Combining caesar cipher and hill cipher in the generating encryption key on the vigenere cipher algorithm

Z Qowi and N Hudallah

1Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Semarang, Indonesia

Corresponding author: zqowi97@students.unnes.ac.id

Abstract. There is a risk in the process of exchanging information, especially text information. To reduce that risk, one of cryptographic method can be applied. That is Vigenere Cipher algorithm. But the Vigenere Cipher algorithm have a weakness. The weakness is a repetitive encryption key. It causes the ciphertext to be predictable with the Babbage-Kasiski method. By combining Caesar Cipher and Hill Cipher methods in the process of generating the encryption key, its expected to cover up the weakness of the Vigenere Cipher method. The combination has the capability to hide the character appearance’s frequencies.

1. Introduction

Information technology today is developing at a very fast pace. Even information technology became a basic requirement in various fields in the industrial world. With this kind of development, the process of exchanging digital data such as text, images, audios and videos can be done quickly. But, there is also a high risk of duplicating digital data [1]. A message that contains important information can be misused by irresponsible people [2]. To minimize this risk, a security system is needed to maintain the confidentiality of important data and one of them is by applying a cryptographic technique [3]. One of cryptographic methods is vigenere cipher which utilizes a key in the form of characters to encrypt the plaintext [4]. However, this method has a weakness in the form of a repetitive key and as a result, the ciphertext can be easily broken [5]. But in another method, hill cipher, has the advantage of hiding the frequency of appearance of letters [6]. This method using modulo and matrix arithmetic so it can hide the frequency of appearance of letters [7]. Hill cipher method uses a square matrix as a key that used for encryption and decryption [8]. By utilizing hill cipher’s advantage and using this method and combining it with other methods, Caesar cipher, to generate the key, it will be able to minimize the weaknesses of the vigenere cipher.

2. Methods

The vigenere cipher, caesar cipher, and hill cipher were used in this research. Caesar cipher and hill cipher are used to form the initial key to form new keys. Key entered by the user will be repeated so that it has the same length as the message that user has entered. Then the key is encrypted using caesar cipher with the shift value according to the character value in the key. Then the information obtained from caesar cipher will be encrypted again using hill cipher algorithm to hide the frequency of appearance of characters in the formed key. The result will be used to encrypt text data using vigenere cipher algorithm.
The programming language used in this study is C++. Modification of vigenere cipher in this study supports the use of special characters, uppercase and lowercase letters according to ASCII Printable Character.

2.1. Algorithm
Generating of a new algorithm in the application of vigenere cipher begins with processing the key (K) that entered by user. Key that entered by user will be repeated so it has the same character length as the message length. The next process is to form a key that will be used to encrypt the message or plain text (P). The first step to form the key is carried out using caesar cipher algorithm. The repeated key will be encrypted with the value of character position. For example there is a key “ABC”. The “A” will be encrypted with the value 1. The “B” will be encrypted with the value 2 and so on. The result of the first step is called the first key (K1). The final key will be formed on the next step using hill cipher algorithm. the key used in formation of this final key is a 4x4 matrix. It contains the ASCII value of the original key which is repeated so the key matrix has a length of 16 characters. Meanwhile, K1 will be used as plaintext at this step. K1 will be divided into 4 letter blocks because the key matrix used is 4x4. Then the key matrix will be multiplied by each block of K1. The result will be called the final key (FK). The next step is process of encrypting plaintext (P) using vigenere cipher algorithm with the final key (FK).

2.2. Mathematical Model

2.2.1. Caesar Cipher
Caesar cipher is used as a key generator to be used in the message encryption process. The caesar cipher algorithm will shift the value of the message character (P) as far as the key (K), where K is an integer value [9]. For example, there is a message “SIX” and it is shifted as far as K = 3, then the cipher text is “VLA”. In the caesar cipher, letters A, B, C, …, Z will be labeled with the values 0, 1, 2, …, 25 [10]. The mathematical model for the caesar cipher can be calculated using equation (1):

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

Information :
- C = Cipher text
- E(P, K) = Encrypting P using K
- P = Plain text
- K = Shifting key

This study using a modified vigenere cipher so that the caesar cipher algorithm will changed. Changes are made by using the message position value (i) as the key and the modulus value used is 95. Equation (2) is a modification of equation (1).

\[
C = E(Pi, i) = (Pi + i) \mod 95
\]

Information :
- C = Cipher text
- E(Pi, i) = Encrypting Pi using i
- Pi = The \(i^{th}\) plain text character
- i = Shifting key

2.2.2. Hill Cipher
Hill cipher is one of symmetric cryptography that uses \(n \times n\) matrix. Several numbers will be entered into the matrix and used as a key to encrypt the plain text [11]. Mathematically, the encryption process
of plaintext \((P) = (p_1, p_2, p_3, \ldots, p_n)\) with a key matrix \((K) = (k_1, k_2, k_3, \ldots, k_n)\) and will produce ciphertext \((C) = (c_1, c_2, c_3, \ldots, c_n)\).

\[
\begin{pmatrix}
  c_1 \\
  c_2 \\
  c_3 \\
  \vdots \\
  c_n
\end{pmatrix} =
\begin{pmatrix}
  k_{11} & k_{12} & k_{13} & \cdots & k_{1n} \\
  k_{21} & k_{22} & k_{23} & \cdots & k_{2n} \\
  k_{31} & k_{32} & k_{33} & \cdots & k_{3n} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  k_{n1} & k_{n2} & k_{n3} & \cdots & k_{nn}
\end{pmatrix}
\begin{pmatrix}
  p_1 \\
  p_2 \\
  p_3 \\
  \vdots \\
  p_n
\end{pmatrix} \mod 95 \tag{3}
\]

2.2.3. Modified Vigenere Cipher

Vigenere cipher is a polyalphabetic substitution cipher and was developed from a modified Caesar cipher. Vigenere cipher is considered as the most secure encryption system compared to other polyalphabetic substitution ciphers [12]. In the Vigenere cipher, the keys used are in the form of characters entered by user. For example, there is a key “ENCODE” and the message “THE SKY IS FALLING”. The encryption process begins by matching each letter with a number from 0 to 25. The result is obtained by adding the message value with the key. The result will be reduced by 26 if the result is more than 25. The cipher text of the encryption is “XUGGNCMFHOOPMAI” [13]. Mathematical model of the Vigenere cipher encryption algorithm can be calculated as equation (4):

\[
C = E(x) = (x + n) \mod 26 \tag{4}
\]

Information :

- \(E(x)\) = Encryption of character \(x\)
- \(x\) = Character from message
- \(n\) = Character from key

This study using a modified Vigenere cipher so that the Vigenere cipher algorithm will changes. Changes are made by changing the modulus value to 95. Equation (5) is a modification of equation (4).

\[
C = E(x) = (x + n) \mod 95 \tag{5}
\]

Information :

- \(E(x)\) = Encryption of character \(x\)
- \(x\) = Character from message
- \(n\) = Character from key

3. Results and Discussion

3.1. Results

The message used is “IN THE FOREST THERE ARE MANY TREES WITH THE SAME HEIGHT FOR EXAMPLE MANY” with the key “KIRA” [14].

3.1.1. Vigenere Cipher’s Result
Figure 1. Screenshot of the execution result of the console program simulation

Figure 1. shows that the results have two repeated words in the cipher text (“NFRO” and “VMKVP”). This is due to the repetition of words in the final key, so there is a chance to predicting the message using Babbage-Kasiski method [15].

3.1.2. Modified Vigenere Cipher’s Result

```plaintext
<< DISPLAY INPUT
IN THE FOREST THERE ARE MANY TREES WITH THE SAME HEIGHT FOR EXAMPLE MANY
Input Key : KIRA
Key Matrix becomes = KIRA KIRA KIRA
Key becomes = KIRA KIRA KIRA
Final Key = +E/NBdWjSV3 Yp*hY/(lpd?w?9:Vnaome*EK\Rz<sZpSJC8j4k<~8419\lFRP<qMq~suG6j>
<< DISPLAY OUTPUT
Ts/##j4nni)eE-E\wN56d\j^~:08Im:jj07RRe{Cp(rh@9b!"~g)Yz"WfuVwUk-Un?O@;6c,IV
```

Figure 2. Screenshot of the execution result of the console program simulation

Different things can be seen in Figure 2, where the simulation results don’t have repetitive words in the cipher text. This is because the message is encrypted by various character compositions so the message can’t be predicted.

3.2. Discussion

3.2.1. Predictable Information Chance Value

Figure 3. shows that the comparison of the predicted information has some of the same information as the original. The first line in Figure 3. is the original message, while the second line is the prediction result of the message using Babbage-Kasiski method.

```
INTHE  FORES T
TEE  WOEJ
AREM  ANYTR
BAIEJ  AEYQR
THE  SAMEH
TYE  PADEE
OREX  AMPLE
WOEO  AJPCE
```

Figure 3. Results of information prediction using Babbage-Kasiski method

Based on figure 3, it can be seen that the chance to knowing the message is 50.84%. It proves the vigenere cipher algorithm has an opportunity on data/information stealing.
Figure 4. Comparison of message prediction probability between vigenere cipher and modified vigenere cipher

Figure 4. shows that the comparison between vigenere cipher and modified vigenere cipher is too large. The largest probability to predicting information on vigenere cipher algorithm is 67.80% for key with 3 characters length. Meanwhile, the message information encrypted using modified vigenere cipher has 0% chance to be predicted. That proves the modified vigenere cipher algorithm is more secure than vigenere cipher algorithm.

4. Conclusion
The simulation results show the vigenere cipher algorithm has repeated words on its final key so there is a chance to predicting the information. But, the modified vigenere cipher algorithm doesn’t have it so the information can not be predicted.

References
[1] Fan Y, Lin X, Liang W, Tan G and Nanda P 2019 Future Gener. Comput. Syst. 101 127
[2] Subandi A, Meiyanti R, Mestika S C L and Sembiring R W 2017 Adv. Sci. Technol. Eng. Syst. 2 1
[3] Kumari S 2017 Int. J. Eng. Comput. Sci. 6 20915
[4] Kester Q 2013 Int. J. Adv. Technol. Eng. Res. 3 141
[5] Soofi A A, Riaz I and Rasheed U 2016 Int. J. Sci. Technol. Res. 5 141
[6] Chinonso O, Omoruyi O, Okokpujie K and John S 2017 Covenant J. Eng. Technol. 1 65
[7] Hraoui S, Gmira F, Abbou M F, Oulidi A J and Jarjar A 2019 Procedia Comput. Sci. 148 399
[8] Sharma N and Chirgaiya S 2014 Int. J. Comput. Appl. 108 34
[9] Goyal K 2013 Int. J. Comput. Appl. 73 26
[10] Limbong T and Silitonga P D P 2017 Int. J. Eng. Res. Technol 6 175
[11] Putera A, Siahaan U and Rahim R 2016 Int. J. Secur. Appl. 10 173
[12] Salomon D 2012 Data Privacy and Security (Northridge: Springer Science & Business Media)
[13] Prabowo H E and Hangga A 2015 Prosiding SNST 1 77
[14] Kester Q 2012 Int. J. Adv. Res. Comput. Eng. Technol. 1 108
[15] Flaut C, Flaut D, Hoskova-Mayerova S and Vasile R 2019 Adv. Mil. Technol. 14 81