

Steganographic Model for Encrypted Messages Based on DNA Encoding

Alfian Abdul Jalid*¹, Agus Harjoko², Anny Kartika Sari³

¹Master Program in Computer Science, FMIPA UGM, Yogyakarta, Indonesia

^{2,3}Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

e-mail: *alfian.abdul.j@mail.ugm.ac.id, aharjoko@ugm.ac.id, ³kartikasari@ugm.ac.id

Abstrak

Informasi sudah menjadi bagian yang tidak bisa dipisahkan dari kehidupan manusia. Beberapa informasi yang dinilai penting seperti dokumen negara atau perusahaan memerlukan pengamanan lebih untuk menjamin kerahasiaannya. Salah satu cara pengamanan informasi adalah dengan menyembunyikan informasi tersebut ke dalam suatu media tertentu dengan teknik steganografi. Steganografi adalah metode menyembunyikan informasi ke dalam file lain untuk membuatnya tidak terlihat. Salah satu metode steganografi yang sering digunakan adalah Least Significant Bit (LSB).

Pada penelitian ini akan dilakukan modifikasi metode LSB menggunakan DNA Encoding dan Chargaff's Rule. Chargaff's Rule atau complementary base pairing rule digunakan untuk menyusun complementary strand. Modifikasi metode LSB menggunakan DNA Encoding dan Chargaff's Rule diharapkan dapat meningkatkan keamanan dari informasi.

Hasil pengujian MSE menunjukkan nilai rata-rata metode LSB adalah sebesar 0.000236368, sedangkan nilai rata-rata untuk metode Steganografi berbasis DNA Encoding adalah 0.000770917. Nilai rata-rata PSNR untuk metode LSB adalah 76.82 dB sedangkan metode Steganografi berbasis DNA Encoding memiliki nilai rata-rata 70.88 dB. Waktu penyisipan dan ekstraksi pesan dengan metode Steganografi berbasis DNA Encoding relatif lebih lama dibandingkan dengan metode LSB karena kompleksitas algoritmenya yang lebih tinggi. Keamanan pesan dari metode Steganografi berbasis DNA Encoding lebih baik karena terdapat enkripsi dalam algoritmenya dibandingkan dengan metode LSB yang belum memiliki enkripsi.

Kata kunci— Steganografi, LSB, DNA Encoding, Chargaff's Rule

Abstract

Information has become an inseparable part of human life. Some information that is considered important, such as state or company documents, require more security to ensure its confidentiality. One way of securing information is by hiding the information in certain media using steganography techniques. Steganography is a method of hiding information into other files to make it invisible. One of the most frequently used steganographic methods is Least Significant Bit (LSB).

In this study, the LSB method will be modified using DNA Encoding and Chargaff's Rule. Chargaff's Rule or complementary base pairing rule is used to construct a complementary strand. The modification of the LSB method using DNA encoding and Chargaff's Rule is expected to increase the security of the information.

The MSE test results show the average value of the LSB method is 0.000236368, while the average value for the DNA Encoding-based Steganography method is 0.000770917. The average PSNR value for the LSB method was 76.82 dB while the DNA Encoding-based Steganography method had an average value of 70.88 dB. The time of inserting and extracting messages using the Steganography method based on DNA Encoding is relatively longer than the LSB method because of its higher algorithmic complexity. The message security of the DNA Encoding-based Steganography method is better because there is encryption in the algorithm compared to the LSB method which does not have encryption.

Keywords— Steganography, LSB, DNA Encoding, Chargaff's Rule

1. INTRODUCTION

The Information has become an inseparable part of human life. Almost all information has been stored in a data file format that can be stored on digital media such as computers, external storage, and others. Information stored in digital media has several advantages including easy storage, reduced paper usage, more resistance to damage, and others. However, information stored in digital media also has several drawbacks, including its originality, which is easy to change, easy to duplicate, and others. Some information that is considered important, such as state or company documents, require more security to ensure its confidentiality.

There are many ways to secure information stored in digital media. One way of securing information is by hiding the information in certain media using steganography techniques. Steganography is a method of hiding information in other files, such as image files, to hide the information's presence. The application of steganography will provide more security for information security, as well as a challenge for attackers in digital media storage. Attackers who want to know confidential data need to work harder to get that information.

Along with the development of DNA computation[1] emerged DNA cryptography. DNA cryptography is a relatively new technique for securing information in the field of cryptography, using DNA as an information carrier and computation with the help of molecular techniques. DNA cryptography combines computational complexity and biological complexity[2]. DNA cryptography is gaining attention because of its large DNA storage capacity, where one gram of DNA is known to be capable of storing about 108 terabytes of data. This ability to store large amounts of data makes DNA the best candidate for future media storage. The study of DNA can be applied to DNA cryptosystems based on DNA and one-time-pads, and if used correctly, the system is virtually impossible to penetrate. There are various procedures for one-time-pad DNA encryption schemes[2].

The most widely known method of steganography is the Least Significant Bit (LSB) method. This method modifies the smallest bit layer of an image. This technique takes advantage of the fact that the smallest bits in the image can be considered random noise and their alteration will have no effect on the image. Although the image did not appear to change visually after modification, the statistical properties of the image did change significantly. [3]explain that this method operates in the spatial domain of digital images. However, the application of this method usually raises suspicion because sometimes it can be detected by steganography detection applications and is very easy to extract.

Many modified LSB methods that aim to increase security and reduce noise that occurs in the information insertion process. Several LSBs are modified S Simple LSB Substitution, Fibonacci Decomposition LSB Substitution, Prime Number Decomposition LSB Substitution, and Natural Number Decomposition LSB Substitution. However, the algorithm of the modified LSB focuses on the insertion and pays less attention to the encryption aspects of the message. To encrypt the message to be inserted, the modified LSB uses another encryption algorithm outside the steganographic scheme.

This study aims to modify the LSB method using DNA encoding and Chargaff's Rule. Chargaff's Rule, also known as the complementary base pairing rule, states that the DNA base pairs are always adenine with thymine (A-T) and cytosine with guanine (C-G). Chargaff's Rule is used to construct complementary strands. With this complementary strand, it is expected to increase the security of the message. The LSB method will be modified by utilizing DNA Encoding and Chargaff's Rule, the modification will use 2 LSB bits because it is adjusted to the characteristics of DNA Encoding which represents nucleotide bases in 2 bits.

2. METHODS

2.1 DNA Encoding

Deoxyribonucleic Acid (DNA) is an entity that stores information from all types of living things. There are four nucleic acids, namely A (Adenine), C (Cytosine), G (Guanine), and T (Thymine) which are used in the DNA sequence. In the DNA sequence, A is the complement of T and C is the complement of G[4]. These four nucleic acids can be represented in binary numbers, as we know, in the binary system, 0 and 1 complement each other. Therefore, it can be concluded that 00 and 11 are complementary and also 01 and 10 are complementary[5], Table 1 shows the encoding and decoding maps for DNA.

Table 1 DNA Encoding and Decoding Rules

NA	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8
A	00	00	11	11	10	01	10	01
T	11	11	00	00	01	10	01	10
C	10	01	10	01	00	00	11	11
G	01	10	01	10	11	11	00	00

For example, if there are pixels with the value 157, the binary format is (10011101) 2. The DNA codes for all 8 rules in Table 1 are as follows: Rule 1 (CGTG), Rule 2 (GCTC), Rule 3 (CGAG), Rule 4 (GCAC), Rule 5 (ATGT), Rule 6 (TAGA), Rule 7 (ATCT) and Rule 8 (TACA).

In addition to increasing the security of DNA cryptography, [6] proposed several algebraic operations in the form of XOR, addition, and subtraction between nucleic acids, as shown in Table 2, Table 3, and Table 4. Addition and subtraction for DNA were carried out based on a system of addition and subtraction in $Z_2 \pmod{2}$ [7]. For example, $11 + 10 = 01$, $01 - 11 = 10$.

The encryption process using DNA cryptography is done by converting the value from plain text to binary form. Then do the coding based on one of the rules in Table 1. Next, encrypt the coding results with the key using the XOR operation and summation in Table 2 and Table 3. The decryption process uses the XOR operation and subtraction in Table 2 and Table 4.

Table 2 DNA XOR operations

XOR	A	G	C	T
A	A	G	C	T
G	G	A	T	C
C	C	T	A	G
T	T	C	G	A

Table 3 DNA Addition Operation

+	A	G	C	T
A	C	T	A	G
G	T	C	G	A
C	A	G	C	T
T	G	A	T	C

Table 4 DNA Subtraction Operation

-	A	G	C	T
A	C	T	A	G
G	T	C	G	A
C	A	G	C	T
T	G	A	T	C

2.2 Chargaff's Rule

Chargaff's Rule, also known as the complementary base pairing rule, states that DNA base pairs are always adenine with thymine (A-T) and cytosine with guanine (C-G)[8]. Purines always pair with pyrimidines and vice versa. However, A does not pair with C, even though it is purines and pyrimidines. This rule is named after scientist Erwin Chargaff who found that there are basically the same concentrations of adenine and thymine, as well as guanine and cytosine in almost all DNA molecules [9]. These ratios can vary among organisms, but the actual concentration of A is always the same as T and the same as G and C. For example, in humans, there is about 30.9% adenine, 29.4% thymine, 19.8% cytosine, and 19.9% guanine. This supports the complementary rule that A must match T and C must match G [10].

This corresponds to hydrogen bonds joining complementary DNA strands along with the space available between the two strands. There are approximately 20 Å (angstrom, 1 angstrom equal to 10⁻¹⁰ meters) between the two complementary DNA strands. Two purines and two pyrimidines together will only take up too much space to fit into the space between the two strands. This is why A can't bind G and C can't bind to T.

The bonds between purines and pyrimidines are not interchangeable due to the hydrogen bonds that connect the bases and stabilize the DNA molecule. The only pairs that can make hydrogen bonds in that space are adenine with thymine and cytosine with guanine. A and T form two hydrogen bonds while C and G form three bonds. It is these hydrogen bonds that join the two strands and stabilize the molecule, allowing it to form a ladder-like double helix.

By using Chargaff's Rule, you can arrange complementary strands based solely on the order of the base pairs. For example, let's say you know the sequence of one DNA strand is as follows:

AAGCTGGTTTTGACGAC

Using Chargaff's Rule, a complementary strand is obtained:

TTCGACCAAACTGCTG

2.3 Proposed Methods

This study develops an algorithm that combines DNA-based cryptography and steganography using Chargaff's Rule. In general, this research scheme is divided into an insertion scheme and an extraction scheme. In the insertion process, Message M is inserted into the Cover Image, the insertion process will produce a Stego Image. In the extraction process, Message M is extracted from the Stego Image.

2.3.1 Insertion Scheme

The insertion process is carried out by converting the binary message or M-BIN into DNA using DNA encoding based on Rule 2 in Table 1. The DNA message is then processed by Chargaff's Rule into MCHR and the length of the DNA message is used as MLENGTH. At the same time, the RGB value of the Cover Image pixels is taken. Then the value is divided into layer R, layer G, layer B. Then at the end of layer G, the value from MLENGTH is inserted. In this study, the last 24 LSBs on layer G were allocated to store the MLENGTH value. The binary value of each layer is converted into DNA using DNA encoding based on Rule 2 in Table 1.

The DNA G layer is then XORed with a Secret Key based on Table 2 to produce a G DNA Key in the form of DNA that is entered by the user. Figure 2 shows the flowchart of the Insertion Scheme.

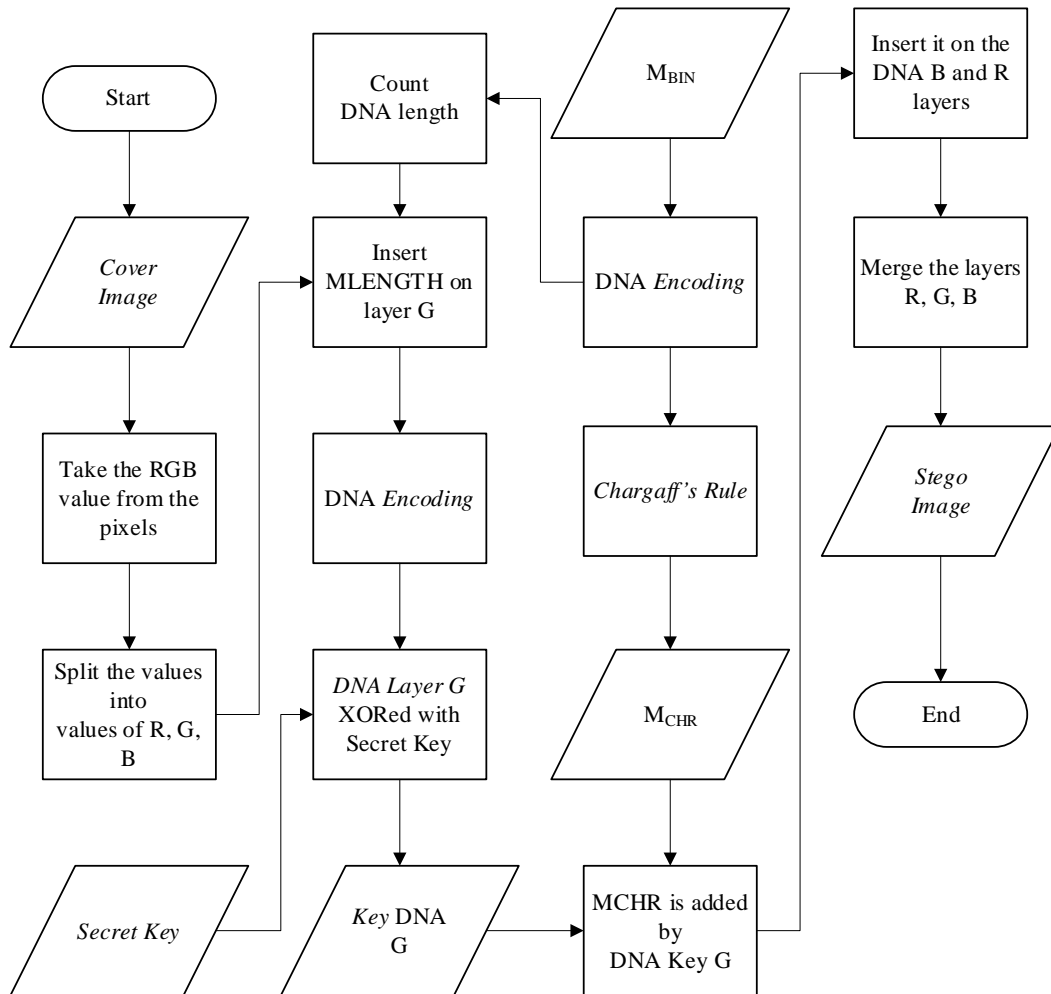


Figure 2 Insertion Scheme

Then the DNA addition operation was performed on the DNA Chargaff's Rule or MCHR message with the DNA G Key according to Table 3. generate MKEYG messages. According to the characteristics of human vision [11], the sensitivity of the three components of a different color image is most sensitive to green, followed by red, which is least sensitive to blue. Therefore, MKEYG messages will be inserted at layer B then on layer R. Meanwhile, at layer G there is no message insertion, layer G is only used to store MLENGTH. After the message is inserted into layers B and R, the Stego layers R and B are generated. From the G and Stego layers, the image is reconstructed to produce a Stego Image.

2.3.2 Extraction Scheme

The extraction process is carried out by taking the RGB value of the Stego Image pixels. Then the RGB value is divided into layer R, layer G, layer B. MLENGTH is obtained from the end of layer G. The binary values of each layer are then converted into DNA using DNA Encoding.

Then take the LSB value from layer B, then layer R along with the value from MLENGTH produces MKEYG. The DNA G layer is then XORed with a Secret Key based on

Table 2 to produce a G DNA Key. Then the MKEYG reduction operation is carried out with the G DNA Key according to Table 4 to produce a DNA message. The DNA obtained from the reduction operation is then carried out by the Chargaff's Rule process and converted into binary by the DNA Decoding process. Figure 3 Shows the Extraction Scheme.

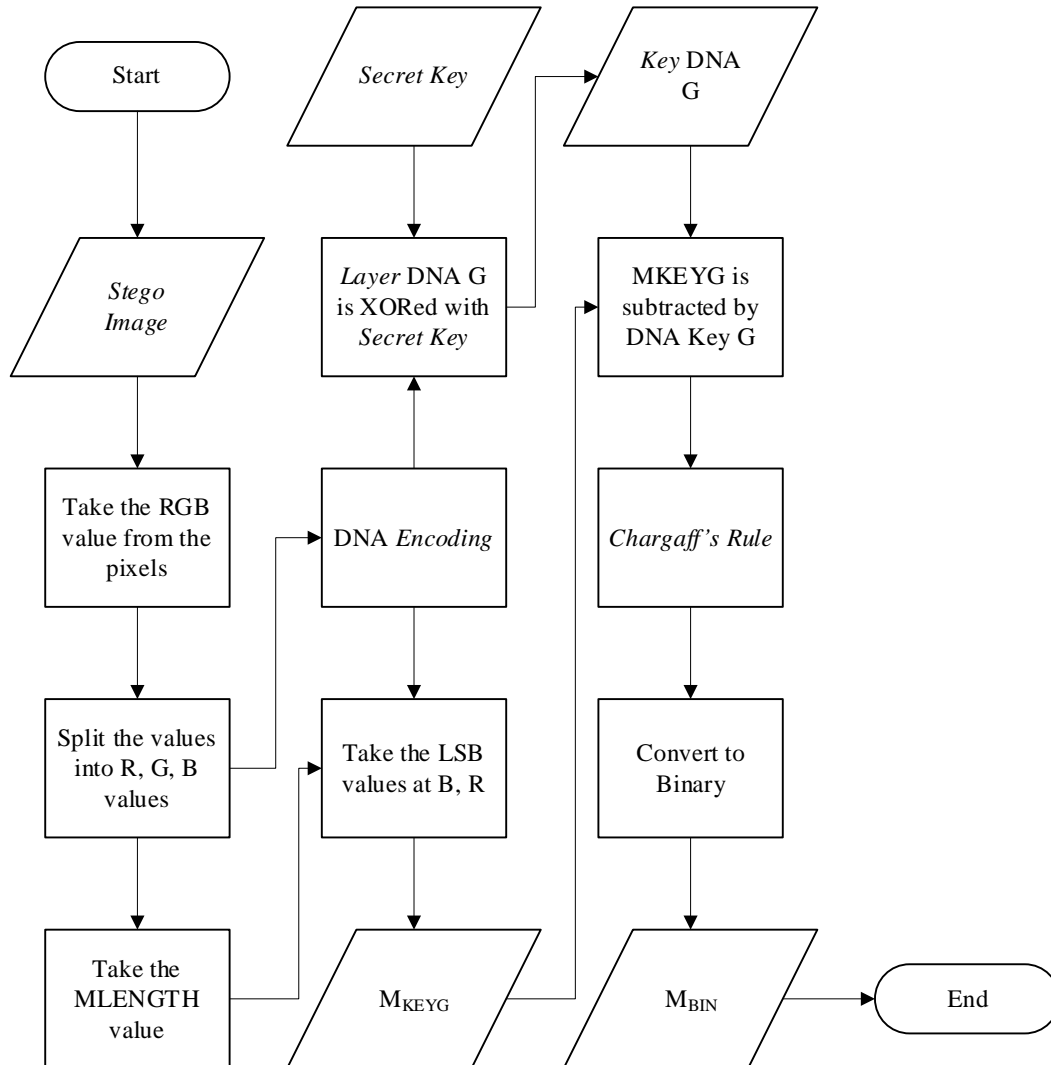


Figure 3 Extraction Scheme

3. RESULTS AND DISCUSSION

Image quality testing is done using MSE calculations and PSNR calculations. The greater the MSE value, the greater the difference between the cover image and the stego image. In contrast to the MSE calculation, in the PSNR calculation, the difference between the cover image and the stego image is greater if the resulting value is getting smaller. PSNR values are usually expressed on a decibel (dB) scale. PSNR values below 30 dB indicate low image quality, while values above 40 dB indicate good image quality. The higher the PSNR value produced, the better the image quality [12]. The results of MSE and PSNR calculations for the DNA Encoding-based Steganography method and the LSB method can be seen in Table 5.

Table 5 MSE And PSNR Results

No	Image Name	MSE		PSNR	
		LSB	DNA Encoding-based Steganography	LSB	DNA Encoding-based Steganography
1	Fence.bmp	0.00045185	0.0014093	73.3435	67.5872
2	BannerUKM.bmp	0.00015444	0.00044416	77.446	71.4648
3	KuliahUmum.bmp	0.00010995	0.00035417	79.4655	73.5951
4	Vredeburg.bmp	0.00041111	0.0014593	73.5697	67.4325
5	NolKM.bmp	0.00017216	0.00054615	77.5685	71.6942
6	Pool.bmp	0.00011435	0.00035625	79.3521	73.4871
7	Lounge.bmp	0.00042685	0.0014481	73.6592	67.5081
8	Bathroom.bmp	0.00017433	0.00055809	77.5355	71.5863
9	Kitchen.bmp	0.00011227	0.00036273	79.4984	73.5686

In Table 5, it can be seen that the comparison of MSE and PSNR values from the DNA Encoding-based Steganography method and the LSB method does not have a significant difference, so it can be said that the method produces the same good image quality.

The message insertion and extraction time test was carried out to determine whether the DNA Encoding-based Steganography method modified by the LSB method still had the same speed or not. LSB is a steganographic method that has a relatively fast message insertion and extraction speed. A comparison of the results of message insertion and extraction time between DNA Encoding-based Steganography method and the LSB method can be seen in Table 6.

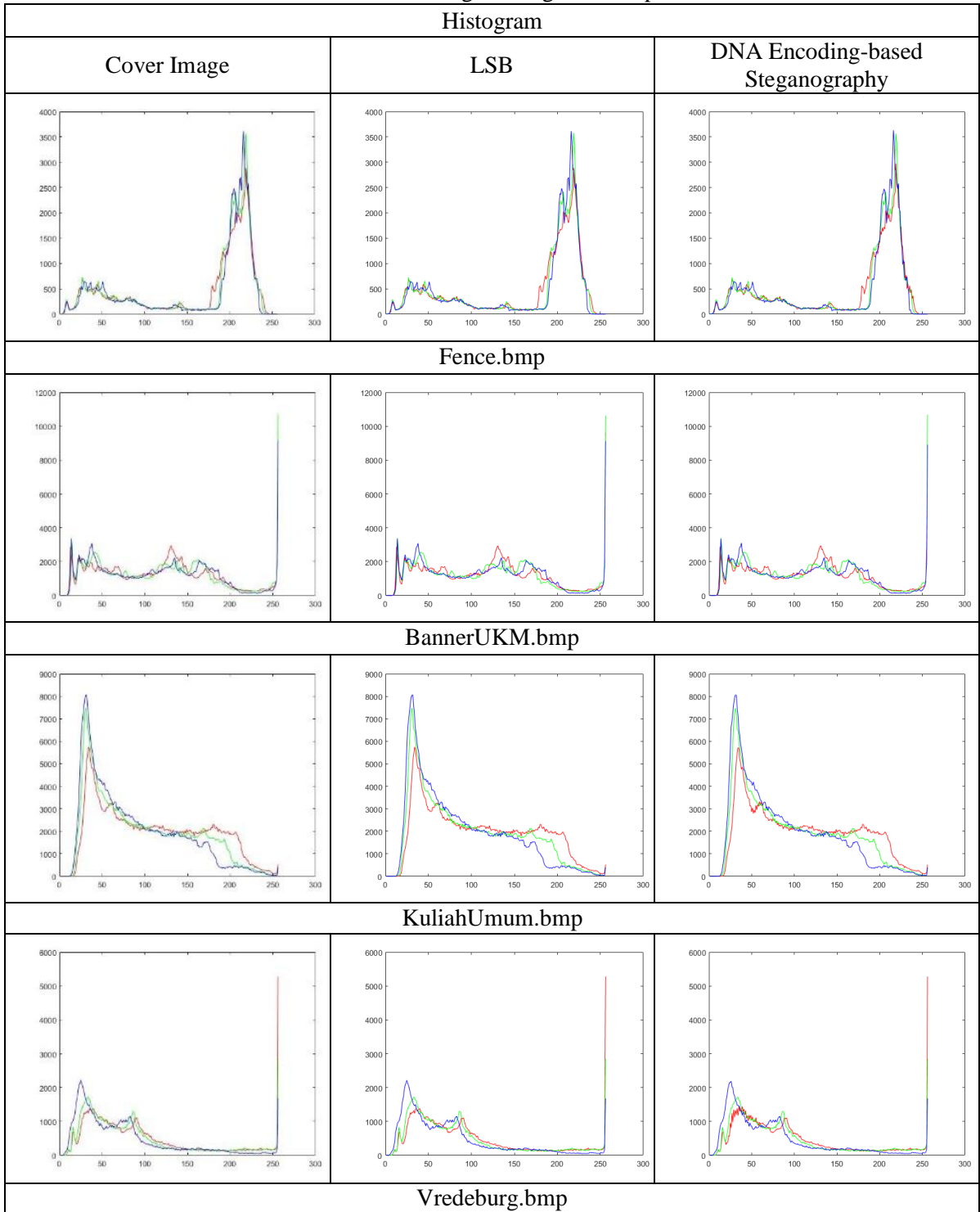
Table 6 Comparison of insertion and extraction times

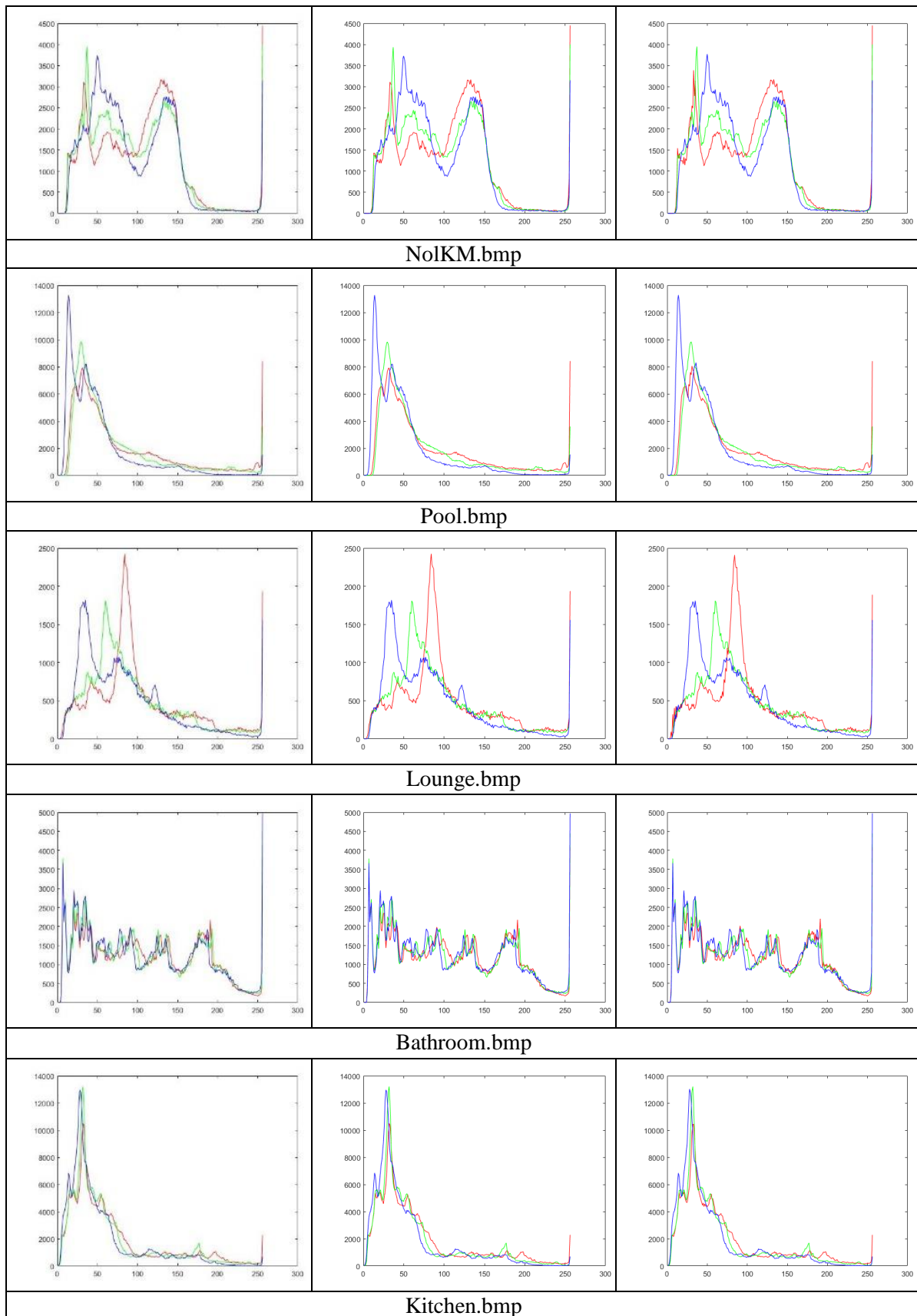
No	Image Name	Insertion		Extraction	
		LSB (s)	DNA Encoding-based Steganography (s)	LSB (s)	DNA Encoding-based Steganography (s)
1	Fence.bmp	0.15649	0.64917	0.10135	0.73068
2	BannerUKM.bmp	0.12047	0.66188	0.10738	0.89006
3	KuliahUmum.bmp	0.13666	0.66406	0.11155	0.70978
4	Vredeburg.bmp	0.12716	0.63781	0.09956	0.70797
5	NolKM.bmp	0.11204	0.64059	0.09749	0.70576
6	Pool.bmp	0.12257	0.77381	0.10051	0.72574
7	Lounge.bmp	0.11018	0.73734	0.10012	0.73275
8	Bathroom.bmp	0.1120	0.73882	0.10546	0.68867
9	Kitchen.bmp	0.12383	0.75297	0.10172	0.71107

In Table 6, it can be seen that the average time required to insert and extract messages using DNA Encoding-based Steganography method is more than LSB method. This is because DNA Encoding-based Steganography method has a higher complexity than LSB method.

The histogram test aims to compare the histogram of the cover image and the stego image from the DNA Encoding-based Steganography method and the LSB method. The histogram shows the appearance frequency of each pixel value. The histogram comparison of the cover image and stego image of DNA Encoding-based Steganography method and the LSB method is shown in Table 7.

Table 7 Image Histogram Comparison





In general, the inserted image using the two techniques tested gave the same good results visually. When the message extraction process is carried out according to the message inserted. To test more accurately, several quantitative test parameter values have been carried out. The parameters taken into account in this test are the value of MSE, PSNR, and processing time.

The MSE value is a parameter that measures the error between the original image and the embedded image. The MSE value is large enough to indicate a decrease in quality or there has been a significant change in the inserted image. Table 5 shows the results of MSE calculations for each of the techniques tested. The average MSE value for the Cha DNA Encoding-based Steganography method is 0.000770917, while the average MSE value for the LSB method is 0.000236368. If viewed from the MSE value, it can be seen that the LSB method provides slightly better insertion results than DNA Encoding-based Steganography method.

Peak Signal to Noise Ratio (PSNR) is a comparison used to compare the value of the cover image with the stego image that has been inserted with a message. The higher the PSNR value, the better the similarity level between the cover image and the manipulated image. To calculate the PSNR value, you must first calculate the Mean Square Error (MSE) value of the two images.

Based on Table 5, it can be seen that the average PSNR value in the LSB method is 76.82 dB while the DNA Encoding-based Steganography method is 70.88 dB. This shows that the LSB method tends to be better than DNA Encoding-based Steganography method. This can be seen from the difference in the average PSNR value of 5.94 dB. The PSNR value distribution for the LSB method is at a minimum of 73.34 dB and a maximum of 79.49 dB. As for the DNA Encoding-based Steganography method, it is between 67.43 dB to 73.59 dB.

Processing time testing aims to review the time required to perform the insertion and extraction in each method. Based on Table 6, it is known that the average insertion processing time for DNA Encoding-based Steganography method is 0.6951 seconds, while the extraction process is 0.7336 seconds. Then for the LSB method, an average insertion processing time was obtained for 0.1246 seconds and extraction for 0.1027 seconds. DNA Encoding-based Steganography method requires a longer processing time because it has a higher algorithmic complexity than the LSB method.

Based on the results and analysis that has been done, it is found that the LSB method has the advantage of a faster processing time than DNA Encoding-based Steganography method. DNA Encoding-based Steganography method has a higher level of complexity than the LSB method. DNA Encoding-based Steganography method has better message security because it has encryption compared to the LSB method which doesn't have encryption. The MSE and PSNR values of the DNA Encoding-based Steganography method and the LSB method do not have a significant difference, so it can be said that the two methods produce the same good image quality. Table 8 shows the advantages of each method in each parameter.

Table 8 Methods Comparison

No	Parameter	Methods	
		LSB	DNA Encoding-based Steganography
1	MSE	✓	✓
2	PSNR	✓	✓
3	Processing Time	✓	
4	Complexity		✓
5	Message Security		✓

4. CONCLUSIONS

DNA Encoding-based Steganography method can insert and extract messages properly as long as the inserted message does not exceed the capacity of the cover image used. The calculation of MSE and PSNR from the stego image using the DNA Encoding-based Steganography method produces a value that is not much different from the stego image using the LSB method. The message security in DNA Encoding-based Steganography method is higher than the LSB method because there is encryption in the insertion so that the message that is inserted cannot be immediately guessed. DNA Encoding-based Steganography method developed in this study is applied to insert text into a bitmap image. In further research, insertion with messages and other media can be carried out. DNA Encoding-based Steganography method requires a relatively longer time to insert and extract messages compared to the LSB method. Future research using parallel computing is expected to solve this problem. The resistance of the stego image generated by the DNA Encoding-based Steganography method is relatively low because the message cannot be extracted if an image is manipulated on the resulting stego image. In further research, it can be done to increase the resistance of the resulting stego image to image manipulation.

REFERENCES

- [1] L. M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems," *Am. Assoc. fir Adv. Sci.*, vol. 266, no. 5187, pp. 1021–1024, 1994 [Online]. Available: <http://www.jstor.org/stable/2885489>
- [2] B. Anam, K. Sakib, M. A. Hossain, and K. Dahal, "Review on the Advancements of DNA Cryptography," 2010 [Online]. Available: <http://arxiv.org/abs/1010.0186>
- [3] R. K. Arya and R. Saharan, "Algorithm to Enhance the Robustness and Imperceptibility of LSB," *Proc. - 2015 2nd IEEE Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2015*, pp. 583–587, 2015, doi: 10.1109/ICACCE.2015.19.
- [4] Q. Zhang, L. Guo, and X. Wei, "A novel image fusion encryption algorithm based on DNA sequence operation and hyper-chaotic system," *Optik (Stuttg.)*, vol. 124, no. 18, pp. 3596–3600, 2013, doi: 10.1016/j.ijleo.2012.11.018. [Online]. Available: <http://dx.doi.org/10.1016/j.ijleo.2012.11.018>
- [5] R. Enayatifar, A. H. Abdullah, and I. F. Isnin, "Chaos-based image encryption using a hybrid genetic algorithm and a DNA sequence," *Opt. Lasers Eng.*, vol. 56, pp. 83–93, 2014, doi: 10.1016/j.optlaseng.2013.12.003. [Online]. Available: <http://dx.doi.org/10.1016/j.optlaseng.2013.12.003>
- [6] Q. Zhang, L. Guo, and X. Wei, "Image encryption using DNA addition combining with chaotic maps," *Math. Comput. Model.*, vol. 52, no. 11–12, pp. 2028–2035, 2010, doi: 10.1016/j.mcm.2010.06.005. [Online]. Available: <http://dx.doi.org/10.1016/j.mcm.2010.06.005>
- [7] P. Wasiewicz, J. J. Mulawka, W. R. Rudnicki, and B. Lesyng, "Adding numbers with DNA," *Proc. IEEE Int. Conf. Syst. Man Cybern.*, vol. 1, pp. 265–270, 2000, doi: 10.1109/ICSMC.2000.885000.
- [8] K. Menaka, "Message encryption using DNA sequences," *Proc. - 2014 World Congr. Comput. Commun. Technol. WCCCT 2014*, pp. 182–184, 2014, doi: 10.1109/WCCCT.2014.35.
- [9] S. Jain, N. Raviv, and J. Bruck, "Attaining the 2nd Chargaff Rule by Tandem Duplications," *IEEE Int. Symp. Inf. Theory - Proc.*, vol. 2018–June, pp. 2241–2245, 2018, doi: 10.1109/ISIT.2018.8437526.
- [10] E. Walsh, "What Is the Complementary Base Pairing Rule?," *sciencing.com*, 2019.

- [Online]. Available: <https://sciencing.com/complementary-base-pairing-rule-8728565.html>. [Accessed: 29-Jun-2019]
- [11] X. Zhou, W. Gong, W. Fu, and L. Jin, "An Improved Method for LSB Based Color Image steganography Combined with Cryptography," *2016 IEEE/ACIS 15th Int. Conf. Comput. Inf. Sci.*, pp. 1–4, 2016, doi: 10.1109/ICIS.2016.7550955.
- [12] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "Digital image steganography: Survey and analysis of current methods," *Signal Processing*, vol. 90, no. 3, pp. 727–752, 2010, doi: 10.1016/j.sigpro.2009.08.010. [Online]. Available: <http://dx.doi.org/10.1016/j.sigpro.2009.08.010>