte, 3426951], [3 byte, 12487486], [3 byte, 10184987] | 12.44 ms | 20.12 kb |

Graphstega using AES and syllabification can be seen in figure 9.

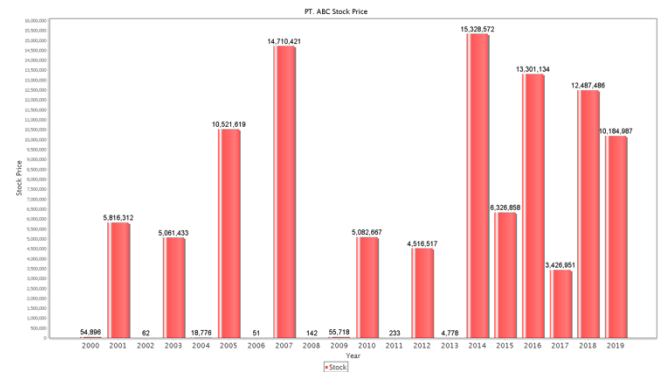


Fig. 9. Graphstega Using AES and Syllabification.

The decimal result from syllabification using AES encryption can be seen in tables X.

TABLE X. VALUE OF DECIMAL BIT DIVIDER ENCODING AND ENCRYPTION AES USING BAR CHART

| Decimal Value Bit Divider and AES Encryption (24 bit) (Number of Plot = 16) | Time | File Size |
|---|----------|-----------|
| [14053464, 12580926, 5061433, 4806816, 9179955, 14710421, 9361830, 5082667, 15287530, 10818218, 15328572, 6326858, 13301134, 3426951, 12487486, 10184987] | 10.03 ms | 16.13 kb |

Graphstega using AES and bit divider factor can be seen in figure 10.

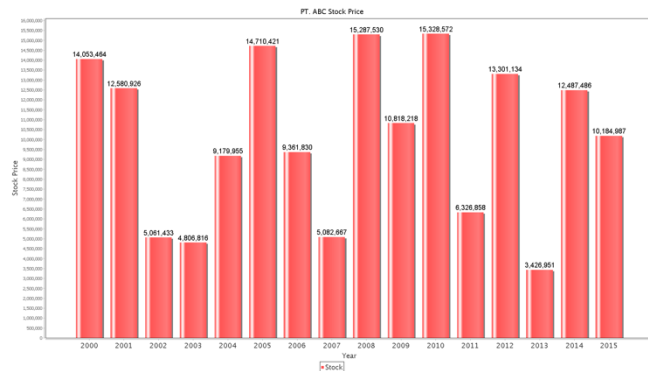


Fig. 10. Graphstega Using AES and Bit Divider Factor

The same syllable word, character or bit piece will not produce the same decimal value, so encoding methods proposed are safe from statistical analysis attacks. AES encryption can increase the number of plots, because AES encryption is block-byte oriented, while RC4 encryption is byte-oriented so it does not increase the number of plots. The graphstega file size is determined by the type of graph and the number of plots on the graph. The syllabification encoding algorithm produces a larger number of plots than the bit divider factor algorithm, which is determined by the number of syllables in the message. Whereas the bit dividing factor of the plot number is determined by its value of the bit divider, this makes bit divider factor method can produces fewer plots than syllabification method. The ratio range generated from the syllabification algorithm using the Indonesian pattern is 1 byte - 5 bytes. While using bit divider algorithm is not generated ratio gap, because each decimal value is the result of a bit divider. The best computation time uses bit divider factor and RC4 encryption.

B. Testing Increase Message Capacity Without Cause Overlapping

The technique is done by set the size of image resolution, set font size, and split dataset techniques. From the testing, the result without splitting and resolution of image (600x400) that proposed by [7] can be seen in figure 11. The graph on figure 11 is overlapping, to make the graphic image looks more realistic although the character of the message is inserted a lot then used splitting method and setting image resolution. From the testing, the proposed method can increase the message capacity without causing overlapping, so the proposed graphstega method is suitable for use on the image cover, because if the value on graphstega overlapping will not be readable by OCR engine. Let the message be **“nama saya ilham firman ashari”**.

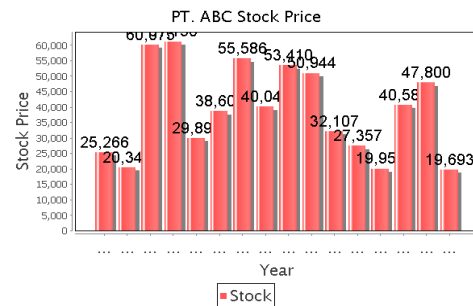


Fig. 11. Graphstega Without Splitting

On Figure 11, if the image resolution is small then the graph can plot only a few characters. Decimal value from encoding using bit divider and AES encryption can be seen in table XI. The results with splitting method can be seen in figure 12 and 13.

TABLE XI. VALUE OF DECIMAL ENCODING AND ENCRYPTION AES USING BAR CHART WITH PLOT DIVIDER

| Split Data | Dataset (Number Of Plot = 16), Encoding Method = Bit Divider (24 bit), Encryption = AES, Font Plot = 35, Image Resolutin (1920 x 1080) | Plot Divider | Image Resolution | Font Size |
|------------|--|--------------|------------------|-----------|
| Graph 1 | [25266, 20346, 60075, 61150, 29894, 38600, 55586, 40044, 53410, 50944] | 10 | 1280 x 720 | 25 |
| Graph 2 | [32107, 27357, 19951, 40585, 47800, 19693] | | 1280 x 720 | 25 |

The graphstega from splitting method can be divided into two images. The first image is the value of the first 10 data plots and the second image is the value of the last 5 data plots.



Fig. 12. The First Graphstega Image (Graph 1)

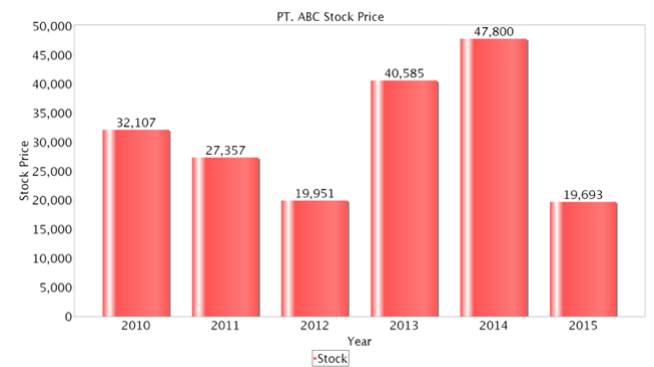


Fig. 13. The Last Graphstega Image (Graph 2)

Split data on the graph makes the values and shape of the graph more legible because it does not cause overlapping of values between graphs. Displaced files can be reassembled and can be extracted again, with requirements file that combined and the encryption key entered are appropriate.

C. Testing Data Reality Improvement

Let the message be = “**nama saya ilham**” using divider bit algorithm (bit divider = 16) and AES. Decimal plot graphstega before and after combination can be seen in table XII. The one way to make the graph more realistic is used combination excel cover. The excel file used as a combination is a file that has been calculated first before being combined with the graphstega plot value.

TABLE XII. THE VALUE PLOT BEFORE AND AFTER COMBINATION USING EXCEL FILE

| Decimal Plot (D1), Font Size = 28 | Excel Cover (E1) | Combination (D1 ⊕ E1) |
|--|--|--|
| 51759, 57873, 16533, 29675, 28310, 42533, 24730, 54075 | 51655, 58973, 17445, 30463, 27630, 41977, 26330, 54687 | 1000, 1100, 1200, 1300, 1400, 1500, 1600 |

Suppose that content of graph relates to an increase indomie price from 2000 to 2007. The graph before being done combination can be seen in figure 14.

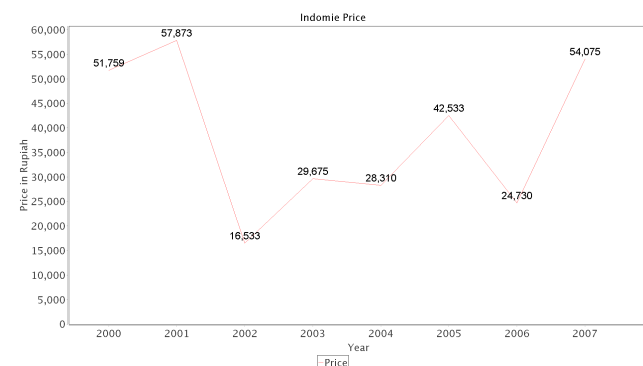


Fig. 14. Graphstega Original Without Combination

Graphstega in figure 14 is the result from encoding method, so the graph looks less realistic in the real world. Graphstega must be combined to make it look more real. The graph after being done combination can be seen in figure 15.

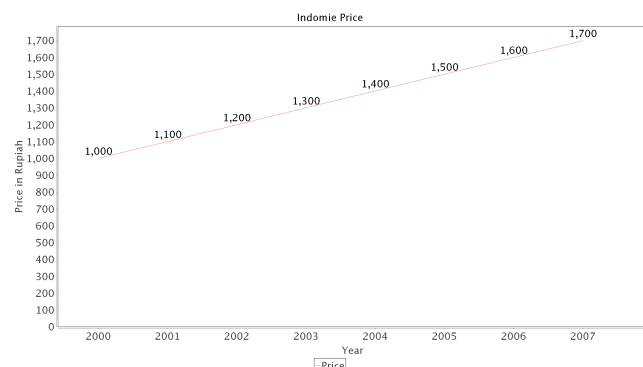


Fig. 15. Graphstega After Combination using Excel Cover

If compared between the figure 14 and 15 then seen in the figure 15, increasing indomie price is more realistic.

V. CONCLUSION AND FUTURE WORKS

The proposed a new graphstega approach provides a better solution than the existing approaches. The experimental results show the proposed Graphstega method approach can improve security, it is seen that each word and the same character will not have the same decimal value.

Data reality can be improved by using encoding syllabification and bit divider factor method so they can reduce the ratio gap in the large value digit between each plot. The plot ratio generated using a silabification algorithm ranges from decimal values from 1 byte - 5 bytes depending on the syllable produced. The more varied words, the more varied the plot value produced.

The ratio of plot values resulting from the bit dividing factor method does not have a significant difference in digit values, because each decimal value in the plot is the result of the division of bits based on the dividing factor. The bit divider factor encoding method is more optimal in reducing the ratio gap and the number of plots on the graph. RC4 encryption method is more efficient in reducing the number of plots is RC4, but in terms, AES more secure. The best computation time is using bit divider factor encoding and RC4 encryption. Furthermore, Plot values on the graph can be made more real because they are combined using an excel file.

The proposed method can improve data capacity without overlapping in the graph by using data splitting, setting image resolution and setting font size. By using font settings and split data, allowing the proposed method is suitable to be implemented using hardcopy images. Message distribution can be improved using some graph cover (excel, image, and image hardcopy) and graph type (bar, scatter, and line chart).

Future works, do experiment to embed image and audio, and add graphstega cover using audio cover.

REFERENCES

- [1] Whitman, M. E., dan Mattord, H. J. (2011). Introduction to Information Security. Principles of Information Security, 1–38. <https://doi.org/112>
- [2] Stallings, W. (2011). Cryptography and Network Security. Network (Vol. 139). <https://doi.org/10.1007/11935070>
- [3] Menezes, A., van Oorschot, P., dan Vanstone, S. (1996). Handbook of Applied Cryptography, 19964964. <https://doi.org/10.1201/9781439821916>
- [4] Katzenbeisser, S., dan Petitcolas, F. (2000). Information hiding techniques for steganography and digital watermarking. (A. House, Ed.). London.
- [5] Desoky, A. (2012). Noiseless Steganography. <https://doi.org/10.1201/b11575>
- [6] Desoky, A., dan Younis, M. (2008). Graphstega: Graph Steganography Methodology. *Journal of Digital Forensic Practice*, 2(1), 27–36. <https://doi.org/10.1080/15567280701797087>
- [7] Akhter, F. (2016). A secured word by word Graph Steganography using Huffman encoding. 2016 International Conference on Computer Communication and Informatics, ICCCI 2016, 9–12. <https://doi.org/10.1109/ICCCI.2016.7480015>
- [8] Liang, F. M. (1983). *Word hyphenation by computer*. Department of Computer Science, Stanford University.
- [9] Prasetyo, E., dan Timoteus, T. (2016). Teknik Pemenggalan Kata Bahasa Indonesia Dalam Dokumen Latex, (January).