khazanah informatika
Jurnal Ilmu Komputer dan Informatika

Analysis and Implementation of Steganography Using Playfair Techniques and DNA Substitution To Improve Message Security

Bambang Harjito [1]*, Dwiko Satriyo.U.Y.S [2], Faishal Rahutomo [3]
*bambang_harjito@staff.uns.ac.id
[1, 2] Informatics Department
Universitas Sebelas Maret
Surakarta, Indonesia
[2] Electrical Engineering Department
Universitas Sebelas Maret
Surakarta, Indonesia

Abstract- Along with the development of the internet, the need for data security is increasing. Ensuring the security of data requires special ways. Among these are cryptography and steganography. Because of the need for this method, many new data security methods have emerged. One of them is using DNA media. DNA cryptography and steganography utilize the properties of DNA to secure and hide secret messages. This paper proposes a combined method between the Playfair cipher cryptographic technique and the DNA substitution steganography algorithm for data security. Comparative analysis was conducted on techniques that are being tested by using similar methods. Four tests are conducted to measure the performance of the proposed method and another method was tested for comparative purposes. Four tests that were conducted are hiding capacity and key length test, speed test, endurance test, and steganalysis test. The result shows that the hiding capacity and key length test is better than other techniques. Steganalysis test result is also better than other tested techniques. Furthermore, in the speed test and endurance test, the results are quite satisfactory, not far from the others technique that is tested. The proposed method had a better performance than any other method that get tested, especially in speed performance and endurance performance.

Keywords: Cryptography DNA, Steganography DNA, Data Hiding, Playfair and substitution DNA

Article info: *submitted May 11, 2022, revised September 02, 2022, accepted September 05, 2022*

1. Introduction

In the digital era where data exchange across the globe became very important in every aspect of human lives, data security became a significant concern. There are many methods to achieve data security, but the methods that are commonly used are cryptography and steganography. Cryptography needs complexity and data randomness and steganography needs randomness and big cover capacity to increase the secureness of private data. DNA (Deoxyribonucleic Acid) approach for Cryptography and Steganography method can become a solution. DNA approach when applied in cryptography can increase the complexity and randomness of ciphertext and when applied in steganography can increase cover capacity and randomness.

DNA cryptography is a study of cryptography technique that exploits the properties of DNA sequence to create new cryptography technique or reinforce existing technique. DNA cryptography simulates transcription and translation from DNA sequences for cryptography [1]. DNA-based cryptography gets much attention because of DNA's complex structure [2-4]. The result of that method is strong against certain attacks, especially brute force attacks [5, 6]. On the other hand,

DNA Steganography is a branch of steganography that uses DNA data as a data hiding medium [7]. DNA steganography utilizes the large size and the complexity of the DNA sequence as a medium for hiding data to prevent other parties reading the message [8, 9]).

DNA is formed from 4 types of nucleic acids, namely A, C, G, and T whose characters form a series of DNA sequences [10]. DNA sequences can be used as an encryption medium because of their complex properties. DNA sequences can be converted into proteins form that is composed of 4 DNA characters. Each of those proteins has certain ambiguities that make them complex. In addition, DNA sequences can also be used as a medium for data concealment [11]. Unlike, audio or video, which can be noticed immediately when the signal is too degraded, DNA is hard to filter to see if there is a watermark in it [12]. Therefore, the attacker is forced to perform a direct match against the existing DNA.

Many studies were done on the topic of DNA cryptography and steganography. The previous study that discussed DNA cryptography are DNA techniques with RSA [13], DNA technique with Playfair cipher [12, 14], DNA-based encryption using pointers [15], DNA encryption using ECC[16], DNA encryption using El Gamal [17], DNA symmetric encryption using Feistel network coding [18], DNA encryption for cloud data sharing using proxy[19], DNA encryption for cloud computing based on genetics technique combined with logical-mathematic function[20], DNA cryptography based on genetic algorithm [21], and DNA encryption using PCR [22]. On the other hand, there is a researcher that discusses DNA steganography. For example, there is a study that proposes the method of insertion, substitution, and complementary pair-rule DNA technique [23]. Other research is the study that proposes the method of least significant base substitution and studies about a method that uses table lookup substitution base for DNA steganography [24]. There are also studies about a lossless DNA data hiding approach [11], and a steganographic scheme based on chaotic maps and DNA computing [25].

Several studies do comparative tests on DNA cryptography algorithms and DNA steganography algorithms. One of the studies that did a comparative test on DNA cryptography was research by Marwan [26]. Marwan et al, compare the Vigenère cipher, Playfair cipher, RSA cipher, and AES cipher when combined with DNA encoding techniques. The result is that in the cryptoanalysis test, the RSA, Playfair, and AES algorithms have the same value. Vigenère cipher has the lowest value in cryptoanalysis testing. For key capacity test, the best key capacity is in the Playfair cipher, followed by AES, RSA, and Vigenère cipher. In speed testing, the best time performance is the Vigenère cipher, followed by AES and Playfair cipher with similar score, and then RSA with the worst time performance. Then there is testing the capacity of the secret message that can be encrypted. The result is that the Playfair cipher, The Vigenère cipher, and AES have the same score and RSA has the smallest score. In addition to DNA cryptography, there are several studies on the comparison of DNA steganography methods. One of them is research by Shiu [23]. In this study, the methods of Insertion, Complementary Pair Rule, and Substitution were compared. The aspects compared are cracking probability with the best order, namely Insertion, Complementary Pair Rule, and Substitution. Another aspect is the data hiding capacity with the best Substitution method, followed by insertion and Complementary Pair Rule. Next, there is the payload aspect that is tested. Payload is the remainder of the initial cover after extraction. The best is the Complementary Pair Rule, followed by Insertion and Substitution.

2. Methods

This section contains an explanation of our proposed method and the methodology used in the research. Our method combined the Playfair cipher cryptographic technique and DNA substitution steganography algorithm to secure secret messages. A general overview of the concealment and extraction stages of the secret message is provided by using the Playfair-DNA cryptography method and DNA substitution steganography. The concealment and extraction model consists of two processes, namely (a) the concealment stage of the secret message and (b) the extracting stage of the secret message. It can be seen in Figure. 1.

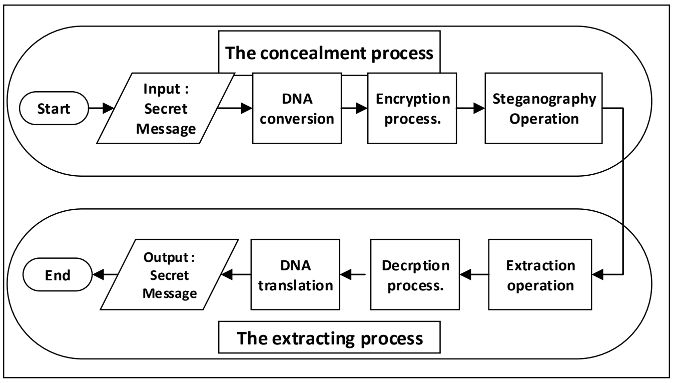


Figure. 1 Concealment and Extraction stages of secret message

a. The Concealment Process

The concealment process consists of three steps, namely (1) DNA conversion that is, converting the secret message to DNA form, (2) the encryption of the DNA form, and (3) the steganography process, which is doing steganography to produce the stego-DNA.

a.1. DNA Conversion

The first stage is the DNA conversion process, which is done by converting the secret message into DNA. Furthermore, encryption and steganography are performed to produce the stego-DNA. Stego-DNA is what will be sent to the recipient.

The plaintext is converted to DNA which is then converted to protein. To convert plaintext to protein form, the incoming data (binary form) is first converted to DNA sequence form. Converting binary form into DNA sequences is done in pairs, i.e. 2 bits are converted to one DNA character. The conversion rules are '00' to 'A', '01' to 'C', '10' to 'G', and '11' to 'U'. The complete flow is depicted in Figure 2.

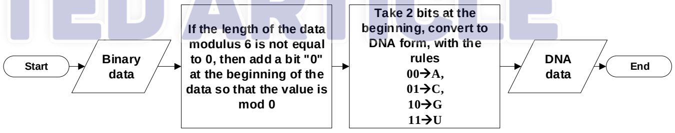

Start
Binary data
If the length of the data modulus 6 is not equal to 0, then add a bit "0" at the beginning of the data so that the value is mod 0
Take 2 bits at the beginning, convert to DNA form, with the rules
00→A,
01→C,
10→G
11→U
DNA data
End

Figure. 2 Binary to DNA Conversion

After obtaining the plaintext in the form of DNA sequences, then the DNA sequences are converted into protein for encryption purposes in the next stage. The conversion of DNA sequence is done by translating 3 DNA to 1 protein. The conversion of 3 DNA is listed in Table 1.

Table 1. Sequence to 3 DNA conversion

CODE	CONVERSION	CODE	CONVERSION
A	GCU, GCC, GCA, GCG	L	CUU, CUC, CUA, CUG
R	CGU, CGC, CGA, CGG	K	AAA, AAG
N	AAU, AAC	M	AUG
D	GAU, GAC	F	UUU, UUC
C	UGU, UGC	P	CCU, CCC, CCA, CCG
Q	CAA, CAG	S	UCU, UCC, UCA, UCG
E	GAA, GAG	T	ACU, ACC, ACA, ACG
G	GGU, GGC, GGA, GGG	W	UGG
H	CAU, CAC	Y	UAU
I	AUU, AUC, AUA	V	GUU, GUC, GUA, GUG
B	UAA, UGA, UAG	O	UUA, UUG
U	AGA, AGG	X	AGU, AGC
Z	UAC		

One character of protein code is converted into three characters of DNA. For example, the Protein code of A can be converted into 3 DNA codes GCU, GCC, GCA, or GCG. The ambiguity number determines the DNA code that gets selected from the conversion step. The conversion stage from DNA form into protein form is depicted in Figure 3.

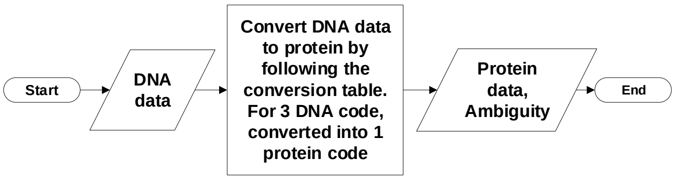

Start
DNA data
Convert DNA data to protein by following the conversion table. For 3 DNA code, converted into 1 protein code
Protein data, Ambiguity
End

Figure. 3 Converting DNA to Protein Form

The input of the conversion stage is DNA data. DNA data is then converted into protein with 3 DNA codes into 1 protein code and 1 ambiguity number. The output of the conversion stage is protein data and the ambiguity of each protein code.

a.2. Encryption of the DNA form

The second stage is to perform the encryption process. After obtaining the plaintext in the form of protein, The encryption process with the Playfair cipher is carried out by generating the key table and substituting the plaintext using the key table created. The flow is described in Figure 4.

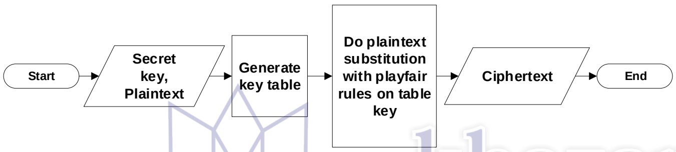


Figure. 4 Playfair Cipher Encryption

The input of the encryption stage is a secret key and plaintext. The first step is generating a key table for the substitution cipher. The next stage is encrypting plaintext using the key table. The output of the encryption stage is Ciphertext

a.3. Steganography Operation

The third stage is to do the steganography process. The results of the encryption are first converted into binary form. The technique used in the steganography process is DNA substitution. The data were substituted into the concealment cover media prepared by this technique. The first step is to generate a random number uniquely with the number of numbers with similar size as the inputted plaintext. The second step is the process of substitution of the plaintext on the concealment media/DNA cover media to produce the Stego-DNA. The process is described in Figure 5.

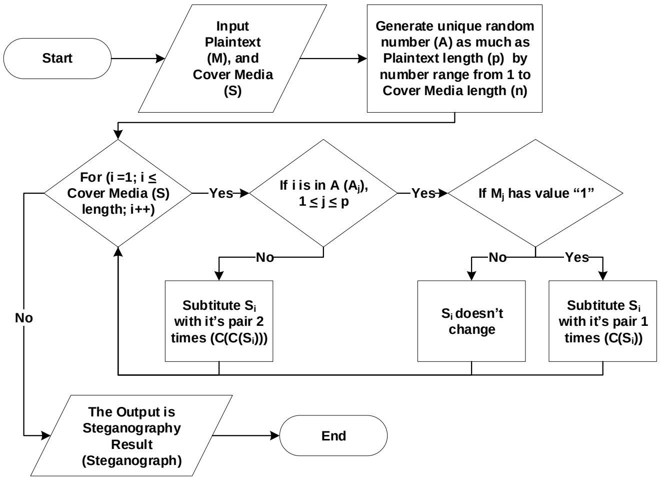

Start
Input Plaintext (M), and Cover Media (S)
Generate unique random number (A) as much as Plaintext length (p) by number range from 1 to Cover Media length (n)
For (i =1; i ≤ Cover Media (S) length; i++)
Yes
If i is in A (Aj), 1 ≤ j ≤ p
Yes
If Mj has value "1"
No
Subtitute Si with it's pair 2 times (C(C(Si)))
No
Si doesn't change
Yes
Subtitute Si with it's pair 1 times (C(Si))
No
The Output is Steganography Result (Steganograph)
End

Figure. 5 Substitution Steganography Embedding Technique

The input of the steganography stage is plaintext and cover media. The first step is generating unique numbers by the size of plaintext ranging from 1 to the length of cover media. And then do a substitution process from plaintext to cover media. The output of the steganography stage is the Steganography result.

b. The Extracting Process

The extracting process consists of three steps, namely (1) the process of extracting messages from the Stego-DNA, (2) the decryption process (3) the DNA translation process.

For the process of getting the secret message back, the steps that are carried out only reverse the previous stages. Starting from the extraction process, decryption process, and finally translation process.

b.1. Extraction of The Stego-DNA

The first process is extraction. Extraction of messages from the Stego-DNA is done by comparing it to the cover media used when embedding the data. Then the DNA substitution extraction operation was carried out to obtain the hidden data. The flowchart form of the DNA substitution extraction process can be seen in Figure 6.

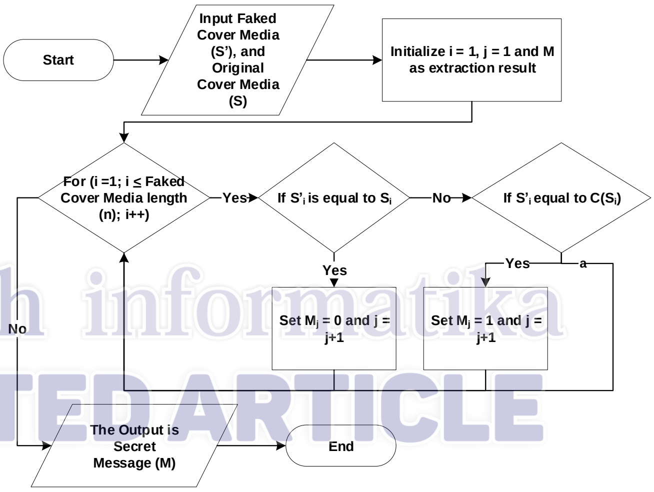

Start
Input Faked Cover Media (S'), and Original Cover Media (S)
Initialize i = 1, j = 1 and M as extraction result
For (i =1; i ≤ Faked Cover Media length (n); i++)
Yes
If S'i is equal to Si
No
If S'i equal to C(Si)
Yes
Set Mj = 0 and j = j+1
Yes
a
Set Mj = 1 and j = j+1
No
The Output is Secret Message (M)
End

Figure. 6 Substitution Steganography Extraction Technique

The input of the extraction stage is faked cover media and original cover media. The first step is initialized variable i for 1, j for 1, and M as extraction results. And then do the extraction process of faked cover media using original cover media. The output of the extraction stage is the secret message.

b.2. Decryption

The next step to getting the secret message is the decryption process. The Playfair cipher decryption process is almost the same as the encryption process. Only the substitution operations on the table are reversed when compared to the encryption process. The process can be seen in Figure 7.

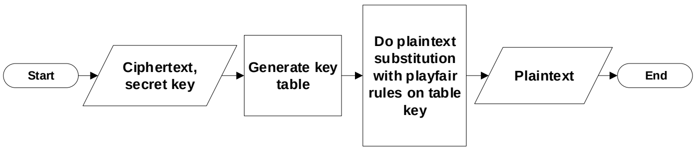


Figure. 7 Playfair Cipher Decryption

b.3. DNA Translation

The process of translating protein to DNA is not much different from the process from DNA to protein. What makes the difference this time is the use of ambiguity values obtained during the translation process from DNA to protein. The ambiguity value determines the converted DNA value of a protein. The flowchart of the process is depicted in Figure 8.

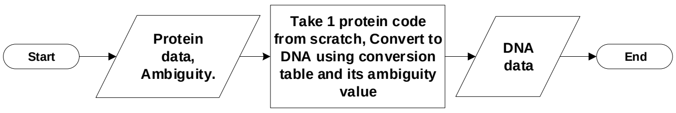

Start
Protein data, Ambiguity.
Take 1 protein code from scratch, Convert to DNA using conversion table and its ambiguity value
DNA data
End

Figure. 8 Protein to DNA Conversion

Just like the process of translation from protein to DNA, the process of translation from DNA to binary-only reverses the process of encoding. DNA is translated according to its rules. That is “A” to “00”, “C” to “01”, “G” to “10”, and “U” to “11”. The flowchart can be seen in Figure 9.

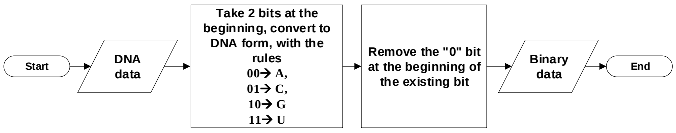

Start
DNA data
Take 2 bits at the beginning, convert to DNA form, with the rules
00→ A,
01→ C,
10→ G
11→ U
Remove the "0" bit at the beginning of the existing bit
Binary data
End

Figure. 9 Convert DNA to Binary

2.1. Data Preparation

The data used are secret messages, secret keys, and DNA data as a concealment cover. The secret message can be binary data or written data but for input into the system, the secret message is binary. The secret key is in the form of letters of the alphabet because the Playfair cipher is alphabetic. And the cover media used is in the form of DNA data, it can be original DNA genetic data or digital data that is converted to DNA form.

2.2. Playfair Chipper

Playfair ciphers are included in the category of Polygram substitution ciphers, namely substitution ciphers that substitute plaintext in groups of characters/more than one. Playfair cipher breaks plaintext into diagrams or groups containing 2 characters each and then encrypted against the key table generated at the beginning. The key table is in the form of a 5x5 matrix containing the alphabets that make up the key table [27].

2.3. Steganography

Steganography is an act of secret communication, which means that only the sender, Alice, and the recipient, Bob, are aware of the secret communication [28]. Steganography inserts confidential digital information into cover media (images, audio, video, etc.) without being known by others. Most applications of steganography follow the general principle that Alice, wishing to share a secret message m with Bob, randomly selects (using a private random source r) a harmless message, called a cover- object, which can be transmitted to Bob without causing any harm or suspicion. Alice embeds a secret message into c, using a k key, called the Alice key. Alice then changes cover c to the Stego-object s so that third parties only know that only cover c is transmitted that hasn't been changed to the Stego-object s[29].

The main applications of steganography include confidential communication, copyright protection, integrity verification, and authentication [30]. When designing a steganographic scheme, it is necessary to pay attention to aspects of the properties of the communication channel, the cover source used as a Steganography medium, and also the embedding and extraction functions[28]

2.4 DNA

Computers make us familiar with the concept of information as something that can be counted (in the form of bits). It also makes us aware that information can be stored in other physical forms. Living cells, like computers, can store and process information. This information is stored in the form of long-chain structures that make up DNA [31]. In 1953, the structure of DNA was correctly predicted by Watson and Francis Crick that the DNA molecule consists of two long polynucleotide chains, each of which is known as a DNA chain, or a DNA sequence made up of simple subunits, called nucleotides. Each nucleotide consists of a sugar-phosphate molecule with a nitrogen- containing side group, or base. Bases are of four types (adenine, guanine, cytosine, and thymine), corresponding to four different nucleotides, labeled A, G, C, and T[31]. Each nucleotide can represent by two bits[32]. Assumed that A = 00, C = 01, G = 10, T = 11. Then each alphabet can be represented by two bits. Every three adjacent nucleotides form a codon. Given that each nucleotide can have one of four chemical bases and each codon consists of three nucleotides, there is a total of 43 or 64 different possible combinations [8, 33]. This combination determines the amino acids to be used by living organisms, the arrangement of which determines the structure and function of the resulting protein[34].

DNA is used for cryptography to serve as the information carriers and also implement biological operation DNA as an arithmetic principle [35]. Biological operation DNA that can be used for cryptography is DNA encoding and DNA translation [36]. DNA cryptography gained popularity because it works like natural operation DNA but does not involve the actual operation in the laboratories [37].

In the DNA Steganography technique, DNA sequences are used as a hiding medium from secret data. DNA data is inserted into secret data with a certain steganographic algorithm so that the secret data can only be known by the intended recipient. DNA is very suitable for data hiding medium because of its big hiding capacity [33]. The use of DNA data itself started from research in 1999[38] by hiding the secret data into DNA data which is then made into microdot.

The utilization of DNA data as a medium for hiding messages is carried out in various forms. Some examples are in the form of microdots, microarrays, or a new technique, namely engineered cells. Microdot is a message hiding technique used by German spies in World War 2 by shrinking written messages so that they are invisible and pasted on paper media such as letters, documents, or others [38]. Then, the microarray technique is to duplicate the sequence of a DNA sample on certain media such as agar and silicon. The representation can be in 2 dimensions or 3 dimensions[39]. The last technique is engineered cell or synthesizing the nucleus of a genome to insert genetic data in the form of a secret message [40].

3. Result

In this section, the results of the research are explained and, at the same time is given a comprehensive discussion. Results can be presented in figures, graphs, tables, and others that make the reader understand easily. The discussion can be made in several sub-chapters. It is strongly suggested that a comparison with results from other published articles is provided to give more context and to strengthen the claim of novelty.

3.1. Testing Result

The first test is the speed test. What is tested in this experiment is the encoding time and decoding time. Encode time is the time spent in the process from initial input, DNA conversion, and encryption to steganography embedding. Meanwhile, the decoding time is the time taken from the steganography extraction process, decryption, until the conversion back to the original data form when inputted. The graph of the results of the encoding test is listed in Table 2.

Table 2. Speed test of several methods

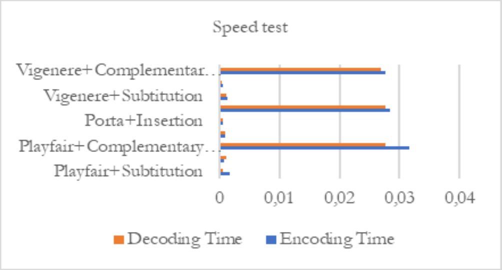
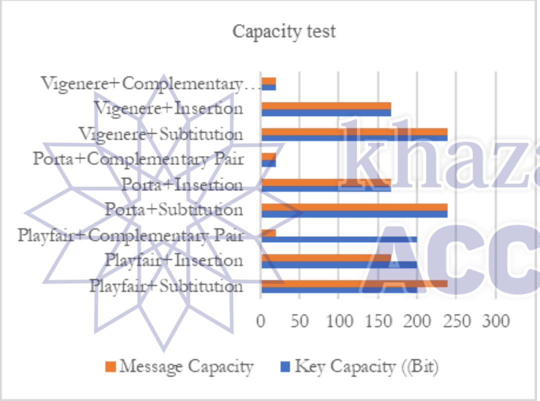


Table 2 shows the speed comparison between the tested methods. The proposed method is a Playfair cipher with DNA substitution in the far-left position. While the other method is a combination of other similar methods. The difference in speed between the methods is only slightly differ except for 3 methods, namely "Playfair + Complimentary Rule", "Porta+ Complimentary Rule", and "Vigenère + Complimentary Rule". The second test is the capacity test. Capacity testing tests the capacity of each algorithm. Trials were carried out on combination techniques of cryptography and steganography; the results are as in Table 3.

Table 3. Capacity Test of Several Methods



The third test is the endurance test. The complexity of each technique tested is in Table 4.

Table 4. Endurance test of several methods

Algorithm	Complexity
Playfair+Subtitution	$O(n^{115.5})$
Playfair+Insertion	$O(n^{117.5})$
Playfair+Complementary Pair	$O(n^{122.1})$
Porta+Subtitution	$O(n^{35.5})$
Porta+Insertion	$O(n^{37.5})$
Porta+Complementary Pair	$O(n^{42.1})$
Vigenere+Subtitution	$O(n^{36.5})$
Vigenere+Insertion	$O(n^{38.5})$
Vigenere+Complementary Pair	$O(n^{43.0})$

The last one is the steganalysis test. In the steganalysis test, the histogram of the result and similarity of each method to the original media coverage of the result of each method is calculated and compared for each other. If steganalysis in most steganography techniques, be it images, videos, or sounds, is looking for the level of similarity to the cover media, DNA steganography is sought for the level of dissimilarity with the initial cover media. That is because, in DNA steganography, the key to extracting a secret message is the cover media itself. The histogram of several methods is listed in Table 5.

Table 5. The Steganalysis Test Frequency Several Method

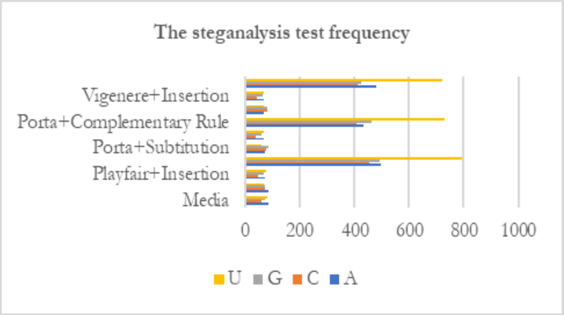
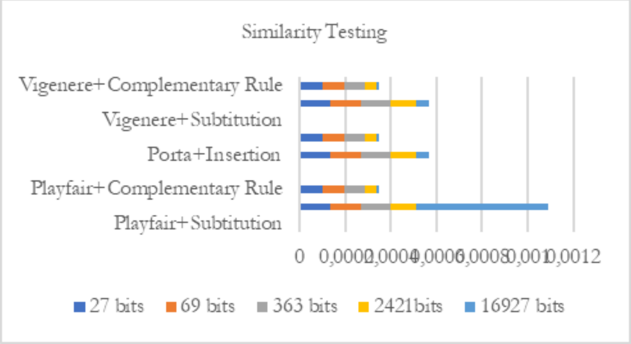


Table 5 shows a frequency table of the results of the compared techniques. The statistics of the amount of each DNA are almost close, except for the technique that uses the Complementary Pair Rule.

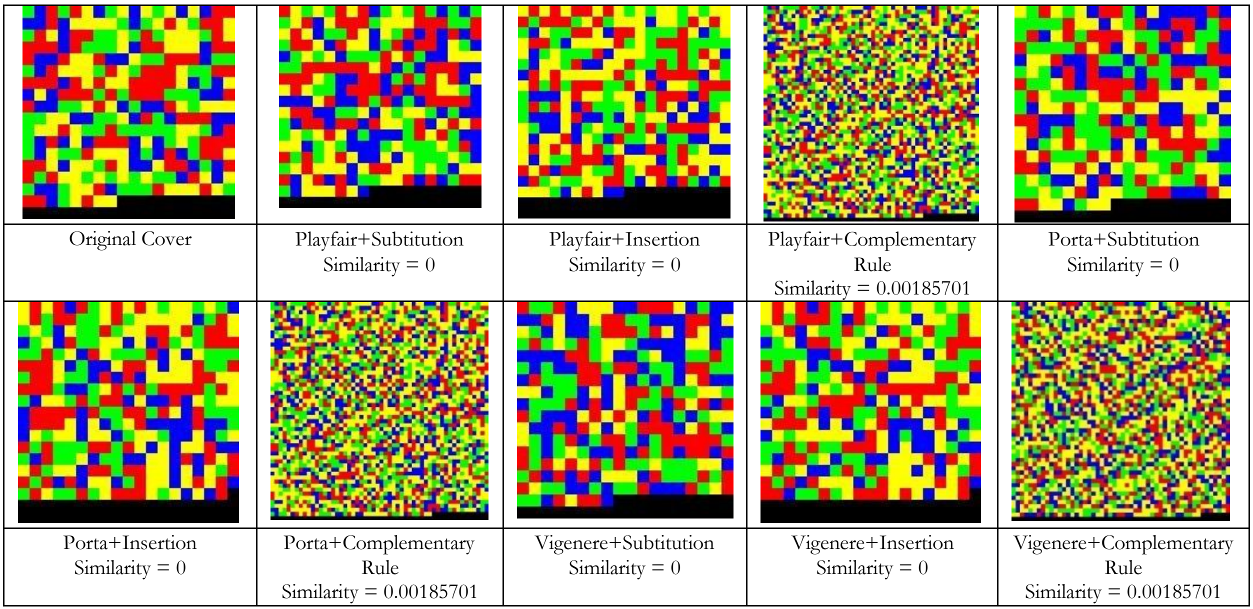
Next, media coverage is tested before the data insertion operation is performed compared to the media coverage that has already been inserted data. The similarity value before and after being inserted with a certain algorithm. It can be seen in Figure 10. The encode similarity with the original media cover is very little. There is a value of 0 or not similar at all. For the rest, the similarity value is below 0.003 or 0.3 percent. Furthermore, to clarify, an illustration is given with a picture of the DNA sequence pattern from DNA data, it can be seen in Figure 10. The illustration above is arranged in the following order: red pixels for DNA "A", green for "G", blue for "C", and yellow for "U". It can be seen from the comparison of the images that the distribution pattern of each image is much different from the original cover media. This is evidenced from the previous similarity table that the similarity with the original media cover is at most 0.0026 or about 0.26 percent. Thus, it is difficult to know the DNA medium used as the cover medium.

Finally, the similarity value is calculated with different inputs. The media cover being tested is MN988668 with a character length of 29881. There are several lengths of plaintext in binary that will be tested from the short one, which is 27 bits, to the long one, which is 16827 bits. The test aims to find the effect of the similarity value on the cover media with plaintext length. It can be seen in Table 6.

Table 6. Similarity Testing With Different Plaintext Lengths



The test results show that the addition of the plaintext length can affect the similarity value. The longer the inserted plaintext, the smaller the similarity value or similarity value. Except for techniques that use the substitution method, because the similarity value is already 0.



Original Cover

Playfair+Subtitution
Similarity = 0

Playfair+Insertion
Similarity = 0

Playfair+Complementary Rule
Similarity = 0.00185701

Porta+Subtitution
Similarity = 0

Porta+Insertion
Similarity = 0

Porta+Complementary Rule
Similarity = 0.00185701

Vigenere+Subtitution
Similarity = 0

Vigenere+Insertion
Similarity = 0

Vigenere+Complementary Rule
Similarity = 0.00185701

Figure. 10. The similarity value and visualization of Cover Media and Tested Techniques

4. Discussion

From the test results, several analyzes can be carried out on the results of the speed test, capacity test, endurance test, and steganalysis test

In terms of speed, almost all techniques have almost the same speed except for techniques that use the Complementary Pair Rule steganography method. So from the 3 methods of steganography used, namely the method of Substitution, Insertion, and Complementary Pair Rule. The Complementary Pair Rule method has the slowest speed, far from the other two methods. Meanwhile, for the speed of the cryptographic methods tested, namely Playfair cipher, Vigenère cipher, and Porta cipher, the speed does not differ much.

The Insertion method has a fast time because the essence of the method is to insert a secret message into the media cover. The substitution method performs substitutions on each character of the DNA cover media following the rules. Thus, the speed is slightly longer. As for the Complementary Pair Rule method, the speed is much longer because there is a search process for the longest DNA pair, both in insertion and extraction.

Next is the analysis of the test results on the plaintext capacity, key, and hiding capacity. The capacity of 3 cryptographic methods tested, namely Playfair cipher, Vigenère cipher, and Porta cipher, has a maximum of infinity. On key capacity, some algorithms depend on the size of the plaintext. Namely the Vigenère cipher and Porta cipher algorithms. This is because the two algorithms are included in the Shift cipher, which is a cryptographic algorithm that works by shifting the position of the message in the alphabet table. Thus, it takes several keys equal to the size of the plaintext. But if it is less than plaintext, you can repeat the key to the plaintext that does not get the key to shifting. Next, is testing the hiding capacity. Of the 3 algorithms used, message hiding capacity has no limitations except for the substitution technique. Two techniques other than substitution have a core operation by inserting messages directly between the media covers. Thus, the number of inserts does not have a length limit. However, the more data or messages that are inserted, the visibility of the message will be easier to see. For the last technique, namely substitution, the way the algorithm works is to replace or substitute the media cover with the message you want to insert into the media cover. Thus, the maximum limit that can be replaced is as large as the media cover. This technique also has limitations like the previous 2 techniques, namely the more data that is substituted for the cover media, the easier it is to see changes to the cover media

The next analysis is the endurance test. Resilience means the complexity of the algorithm while trying to disassemble it. In the encryption algorithm, the amount of complexity varies. For the Playfair cipher algorithm, the complexity is 25!. The value is obtained from the number of possibilities in the key table. The key table is 5x5 so there are 25 members. So the possible variations of the key table are 25! time. While the algorithm Vigenère cipher and Porta cipher. Complexity is highly dependent on key length. The complexity is n*13 for the Porta cipher and n*26 for the Vigenère cipher where n is the key length. 26 is obtained from the number of the alphabet, which is 26 characters. While the Porta cipher uses half so that it is worth 13. Then, in the test of the complexity of the steganography algorithm. The security of the DNA steganography algorithm is the media cover and the position of the secret message placement. Because of that, the complexity of DNA steganography ins depends on how big the DNA database is and the method that is used to embed secret messages to cover media.

The last test analysis is the steganalysis test. What is tested is the histogram statistic and the level of change in the media cover after insertion is carried out. When viewed from the histogram of the comparison between the insertion result and the initial cover media, each DNA character has an approximate length except for the technique that uses complementary pair-rule steganography. The ratio between characters, in each technique, is almost the same. Next is testing the total similarity between the insertion results and the initial cover media. Of all the techniques tested, the similarity value of the embedding results to the media coverage is very small. Ranges from 0 to 0.26 percent. This means that it is very difficult to know the cover media initially if you only look at the cover media that has been inserted by a secret message. So that the confidentiality of the media cover is maintained from other than the recipient who knows the cover media used. The last is similarity testing with different input plaintext lengths. The length of the plaintext affects the similarity value of the technique being tested. However, because the similarity value is already small, the difference is only slight.

5. Conclusion

Cryptography and DNA Steganography are new branches of science that still need a lot of development going forward. DNA cryptography still has the potential to be developed due to the unique nature of DNA.

In this study, a data hiding technique was tested using the Playfair cipher cryptography method and DNA substitution steganography. The method was tested with a similar technique, namely the Vigenère cipher cryptography method, and the Porta cipher. As for the steganography method, the comparison method is the Insertion and Complementary Pair method. What is tested are speed, capacity, endurance, and similarity.

In the speed test, the proposed algorithm gets relatively fast results. That is about 0.0011 seconds for the encode process, slightly adrift from the fastest method achieved by the Porta method with Insertion with a time of 0.0005 seconds.

In the capacity test, the key capacity of the proposed method is 25 characters. This number is smaller than the method that uses Vigenère and Porta encryption, namely the key capacity follows the length of the plaintext. The secret message length capacity is also smaller than other methods. The secret message length capacity of the proposed algorithm is smaller or equal to the length of the media coverage. This capacity is smaller than the method that uses the Insertion and Complementary Pair steganographic algorithms with no limit on the secret message capacity.

The endurance test shows that the complexity of the proposed method is quite high, which is O(n^122.1). This number is higher than the method that uses the Vigenère and Porta cryptographic methods. However, under a technique that uses the same encryption but a different steganography method.

The last one is the steganalysis test. The histogram values of all the tested techniques are almost the same except for those using the complement pair-rule steganography method. The ratio between the number of characters A, C, G, and U is almost the same for each technique tested. The similarity value of the proposed technique, namely the technique using the Playfair fair and substitution method, is 0.24 percent. This value is also not much different from other methods tested with a value range of 0 to 0.26 percent. This small value will make it difficult to find the initial media cover.

To increase the ability to hide the presence of secret messages in the media coverage, it is recommended to use additional visible media such as images, videos, or pdf documents. In addition, DNA techniques can be applied to asymmetric cryptography such as RSA, Elgamal, ECC, or POST quantum cryptography methods such as NTRU, SIDH, and McBits.

Reference

[1] B. Sayantani, K. Marimuthu, N. Mita, H. Anup Kumar, and R. Niranchana, "Bio-inspired cryptosystem with DNA cryptography and neural networks," Journal of Systems Architecture, vol. 94, pp. 24-31, 2019.

[2] S. Namasudra, "Fast and Secure Data Accessing by using DNA Computing for the Cloud Environment," IEEE Transactions on Services Computing, pp. 1-1, 2020.

[3] S. Namasudra, D. Devi, S. Kadry, R. Sundarasekar, and A. Shanthini, "Towards DNA based data security in the cloud computing environment," Computer Communications, vol. 151, pp. 539-547, 2020/02/01/ 2020.

[4] A. Majumdar, A. Biswas, A. Majumder, S. K. Sood, and K. L. Baishnab, "A novel DNA-inspired encryption strategy for concealing cloud storage," Frontiers of Computer Science, vol. 15, p. 153807, 2020/12/31 2020.

[5] A. Khalifa, "LSBase: A key encapsulation scheme to improve hybrid crypto-systems using DNA steganography," in 2013 8th International Conference on Computer Engineering & Systems (ICCES), 2013, pp. 105-110.

[6] K. Ning, "A Pseudo DNA Cryptography Method," Computing Research Repository - CORR, 2009.

[7] B. Shimanovsky, J. Feng, and M. Potkonjak, "Hiding Data in DNA," in Information Hiding, Berlin, Heidelberg, 2003, pp. 373- 386.

[8] O. A. Al-Harbi, W. E. Alahmadi, and A. O. Aljahdali, "Security analysis of DNA based steganography techniques," SN Applied Sciences, vol. 2, p. 172, 2020/01/09 2020.

[9] S. Singh and Y. Sharma, "A Review on DNA based Cryptography for Data hiding," in 2019 International Conference on Intelligent Sustainable Systems (ICISS), 2019, pp. 282-285.

[10] P. Pavithran, S. Mathew, S. Namasudra, and P. Lorenz, "A novel cryptosystem based on DNA cryptography and randomly generated mealy machine," Computers & Security, vol. 104, p. 102160, 2021/05/01/ 2021.

[11] M. S. Rahman, I. Khalil, and X. Yi, "A lossless DNA data hiding approach for data authenticity in mobile cloud-based healthcare systems," International Journal of Information Management, vol. 45, pp. 276-288, 2019/04/01/ 2019.

[12] Richard M. Marzan and Ariel M. Sison, " An Enhanced Key Security of Playfair Cipher Algorithm," ICSCA '19: Proceedings of the 2019 8th International Conference on Software and Computer Applications.

[13] X. Wang and Q. Zhang, "DNA computing-based cryptography," in 2009 Fourth International on Conference on Bio-Inspired Computing, 2009, pp. 1-3.

[14] A. Elhadad, A. Khalifa, and S. Rida, "DNA-Based Data Encryption and Hiding Using Playfair and Insertion Techniques," Journal of Communications and Computer Engineering, vol. 2, p. 44: 49, 12/17 2011. DOI:10.20454/JCCE.2012.242

[15] UbaidurRahman, Noorul Hussain et al. "A Novel String Matrix Data Structure for DNA Encoding Algorithm." Procedia Computer Science 46 (2015): 820-832.

[16] A. Jose and K. Subramaniam, "DNA based SHA512-ECC cryptography and CM-CSA based steganography for data security," Materials Today: Proceedings, 2020/11/02/ 2020.

[17] M. Thangavel and P. Varalakshmi, "Enhanced DNA and ElGamal cryptosystem for secure data storage and retrieval in cloud," Cluster Computing, vol. 21, pp. 1411-1437, 2018/06/01 2018.

[18] E. Şatir and O. Kendirli, "A symmetric DNA encryption process with a biotechnical hardware," Journal of King Saud University - Science, vol. 34, p. 101838, 2022/04/01/ 2022.

[19] A. Elhadad, "Data sharing using proxy re-encryption based on DNA computing," Soft Computing, vol. 24, pp. 2101-2108, 2020/02/01 2020.

[20] F. Thabit, S. Alhomdy, and S. Jagtap, "A new data security algorithm for the cloud computing based on genetics techniques and logical-mathematical functions," International Journal of Intelligent Networks, vol. 2, pp. 18-33, 2021/01/01/ 2021.

[21] M. Tahir, M. Sardaraz, Z. Mehmood, and S. Muhammad, "CryptoGA: a cryptosystem based on genetic algorithm for cloud data security," Cluster Computing, vol. 24, pp. 739-752, 2021/06/01 2021.

[22] D. Prabhu and M. Adimoolam, "Bi-serial DNA Encryption Algorithm(BDEA)," Computing Research Repository - CORR, 01/13 2011.

[23] H. J. Shiu, K. L. Ng, J. F. Fang, R. C. T. Lee, and C. H. Huang, "Data hiding methods based upon DNA sequences," Information Sciences, vol. 180, pp. 2196-2208, 2010/06/01/ 2010.

[24] J. S. Taur, H.-Y. Lin, H.-L. Lee, and C. W. Tao, "Hiding In Dna Sequences Based On Table Lookup Substitution," 2012.

[25] B. Mondal, "A Secure Steganographic Scheme Based on Chaotic Map and DNA Computing," in Micro-Electronics and Telecommunication Engineering, Singapore, 2020, pp. 545-554.

[26] S. Marwan, A. Shawish, and K. Nagaty, "DNA-based cryptographic methods for data hiding in DNA media," Biosystems, vol. 150, pp. 110-118, 2016/12/01/ 2016.

[27] A. J. Menezes, S. A. Vanstone, and P. C. V. Oorschot, Handbook of Applied Cryptography: CRC Press, Inc., 1996.

[28] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, Digital Watermarking and Steganography: Morgan Kaufmann Publishers Inc., 2007.

[29] S. Katzenbeisser and F. A. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking: Artech House, Inc., 2000.

[30] S. K. Ghosal and J. K. Mandal, "On the use of the Stirling Transform in image steganography," Journal of Information Security and Applications, vol. 46, pp. 320-330, 2019/06/01/ 2019.

[31] J. Boyle, "Molecular biology of the cell, 5th edition by B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter," Biochemistry and Molecular Biology Education, vol. 36, pp. 317-318, 2008/07/01 2008.

[32] N. Kar, K. Mandal, and B. Bhattacharya, "Improved chaos-based video steganography using DNA alphabets," ICT Express, vol. 4, pp. 6-13, 2018/03/01/ 2018.

[33] Y. Wang, Q. Han, G. Cui, and J. Sun, "Hiding Messages Based on DNA Sequence and Recombinant DNA Technique," IEEE Transactions on Nanotechnology, vol. 18, pp. 299-307, 2019.

[34] J. Madison, I. Techreport, and S. Dickman, "An Overview of Steganography," 08/01 2007.

[35] J. Chen, L. Chen, and Y. Zhou, "Cryptanalysis of a DNA-based image encryption scheme," Information Sciences, vol. 520, pp. 130- 141, 2020/05/01/ 2020.

[36] M. Indrasena Reddy, A. P. Siva Kumar, and K. Subba Reddy, "A secured cryptographic system based on DNA and a hybrid key generation approach," Biosystems, vol. 197, p. 104207, 2020/11/01/ 2020.

[37] M. Sohal and S. Sharma, "BDNA-A DNA inspired symmetric key cryptographic technique to secure cloud computing," Journal of King Saud University - Computer and Information Sciences, vol. 34, pp. 1417-1425, 2022/01/01/ 2022.

[38] C. T. Clelland, V. Risca, and C. Bancroft, "Hiding messages in DNA microdots," Nature, vol. 399, pp. 533-534, 1999/06/01 1999.

[39] R. Bumgarner, "Overview of DNA microarrays: types, applications, and their future," Curr Protoc Mol Biol, vol. Chapter 22, p. Unit 22.1., Jan 2013.

[40] D. Na, "DNA steganography: hiding undetectable secret messages within the single nucleotide polymorphisms of a genome and detecting mutation-induced errors," Microbial Cell Factories, vol. 19, p. 128, 2020/06/11 2020.