Research Article

SOHAG JOURNAL OF SCIENCES

https://doi.org/10.21608/sjsci.2022.165686.1036

Comparison Study Between Simple LSB and Optimal LSB Image Steganography

Marghny H. Mohamed[1], Mahmoud A. Mofaddel[2] and Tarek Y. Abd El-Naser[3,*]

[1] *Computer Science Department, Faculty of Computers and Information, Assuit University, Assuit 71516, Egypt.*
[2] *Computer Science Department, Faculty of Computers and Artificial Intelligence, Sohag University, Sohag 82524, Egypt*
[3] *Computer Science Department, New Cairo Academy, Cairo 11835, Egypt*
**E-mail: tarekyahia2000@gmail.com*

Received: 27th September 2022, **Revised:** 21st November 2022, **Accepted:** 24th November 2022.
Published online: 1st January 2023

Abstract: There is no denying communication's role in the evolution of the internet and other digital technologies. Information comes in many forms, and its conveyance necessitates a wide variety of methods of communication. We must communicate these types of data over a secure communication channel on the internet because they include confidential and sensitive information. One of the most challenging issues was figuring out how to send sensitive information without raising suspicion. Cryptography and steganography are just two examples of many methods used to conceal information from prying eyes, all to protect sensitive data from being misused or altered by unauthorized parties. Using an encryption method, cryptography transforms information from its original, easily decipherable form (plain text) into a completely meaningless one (cipher text). We can also use a decryption technique to change the encrypted text back into plain text. Steganography is a technique used to transmit information through other forms of digital media secretly. Due to the difficulty in visually detecting hidden data, the image will serve as the cover media for this thesis. The well-known method used to conceal confidential information is LSB (Least Significant Bit). The LSB method is based on exchanging bits of the cover medium for bits of private data.

Keywords: Steganography, LSB Method, Optimal LSB, Cryptography, Information Hiding.

1. Introduction

Steganography is used to conceal sensitive information such as text, audio, and video in other digital cover media such as text, audio, and video to facilitate confidential communication. Steganography is used primarily for secret communication. Communication is considered secret if the secret message is concealed during transmission. "Steganography" comes from the Greek word for "covered writing." Simmons' Prisoner's Problem is a famous Steganography story. Two prisoners plot an escape; they must discuss the plan without drawing the 'guards attention. They communicated using a specific method that carried the hidden message. Steganography has a long and ancient history as an idea and practice [1,2].

Ancient Greece used steganography techniques to send secret messages by shaving a slave's head and sending him as a carrier when his hair grew back. Letters were written on silk and dipped in wax in ancient China. Many cultures use steganography to hide messages using secret (invisible) inks [3]. Steganography's main objective is to hide information in other digital cover media so that only the intended sender and recipient (authorized users) know its existence. This prevents other people (unauthorized users) from noticing its presence.

2. Overview of Simple LSB Technique

The Least Significant Bit Substitution technique (LSB) is recognized as one of the pioneering methods of steganography and one of the most widely used methods today. The least significant bit (LSB) is the tiniest bit in the bit sequence, as defined by the field of computer science.

The LSB substitution technique is referred to as "hiding the secret data bit into the LSB of the [4] cover binary sequence" and also is defined as "hiding the secret data bit into the LSB of the cover binary sequence by replacing the LSB value of the cover binary sequence with the secret bit value." This is true regardless of the order of the cover binary sequences used to contain the confidential data and whether the hiding order is sequential or random. The Hide and Seek algorithm is the most straightforward approach for hiding data through LSB substitution. This algorithm embeds hidden bits into the LSBs of the cover binary sequences in a sequential pattern starting at the beginning of the cover medium and working its way through. When we want to embed a byte of confidential data, we take the eight bits of the secret byte and replace the tiniest bit of a series of eight binary sequences of the cover data with these secret bits, and so on. This process is repeated until the byte of secret data has been successfully embedded [5]. For example, if we need to conceal the secret data, we have a cover binary sequence 01101100 and a secret bit with a value of 1. We want to use the LSB substitution technique [6-8], then all we have to do is replace the LSB of the cover binary sequence with the secret bit, and the cover binary sequence becomes 01101101.

2.1 Structure of LSB Technique

The least significant bit, or LSB, is the bit on the rightmost byte. One advantage of employing LSB for hiding data is that changing this bit will not significantly impact the value of the other seven bits: table 1 and table 2 show the process of changing the last bit.

Table 1. The Following 7 Bits Gives Value (3).

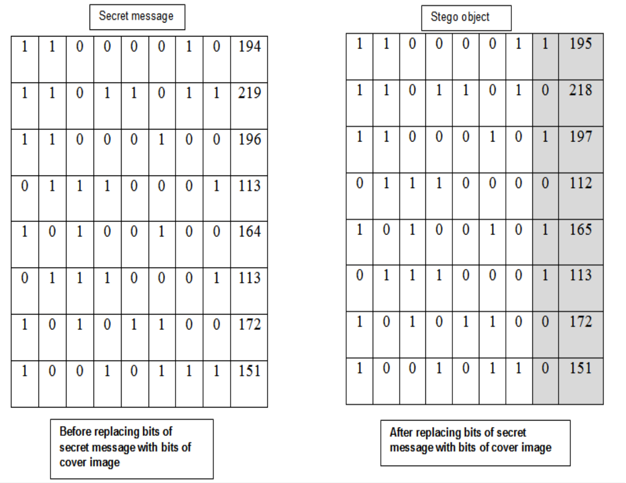
0	**0**	**0**	**0**	**0**	**0**	**1**	**1**

Table 2. After Changing the Last Bit, the Value is (2).

0	**0**	**0**	**0**	**0**	**0**	**1**	**0**

So LSB technique depends on replacing the last bits of the pixels of the cover image with the ones in the secret message without a noticeable change occurring in the cover image [9-11].

Table 3. The Values of Pixels After Being Replaced Using the LSB Technique.



Secret message

1	1	0	0	0	0	1	0	194
1	1	0	1	1	0	1	1	219
1	1	0	0	0	1	0	0	196
0	1	1	1	0	0	0	1	113
1	0	1	0	0	1	0	0	164
0	1	1	1	0	0	0	1	113
1	0	1	0	1	1	0	0	172
1	0	0	1	0	1	1	1	151

Before replacing bits of secret message with bits of cover image

Stego object

1	1	0	0	0	0	1	1	195
1	1	0	1	1	0	1	0	218
1	1	0	0	0	1	0	1	197
0	1	1	1	0	0	0	0	112
1	0	1	0	0	1	0	1	165
0	1	1	1	0	0	0	1	113
1	0	1	0	1	1	0	0	172
1	0	0	1	0	1	1	0	151

After replacing bits of secret message with bits of cover image

From previous results in table 3, we found that changing the value of just one pixel does not make a noticeable difference, and this helps to obtain a less distorted image that can be used safely and is difficult to detect by attackers**.**

3. Improvements of LSB Techniques

3.1 LSB Matching

Mielikainen's LSB matching method, proposed in 2006, improved on the LSB replacement method proposed by Chen et al. in 2004. When the amount of confidential information increases, the image quality of the Chen et alscheme.'s degrades. To improve the image quality of the Chen et alscheme.'s Mielikainen proposed the LSB matching method [11-13].

Their scheme divides a cover image into several 1 X 2 non-repetitive blocks; A and B are represented as two pixels for each block. m_1 and m_2 are the confidential information to be hidden in each group. To determine the appropriate modification rule, the scheme uses the tree flow diagram shown in Figure 1.

To determine the final modified rule, the LSB of A is compared to m1, and then the value of the F function is compared to m2.

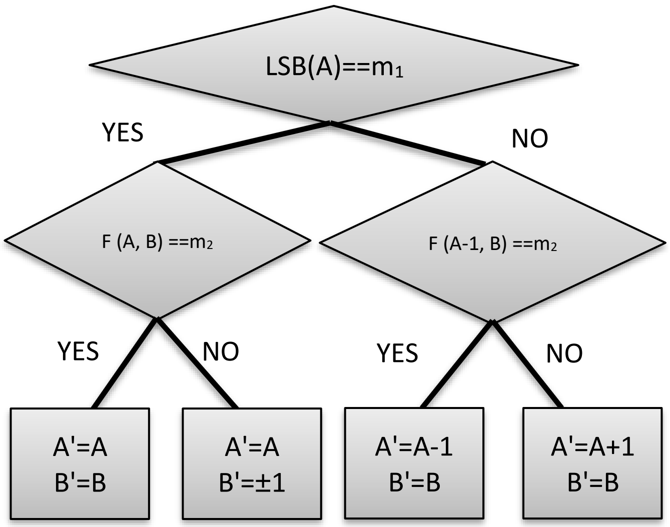


Fig. 1. LSB Matching Flowchart.

The F function is depicted as follows:

$$F\ (A,\ B) = LSB\left(\left\lfloor \frac{A}{2} \right\rfloor + B\right). \qquad (1)$$

The following are the four conditions of the final modify rule:

Case 1: When $LSB(A)=m_1$ and $F(A, B)=m_2$, the pixel pair A and B does not need modification.

Case 2: When $LSB(A)=m_1$ and $F(A, B)\neq m_2$, the pixel A does not change, and B=B+1 or B=B−1.

Case 3: When $LSB(A)\neq m_1$ and $F(A-1, B)=m_2$, the pixel A=A−1, and B does not change.

Case 4: When $LSB(A)\neq m_1$ and $F(A-1, B)\neq m_2$, the pixel A=A+1, and B does not change.

A camouflage image can be obtained by hiding each group of pixel pairs and confidential information in sequence. The confidential information extraction process can begin once the receiver receives the camouflage image. The following is the extraction formula:

$$m_1 = LSB(A'), \qquad (2)$$

$$m_2 = LSB\left(\left\lfloor \frac{A'}{2} \right\rfloor + B'\right). \qquad (3)$$

Finally, the processes of hiding and extracting are completed.

Assume that the pixel pairs are $A=120$, $B=121$, and the secret message is $s=(01)_2$. First, check if LSB (120) is the same as the secret message $m_1=0$. It is noticeable here that LSB (120) equals 0. Therefore, the scheme uses the LSB $(120, 121)$ function to determine whether $LSB = LSB\left(\left\lfloor \frac{120}{2} \right\rfloor + 121\right)$ is the same as $m_2=1$.

According to the hiding flowchart, the result is that

$LSB\left(\left\lfloor\frac{120}{2}\right\rfloor + 121\right)$ = *1*. Hence, the final modification rule is Case 1, where the pixel pair A and B do not need modification. Finally, these results are *A'=20* and *B'=121*.

3.2 GEP Algorithm

The goal of developing a genetic algorithm is to create a group of individuals (solutions) chosen from a specific given population using an evaluation method in which they are assigned points based on their fitness. Then there are the two best individuals who can be used to produce the best offspring. This new offspring may require some mutations based on application's needs, and the results will be evaluated. The process is then repeated until an acceptable solution is found or several of generations have passed. The genetic algorithm has numerous advantages because it is more potent in searching for complex solutions. They are also susceptible to collapse due to minor changes or noise. Other optimization methods are discussed, such as heuristics, praxis, linear programming, first or breadth-first. However, when searching for large multi-modal state spaces, the genetic algorithm produces the best and most important results. Many fields use genetic algorithms, including robotics, automotive design, optimized telecommunications routing, engineering design, and computer-aided molecular design [14].

(GEP) is one of the improved algorithms; it revolutionized solving complex problems with many solutions. It relies on a set of parameters like (Initial population, Selection (Roulette-Wheel), Inversion operator, Reproduction, Fitness Function, Elitism selection). Figure 2 shows the GEP algorithm charts. Figure 3 shows the flowchart of GEP algorithm. Gene expression programming (GEP), in its method of operation, is a combination of the workings of genetic algorithms (GAs) and genetic programming (GP).

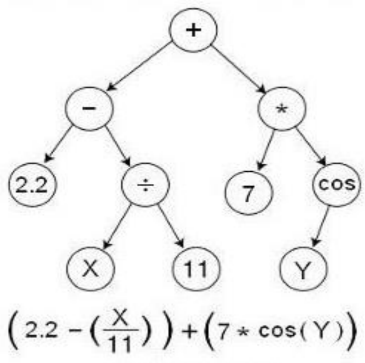


Fig. 2. GEP Algorithm Charts.

A genetic algorithm conducts an experiment on a particular group of individuals, chooses the fit individuals based on an evaluation function that considers their fitness level, and then applies a genetic operator to those chosen individuals. The primary distinction between the three algorithms is based on the characteristics of each individual. With GAs, the individuals are arranged into strings of a fixed length (chromosomes). With GP, the individuals are arranged into nonlinear objects of different sizes and shapes (parse trees).

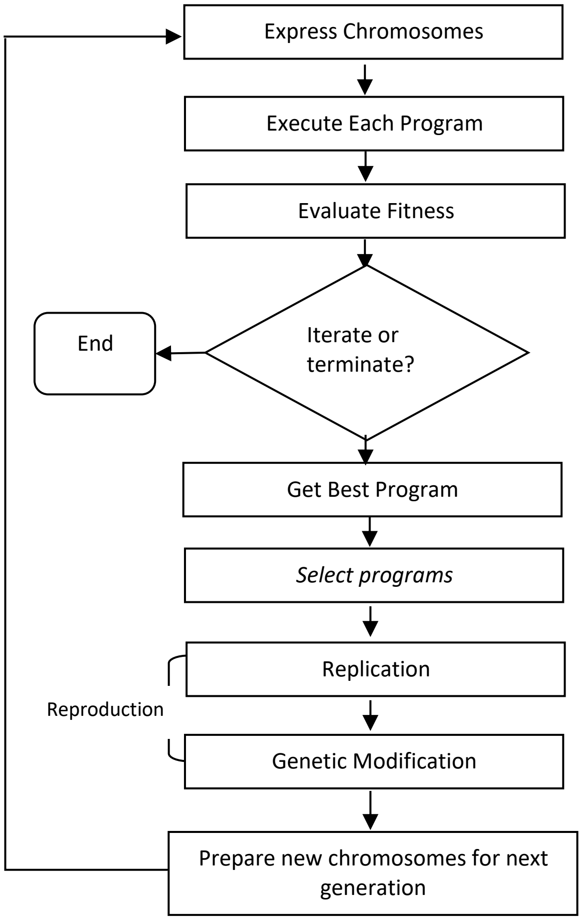
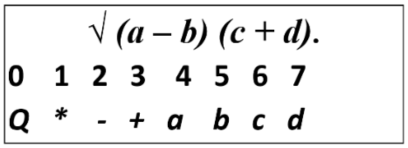


Fig. 3. The Flowchart of the GEP Algorithm

With GEP, the individuals are arranged into strings of a fixed length that are expressed as nonlinear objects with different sizes and shapes [14]. This type of diagram representation is called the phenotype in GEP. "Q" represents the square root function in this representation [15]. From the phenotype, it is easy to figure out the following about the genotype by applying it to the following mathematical expression, as shown in Figure 4.



√ (a – b) (c + d).

0	1	2	3	4	5	6	7
Q	*	-	+	a	b	c	d

Fig. 4. Phenotype and Genotype Structure [16].

4. Fitness Function

In this work, the fitness function is defined as the mean square error (MSE); which calculates differences between the original cover image and the stego-image [17]. The fitness functions are used to measure the quality of the current solution. The fitness function aims to achieve the highest capacity of stego-image with the least amount of distortion possible. The measurement of high capacity and minimum distortion can be evaluated by maximum PSNR, which means minimum MSE. As a result, our objective is to choose a solution with values that are as high as possible for PSNR. Estimates of the PSNR are given in decibels (dB), which are defined as follows:

$$PSNR = 10*log10\frac{255*255}{MSE}, \quad (4)$$

and MSE is defined as:

$$MSE=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-y_{ij}). \quad (5)$$

Where, x_{ij} refers to the original pixel value, and y_{ij} refers to the processed pixel value, and m and n denote the width and height of the image respectively [15].

5. Implement and Evaluation

In the following experiment, two covers are used, "Babbon" and "Lena," with a resolution of 512X512 pixels. One secret image is used, "Tiffany," with a resolution of 128X128 pixels. The parameters of GEP are as follows:

- Maximum of generation = 50,
- Population size = 100,
- Inversion rate = 0.3,
- No. of MGFs=1

The used cover images were Babbon with 512 X 512 pixels and Lena with a resolution of 512 X 512 pixels. The used secret image of Tiffany with a resolution of 128 X 128 pixels is shown in table 4. The results are shown in table 5, table 6, and table 7, resulting from the simple LSB method. The results are shown in table 8, table 9, and table 10, resulting from the optimal LSB method. The comparison between simple LSB and optimal LSB is shown in table 11.

Table 4. The Used Cover-Image and Secret Image.

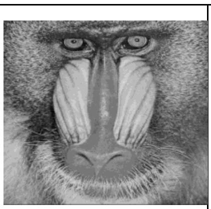
Covers	 Babbon 512 X 512 Pixels	Lena 512 X 512 Pixels
Secret	Tiffany 128 X 128 pixels	

Table 5. The Results of simple LSB 1.

Simple LSB 1		
Cover-image	**Secret-image**	**PSNR**
Baboon	Tiffany 128 X 128 pixels	54.1553
Lena	Tiffany 128 X 128 pixels	54.1679

Table 6. The Results of simple LSB 2.

Simple LSB 2		
Cover-image	**Secret-image**	**PSNR**
Baboon	Tiffany 128 X 128 pixels	46.1223
Lena	Tiffany 128 X 128 pixels	45.9163

Table 7. The Results of simple LSB 4.

Simple LSB 4		
Cover-image	**Secret-image**	**PSNR**
Baboon	Tiffany 128 X 128 pixels	33.4324
Lena	Tiffany 128 X 128 pixels	33.0581

Table 8. The Results of the Optimal LSB 1.

Optimal LSB 1		
Cover-image	**Secret-image**	**PSNR**
Baboon	Tiffany 128 X 128 pixels	56.1464
Lena	Tiffany 128 X 128 pixels	55.1614

Table 9. The Results of the Optimal LSB 2.

Optimal LSB 2		
Cover-image	**Secret-image**	**PSNR**
Baboon	Tiffany 128 X 128 pixels	53.1434
Lena	Tiffany 128 X 128 pixels	54.1400

Table 10. The Results of the Optimal LSB 4.

Optimal LSB 4		
Cover-image	**Secret-image**	**PSNR**
Baboon	Tiffany 128 X 128 pixels	50.1307
Lena	Tiffany 128 X 128 pixels	49.1308

Table 11. Comparison of the Results.

Cover-image	**K-LSB**	**Simple LSB PSNR**	**Optimal LSB PSNR**
Baboon	1	54.1553	56.1464
	2	46.1223	54.1434
	4	33.4324	50.1307
Lena	1	54.1679	55.1614
	2	45.9163	53.1400
	4	33.0581	49.1308

Comparison Between Simple LSB and Optimal LSB Using Baboon as a Cover Image is shown in Figure 5. Comparison Between Simple LSB and Optimal LSB Using Baboon as a Cover Image is shown in Figure 6.

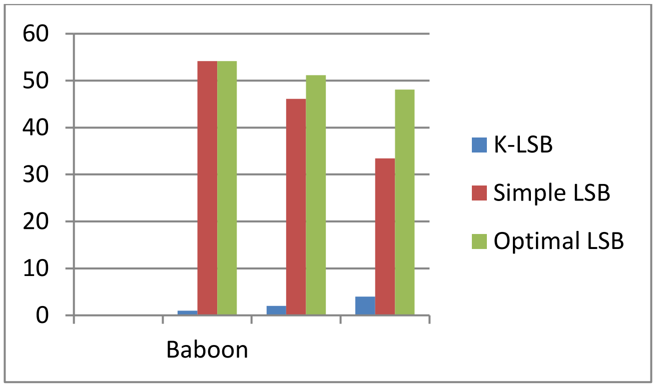


Fig. 5. Comparison Between Simple LSB and Optimal LSB Using Baboon as a Cover Image.

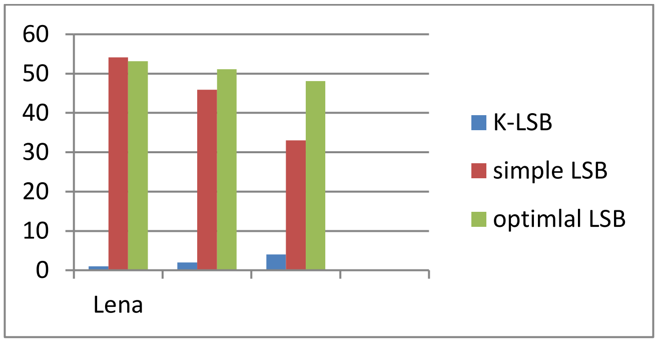


Fig. 6. Comparison Between Simple LSB and Optimal LSB Using Lena as a Cover Image.

6. Conclusion

The most common technique for image steganography is the least significant bit (LSB) substitution method, which can conceal secret information in an image with a large payload without the human visual system (HVS) detecting any distortion. In addition to the GEP algorithm, the proposed scheme uses the LSB matching method as a primary step. The results demonstrated the method's effectiveness in reducing distortion in the stego image, as it became difficult to notice any change with the naked eye, demonstrating how difficult it is for attackers to obtain confidential information.

References

[1] Mohamed, M. H., El-Gendi, S. E., Al-Afari, F., El-Melegy, M., MSc Thesis, Faculty of Science, Assiut University, 7 (2009) 19-28.

[2] M. H. Mohamed, F. Al-Afari, M. A. Bamatraf, International Arab Journal of e-Technology, 2 (2011) 11-17.

[3] C. Irawan, D. I. M. Setiadi, C. A. Sari, E. H. Rachmawanto, Informatics and Computational Sciences (ICICoS), International Conference on, IEEE, Indonesia (2017) 1-6.

[4] M. H. Mohamed and H. Abul-Kasim, International Journal of Computer Applications 45 (2012) 13-20.

[5] M. H. Mohamed, M. N. AL-Aidroos, A. M. Bamatraf, International Journal of Engineering Research and Technology, 1 (2012) 1-7.

[6] Ki-Hyun. Jung and Kee-Young. Yoo, Multimedia Tools and Applications, Springer, 74 (2015) 2143-2155.

[7] M. H. Mohamed, and L. M. Mohamed, Applied Mathematics & Information Sciences, 10 (2016) 269-266.

[8] K. U. Singh, Journal of Engineering Research and Applications 4 (2014) 105-108.

[9] J. Fridrich, and M. Goljan, Security, steganography, and watermarking of multimedia contents VI, International Society for Optics and Photonics, 5306 (2004) 35-45.

[10] C. Parthasarathy, and S. K. Srivatsa, Journal of Theoretical and Applied Information Technology, 7 (2005) 080-086.

[11] J. Mielikainen, IEEE signal processing letters, 13 (2006) 285-287.

[12] M. H. Mohamed, N. M. AL-Aidroos, M. A. Bamatraf, MIS Review, an International Journal, 18 (2012) 57-69.

[13] T.C. Lu, C.Y. Tseng, J.H. Wu, Signal Process, (2015) 77–89.

[14] C. Ferreira, In advances in soft computing. Springer, London, (2003) 257-265.

[15] C. Ferreira, Journal of Complex Systems 13 (2001) 87-129.

[16] C. Ferreira, In recent developments in biologically inspired computing. Igi Global, (2005) 82-103.

[17] K. Thangadurai, and G. S. Devi, Journal of Computers & Security, ScienceDirect, 32 (2014) 192-206.