Optimization Substitution Cipher And Hidden Plaintext In Image Data Using LSB Method

Dony Ariyus¹, Ardiansyah²
¹,² Magister of Informatics Engineering, AMIKOM Yogyakarta University, Indonesia.
E-mail: dony.a@amikom.ac.id, ardiansyah.0948@students.amikom.ac.id

Abstract. Cryptography is a way to protect certain messages that have confidential properties or have properties that must be protected from unrelated parties. Substitution techniques have several algorithms such as hill cipher, vigenere cipher, one-time pad and others. This research is a development of previous research that has been carried out many combine substitution techniques. This research combines 4 substitution techniques that have been modified and stored messages that have been hidden into an image with the Least Significant Bit (LSB) algorithm. This research provides additional before the process of standard substitution techniques so that it will provide a more level of security even though the encryption process uses simple techniques and inserts messages into an image with PNG extension using LSB to improve security in messages. This is evidenced by the absence of color differences in the image so that it will not arouse suspicion. However, hex and the size of the image will change. Besides this research found that images can only be implemented on social media such as Google+, Telegram, and Google Drive. Because social media like Facebook, WhatsApp, Instagram, and Pinterest will automatically convert, resizing to be smaller than the actual size.

1. Introduction
Cryptography is a way to protect certain messages that have confidential properties or have properties that must be protected from unrelated parties. Cryptography comes from fragments of two words crypto and graphein, both of which are from Greek. The word crypto in Greek is secret while graphein means writing [1][2]. In addition, cryptographic methods purpose to make the messages sent change so that the security of the message will increase [3][4]. Cryptographic algorithms are very numerous, but one of them is a substitution technique that has a process of changing plaintext that is replaced or exchanged with other characters from the alphabet sequence which aims to improve message security [1][2][5]. In substitution techniques, there are several algorithms such as hill cipher, genetically engineered chip, caster cipher, one-time pad. Hill ciphers use modulo so hill cipher is called a substitution technique that combines linear and alphabetical characters to produce alphabets of new characters [6][7]. Vigenère cipher or what is often referred to as a polyalphabetic substitution cipher due to the vigenère cipher uses a 26 X 26 matrix that shifts [8]. Caesar cipher or often called the ROT3 algorithm is the most famous substitution technique by replacing the initial letters of the alphabet sequence which aims to improve message security [1][2][5]. In substitution techniques, there are several algorithms such as hill cipher, genetically engineered chip, caster cipher, one-time pad. Hill ciphers use modulo so hill cipher is called a substitution technique that combines linear and alphabetical characters to produce alphabets of new characters [6][7]. Vigenère cipher or what is often referred to as a polyalphabetic substitution cipher due to the vigenère cipher uses a 26 X 26 matrix that shifts [8]. Caesar cipher or often called the ROT3 algorithm is the most famous substitution technique by replacing the initial letters of the alphabet sequence which aims to improve message security [1][2][5]. In substitution techniques, there are several algorithms such as hill cipher, genetically engineered chip, caster cipher, one-time pad. Hill ciphers use modulo so hill cipher is called a substitution technique that combines linear and alphabetical characters to produce alphabets of new characters [6][7]. Vigenère cipher or what is often referred to as a polyalphabetic substitution cipher due to the vigenère cipher uses a 26 X 26 matrix that shifts [8]. Caesar cipher or often called the ROT3 algorithm is the most famous substitution technique by replacing the initial letters of the alphabet sequence which aims to improve message security [1][2][5]. In substitution techniques, there are several algorithms such as hill cipher, genetically engineered chip, caster cipher, one-time pad. Hill ciphers use modulo so hill cipher is called a substitution technique that combines linear and alphabetical characters to produce alphabets of new characters [6][7]. Vigenère cipher or what is often referred to as a polyalphabetic substitution cipher due to the vigenère cipher uses a 26 X 26 matrix that shifts [8]. Caesar cipher or often called the ROT3 algorithm is the most famous substitution technique by replacing the initial letters of the alphabet sequence which aims to improve message security [1][2][5]. In substitution techniques, there are several algorithms such as hill cipher, genetically engineered chip, caster cipher, one-time pad. Hill ciphers use modulo so hill cipher is called a substitution technique that combines linear and alphabetical characters to produce alphabets of new characters [6][7]. Vigenère cipher or what is often referred to as a polyalphabetic substitution cipher due to the vigenère cipher uses a 26 X 26 matrix that shifts [8]. Caesar cipher or often called the ROT3 algorithm is the most famous substitution technique by replacing the initial letters of the alphabet sequence which aims to improve message security [1][2][5]. In substitution techniques, there are several algorithms such as hill cipher, genetically engineered chip, caster cipher, one-time pad. Hill ciphers use modulo so hill cipher is called a substitution technique that combines linear and alphabetical characters to produce alphabets of new characters [6][7]. Vigenère cipher or what is often referred to as a polyalphabetic substitution cipher due to the vigenère cipher uses a 26 X 26 matrix that shifts [8]. Caesar cipher or often called the ROT3 algorithm is the most famous substitution technique by replacing the initial letters of the alphabet sequence which aims to improve message security [1][2][5]. In substitution techniques, there are several algorithms such as hill cipher, genetically engineered chip, caster cipher, one-time pad. Hill ciphers use modulo so hill cipher is called a substitution technique that combines linear and alphabetical characters to produce alphabets of new characters [6][7]. Vigenère cipher or what is often referred to as a polyalphabetic substitution cipher due to the vigenère cipher uses a 26 X 26 matrix that shifts [8]. Caesar cipher or often called the ROT3 algorithm is the most famous substitution technique by replacing the initial letters of the alphabet sequence which aims to improve message security [1][2][5]. In substitution techniques, there are several algorithms such as hill cipher, genetically engineered chip, caster cipher, one-time pad. Hill ciphers use modulo so hill cipher is called a substitution technique that combines linear and alphabetical characters to produce alphabets of new characters [6][7]. Vigenère cipher or what is often referred to as a polyalphabetic substitution cipher due to the vigenère cipher uses a 26 X 26 matrix that shifts [8]. Caesar cipher or often called the ROT3 algorithm is the most famous substitution technique by replacing the initial letters of the alphabet sequence which aims to improve message security [1][2][5].
Image Using LSB Steganography and Cryptography Vigenere Cipher Caesar" still need for improved data security and durability for the message in the image [10]. This research combines 4 modified substitution techniques and stored messages that have been hidden into an image with the LSB. This research aims to make improvements in using substitution techniques that have been carried out by many other types of research and also implement the results of encryption into image pixels using LSB algorithm so that the secret message reaches the goal of cryptography which is increasing message security [1][2][5].

2. Theoretical

2.1. Hill Cipher

In 1929 Hill cipher or Polyalphabetic crypto was discovered by a mathematician Lester S [1][2][8]. Hill ciphers use Modulo 26 and matrices to convert plaintext to another character [1][2][8]. These the formula from hill cipher [1][2][8]:

\[ C = (KP) \mod (26) \]  
\( P = (k^{-1}C) \mod (26) \)

In Equation 1 describes the mathematical process of encryption on a hill cipher while mathematical equations 2 explain the process of decryption on the hill cipher. The corresponding element in the equation 1 and 2 is the key "K", plaintext "P" and Ciphertext "C" [1][2][8].

2.2. Vigenere Cipher

Vigenere cipher is a substitution technique that each ciphertext can have many plaintext possibilities and this technique can be done in 2 ways, namely: numbers and letters [1][2]. The following Fig. 1 example of a vigenere using numbers as a process on the vigenere whereas Fig. 2 is the vigenere that use letters:

Fig 1. Vigenere use numbers
2.3. Caesar Cipher

Caesar cipher, known as the emperor's code in the reign of Julius Caesar and the emperor's code, was the first substitution code in the world of encoding [1][2]. The emperor code uses the ROT3 algorithm which has a pattern replacing the initial letter position of the alphabet. The formula on Caesar cipher is as follows [1][2]:

\[
C = E(P) = (P + K) \mod 26
\]  

In Equation 3 is a P as plaintext. Ciphertext as C, and K as the key that will be used in the process of the Caesar cipher [1][2].

2.4. One Time Pad

One time pad (OTP) is a cipher substitution that is included in symmetry cryptographic algorithms [1][2]. OTP was first put forward by Major J. Maag Borne and G. Vernam in 1917 [1][2][8]. The encryption formula on the OTP is as follows [1][2]:

\[
i_c = (p_i + k_i) \mod 26
\]  

Equation 4 describes the mathematical process in OTP. While \( p_i \) represents the plaintext and \( k_i \) have been determined keyword representing [1][2].

In OTP there are 2 basic difficulties, namely [8]:
   a. The problem with creating large numbers of random keys
   b. Key distribution and protection issues.

2.5. LSB

LSB is a steganography algorithm that aims to hide messages in pixels that are owned by images [14][15][16][17]. Steganography is different from cryptography, this is because steganography hides messages into images so that they do not arouse suspicion on the other side [17].
The embed process carried out in the picture is as follows [14]:

a. Calculates pixel image of object
b. Repeats according to the pixel of the image
c. Take RGB from the pixel that has been obtained
d. Make LSB at each pixel zero
e. Convert messages into binary form and store them in 8-bit pixels
f. Repeat all processes according to many messages that will be hidden.

The process of extracting embeds messages from images that have been inserted secret messages as follows [14]:

a. Calculates pixel image of object
b. Repeats according to the pixel of the image to get 8-bit continuously
c. Retrieves LSB that has been obtained based on pixel objects and converts it into characters.

3. Result and Discussion

In this study the author made several changes, especially in the cipher substitution process and implemented it into an image. The following are the paths carried out in this study:

![Flowchart Optimization Substitution Cipher and Implementation Hidden Text](image)

**Fig 3.** Flowchart Optimization Substitution Cipher and Implementation Hidden Text

3.1. **Optimization Substitution Cipher**

3.1.1. **Modification Vigenere cipher**

In the process of making key authors combine 2 algorithms namely the vigenere cipher and hill. 2 the algorithm-specific author does some enhancements and changes so that the implementation will experience a significant difference as well as modify tables that are frequently used in the font table namely vigenere shifted from letters leading to the last letter. The following Fig. 4 that describes the use of tables for the vigenere cipher process is conducted in the research done, besides the table replaces the table of standards that are commonly used:
3.1.2. Modification Hill Cipher

Especially for hill cipher, the authors make changes in the stages before entering into the matrix calculation. On the calculation of the Hill cipher, the writer keeps using a standard calculation of the hill cipher itself. The following modifications have done cipher hill. The following modifications to the hill cipher conducted by researchers:

Plaintext: SELAMAT PAGI INDONESIA

The plaintext is just an assumption to explain what is meant by researchers. In the implementation of the hill after the writer enters a cipher plaintext into a matrix, the next process is taking in vertical i.e. SAIN, ETIE, LPNS, AADI, MGOA and performed calculations in accordance with the standard of the Hill cipher. The calculation results obtained are organized in vertical form the original matrix. But the process is far from complete because researchers do retrieval based on the character from the matrix of horizontal section. In Fig. 5 and Fig. 6 will give an overview of the associated modifications done by researchers.

$$\begin{bmatrix}
6 & 1 & 8 & 5 \\
4 & 10 & 8 & 1 \\
8 & 5 & 10 & 7
\end{bmatrix} \times \begin{bmatrix}
12 \\
20 \\
7
\end{bmatrix} = \begin{bmatrix}
72 & 20 & 16 & 35 \\
140 & 200 & 18 & 7 \\
24 & 80 & 8 & 49
\end{bmatrix} = \begin{bmatrix}
149 \\
171 \\
161
\end{bmatrix} \mod 26 = \begin{bmatrix}
3 \\
12 \\
5
\end{bmatrix}$$
Fig 5. Modification hill cipher keyword 1

Fig 6. Modification hill cipher keyword 2

What needs to be considered is when keyword 2 in the character matrix changes to a hex number. This occurs because of a change when it does process vigenere cipher. How to cope with the changes of the researchers doing the conversion from hex numbers to alphanumeric characters.

3.1.3. Modification Caesar Cipher

In the Caesar cipher, the researcher made the additional process of calculating the characters from key 1 and the results of key 2 which were previously obtained from the generated keyword process. Once the calculation is done it will be done again the same character deletion key 1 and the result of the key 2. The calculation of a lot of the same character would be made to shift the parameters of the processes of the Caesar cipher. Fig. 7 will explain the results of generated keywords as follows:

Fig 7. Result generated keyword

3.1.4. Modification One-Time Pad (OTP)

The next process is One-Time Pad, researchers do some additions, namely converting hex to decimal in Fig.8. This is due to the result of a Caesar cipher will be letters while Key 2 will either be hex.

Fig 8. Modification convert hex to decimal

The OTP process is the last process for cryptography in this study. The results obtained from the plaintext that has been specified are "GQXVDAQSJLETYKZFACZD".

3.2. Implementation

This research has been done on some of the changes as mentioned above. The result of the substitution cipher is stored into the data of the image using the algorithm of LSB so that the security of the message that will be sent to be increased because the image has been inserted into the pixel object. App view can be seen in Fig. 9 as follows.
Fig. 9 is an application that stores messages into images with PNG extensions which originally had the image of the object having JPG extension. In this research when storing messages made using RGB but will not physically change the image that is the object of storing the message. In Fig. 10 will explain the difference between a picture before the inserted message and Fig. 11 describes the image after the message is inserted.

Fig. 10. Check color the image before the inserted ciphertext
Although the image does not change physically, the image will change hex in the image itself. The following hex on images into objects insertion messages in Fig. 12.

**Fig 11.** Check color the image after the inserted ciphertext

**Fig 12.** Hex image object

Fig. 12 may explain the hex changed significantly however when done checking color against image object does not undergo changes this can be an excess of this research. Pictures that have been pasted messages do not experience physical changes and does not engender suspicion to the other party. However, in this research to get some difference between pictures. It is described in table 1 What are the different image is present on the object.

**Table 1.** Detail image source and detail image result

|               | Image Source | Image Result |
|---------------|--------------|--------------|
| Size          | 83.5 KB      | 268 KB       |
| Dimensions    | 650 x 400    | 650 x 400    |
Table 1 has explained that the differences on the original image with the result image or pictures that have been pasted message did not experience the difference was striking. It can be seen in the histogram of each image that there is absolutely no difference in numbers occurred even though the image has been inserted a secret message. In addition, this research was done in some implementations a picture that has been inserted a message on social media trend today as seen in table 2:

| Social media          | Result Decrypt Message |
|-----------------------|------------------------|
| Facebook Chat / Post Facebook | No                     |
| Whatsapp              | No                     |
| Google+               | Yes                    |
| Google Drive          | Yes                    |
| Pinterest             | No                     |
| Instagram             | No                     |
| Telegram              | Yes                    |

4. Conclusion

This research concluded that the optimization is done has been successful as seen when testing colors on an image that hasn't been inserted a message with images already pasted message does not suffer changes to images into objects insertion of messages. It will not give the impression there is an oddity in the image due to changes at all does not look physically but when testing the hex index in the image. This research also gets some of the facts that the insertion of the message in the picture will be lost when the implementation on social media such as Facebook, Whatsapp, Pinterest, and Instagram. Implementation on social media will damage the contents of the message because social media has a system that converts images into smaller compared to the actual image. But implementation on social media that support the research is Google +, google drive, and the Telegram. The last conclusion was this research can save the size of the image that has been inserted the message. Image used 650 x 400 get the size only of 268 KB this indicates this research has been perfected at the research done by previous researchers.

5. Future work
Suggestions for the next researcher can develop media insertion a more effective message to implementation on social media. so as to provide a direct impact to the lay person though.

References

[1] Ariyus, Dony. “Pengantar Ilmu Kriptografi: Teori Analisis & Implementasi”. Penerbit Andi, 2017.
[2] Ariyus, Dony. "Kriptografi keamanan data dan komunikasi." Yogyakarta: Graha Ilmu, 2006.
[3] Katz, Jonathan, et al. Handbook of applied cryptography. CRC press, 1996.
[4] Paar, Christof, and Jan Pelzl. Understanding cryptography: a textbook for students and practitioners. Springer Science & Business Media, 2009.
[5] Anderson, Ross J. Security engineering: a guide to building dependable distributed systems. John Wiley & Sons, 2010.
[6] Siahaan, Andysah Putera Utama. "Genetic Algorithm in Hill Cipher Encryption." 2017.
[7] Khalaf, Ashraf AM, Mona S. Abd El-Karim, and Hesham FA Hamed. "A triple hill cipher algorithm proposed to increase the security of encrypted binary data and its implementation using FPGA." Advanced Communication Technology (ICACT), 2016 18th International Conference on. IEEE, 2016.
[8] Stallings, William. Cryptography and Network Security, 4/E. Pearson Education India, 2006.
[9] Lubis, Fahrul Ikhsan, et al. "Combination of Caesar Cipher Modification with Transposition Cipher." Advances in Science, Technology and Engineering Systems Journal, 2017. 2.5: 22-25.
[10] Dewangga, I. Gede Arya Putra, Tito Waluyo Purboyo, and Ratna Astuti Nugrahaeni. "A New Approach of Data Hiding in BMP Image Using LSB Steganography and Caesar Vigenere Cipher Cryptography." International Journal of Applied Engineering Research 12.21, 2017: pp. 10626-10636.
[11] Saraswat, Aditi, et al. "An Extended Hybridization of Vigenere and Caesar Cipher Techniques for Secure Communication." Procedia Computer Science 92, 2016. pp. 355-360.
[12] Saputra, Imam, Nelly Astuti Hasibuan Mesran, and Robbi Rahim. "Vigenere Cipher Algorithm with Grayscale Image Key Generator for Secure Text File." International Journal of Engineering Research & Technology (IJERT) 6.1, 2017. pp.266-269.
[13] Nasution, Surya Darma, et al. "Data Security Using Vigenere Cipher and Goldbach Codes Algorithm." Int. J. Eng. Res. Technol 6.1, 2017. pp. 360-363.
[14] Panghal, Sandeep, Sachin Kumar, and Naveen Kumar. "Enhanced Security of Data using Image Steganography and AES Encryption Technique." International Journal of Computer Applications (0975–8887) Recent Trends in Future Prospective in Engineering & Management Technology 2016.
[15] Al-Afandy, Khalid A., et al. "High security data hiding using image cropping and LSB least significant bit steganography." 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt). IEEE, 2016.
[16] Bhardwaj, Rupali, and Vaishali Sharma. "Image steganography based on complemented message and inverted bit LSB substitution." Procedia Computer Science 93, 2016. pp. 832-838.
[17] Rifqi, Muhammad, and Muhammad Nashar. "COMBINING STEGANOGRAPHY AND CRYPTOGRAPH TECHNIQUES FOR DATA SECURITY (CASE STUDY IN PT XYZ)." world 5.01, 2018.