The insertion of confidential information into digital images using blowfish cryptography and end of file steganography

E Aribowo* and A Suryadi

Department of Informatics Engineering, Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

*Email: ekoab@tif.uad.ac.id

Abstract: Identity Theft Resource Centre revealed there were 668 cases of cybercrime occurring with a total number of lost data 22,408,258 started from January to July 2018. It indicates that there was a vulnerability in the defences from attacks carried out through the internet. Thus, the world of information systems should also be accompanied by a high level of security. The stages of system development were started from the analysis of system requirements, the design of the UML, implementation, and testing. The testing process into two-phase. Firstly, system output testing in the form of stego image with various criteria, namely imperceptible, fidelity, recovery, robustness, and histogram. Secondly, system feasibility testing using the SUS method. Based on the results, it revealed that the merging of the blowfish cryptographic method and the end of file steganography was not very effective because, based on the output testing with robustness criteria; it was proven that the message inserted into the image was damaged during extraction. The changes in the size of the original cover with the resulted stego image had increased file size with a ratio of 1: 5.5, meaning that each created stego image was five times the size of the original image.

1. Introduction
The rapid development of the Internet provides complete facilities and easiness for its users, especially for a medium of communication and data transmission. If you look at the existed data, along with the development of the internet and applications that require the internet, the number of crimes in the world of information systems is also increasing [1]. Identity Theft Resource Centre (ITRC) revealed there were 668 cases of cybercrime occurring with a total number of lost data 22,408,258 started from January to July 2018 [2]. Based on the fact from ITRC, the security of information is not a trivial topic. There is much information that should be kept its confidentiality, especially personal information, financial data, and nation data. Along with the rapid development of the internet, the security of information on the internet should be more increased. Various techniques are used to protect the information from the third person in terms of ownership, access, and use. Two of the popular technique are cryptography and Steganography [3]. The cryptography technique will encrypt the messages that will be sent into random
information that is meaningless or often called as chipper text using specific algorithms and keywords before being transmitted over the internet. Then, it will be decrypted into meaningful information that can be understood (plaintext) by the recipient when the information or message has been received using the key owned.

However, the security of information by using cryptography has some weaknesses. Firstly, the random information (chipper text) can be seen clearly, so it attracts people's curiosity to break the encryption. Secondly, the chipper text can be solved in a short time, along with the rapid growth of hardware and increasingly efficient optimization software makes computing capabilities faster [4]. Therefore, another technique is needed to address the weaknesses of the cryptographic technique. This technique is called Steganography. Steganography is the study and art of hiding messages into a media in such a way so that no one knows or realized that there is a message contained in the media because it is not visible to the human sense of sight [5].

2. Literature review

2.1. Review of previous relevant studies

Based on result of research [6], after conducting various tests such as MSE (mean square error), PSNR (peak signal to noise ratio) and coefficient, it was found that the LSB method (least significant bit) proved to be better in terms of quality while the resistance in the form of salt and pepper attacks MSB (most significant bit) is superior.

Moreover, other research by [7] is an application that can insert messages into images using blowfish cryptographic techniques and End Of File steganography. In this application, the information that will be sent is encrypted and then inserted into an image. However, in this study, it was only limited to the process of inserting messages into a cover without being decrypted into plaintext.

2.2. Review of literature

2.2.1. Cryptography

In general, cryptography is defined as the study and art of encoding, aiming to maintain the security and confidentiality of a certain message. Cryptography is the study of encryption techniques where random data using certain encryption key, so it becomes hard to read by someone who does not have a decryption key. Cryptography is the study and art of maintaining the security and confidentiality of messages by encoding them into forms that cannot be understood [8]. Furthermore, there is also another definition; it is a study of mathematical techniques related to security aspects such as data confidentiality, data validity, data integrity, and data authentication.

Blowfish is a type of symmetric-key cryptography that has encryption and decryption processes using the same key, which an efficient algorithm security [9]. It's an encryption method similar to the DES cryptographic algorithm. However, after a test was conducted to evaluate the DES algorithm with Blowfish, it can be concluded that the encryption security and encryption process speed of the blowfish algorithm is superior, whereas for the power consumption used for the two algorithms has the same magnitude [10].

Blowfish algorithms are divided into two important phases. They are expansion and encryption data.

a. Key Expansion has a function to change keys (32-bit minimum and 448-bit maximum) into sub-key arrays with a total of 4168 bytes.

b. Data encryption Consists of simple function iterations (Feistel Network) of 16 cycles. Each round consists of key-dependent mutation and key-and data-dependent substitution. All operations are addition and XOR to 32-bit variables. Another additional operation is four indexed array table lookups for each round.
2.2.2. **Blowfish algorithm**

Since Blowfish uses large sub keys, the key must be counted before encrypting or decrypting data. Blowfish is an algorithm that applies a Feistel network consisting of 16 rounds [6]. For the flow of the encryption algorithm, the Blowfish method is explained as stated:

a. The initials of the P-array are 18 (P1, P2, ..........., P18), each initial consist of 32-bit. The P array consists of eighteen key, 32-bit sub key keys: P1, P2, ......., P18.

b. 4 S-boxes are 32-bits, each with 256 entries. Four 32-bit S-boxes each have 256 entries:
   - S1,0, S1,1, .......... S1,255
   - S2,0, S2,1, .......... S2,255
   - S3,0, S3,1, .......... S3,255
   - S4,0, S4,1, .......... S4,255.

c. The Plaintext to be encrypted is assumed to be an input, the Plaintext is taken as much as 64-bit, and if it is less than 64-bit, then we add more bits so that later the operation matches the data.

d. The results of the retrieval were divided into 2, the first 32-bit is called XL, and the second 32-bit is called XR.

e. Next do the XL = XL xor Pi and XR = F (XL) xor XR operations.

f. The results of the operations above then were exchanged XL to XR and XR to XL.

g. Do the process 16 times. In 16th repetition, then repeat do the XL and XR exchange process.

h. In the 17th process, do XR = XR xor P17 and XL = XL xor P18 operations.

i. The last process is to bring back XL and XR so that it becomes 64-bit again.

Blowfish uses a Feistel network consisting of 16 rounds. To be more precise, the Feistel network scheme can be seen in Figure 1.

![Figure 1. Feistel Network Scheme](image-url)

The blowfish algorithm is unique in terms of the decryption process. The decryption process is carried out in the same order as the encryption process. However, decryption process P1, P2, ..., P18 is used in reverse order. In the Blowfish algorithm, there is also an F function. F function is:
a. Divide XL into four 8-bit parts: a, b, c, and d.
b. \( F(XL) = ((S1, a + S2, b \mod 232) \text{ xor } S3, c) + S4, c\mod 232 \)

Subkeys are calculated using the blowfish algorithm, with the method as follows:

a. First, initialize the P-array and then four S-boxes in sequence with a fixed string. This string consists of hexadecimal digits of Pi.
b. XOR P1 with the first 32-bit key, XOR P2 with the second 32-bit key, and so on for each bit of the key (up to P18). Repeat the key bits until the entire Parray of XOR is changed with the key bits.
c. Encrypt all zero strings with the Blowfish algorithm by using subkeys as described in steps (1) and (2).
d. Replace P1 and P2 with the output from step (3).
e. Encrypt the output from step (3) with the Blowfish algorithm with a modified subkey.
f. Replace P3 and P4 with the output from step (5).
g. Continue the process, replacing all elements of the P-array, then all four S-boxes in sequence, with the output changing continuously from the Blowfish algorithm. The F function in blowfish can be seen in Figure 2

![Figure 2. Function in Blowfish Algorithm [6]](image)

2.2.3. Steganography
The word steganography (steganography) derived from the Greek word \textit{steganos}, which means hidden or veiled and \textit{graphics} means writing. It is defined as veiled writing [5]. Steganography is the study and art of hiding confidential messages so that they are not detected by the human senses. Steganography consists of two main processes. They are the process of insertion and extraction. The insertion process is the process of hiding information on the media to be boarded or often with media coverage.

The extraction process is the process of removing secret messages from the stego medium. Data that becomes stego media is usually in the form of general data that is often sent like audio, photo, or video. In an ideal situation, anyone who scans the data will not know that the data contains confidential information so that the retrieval of information can only be done by those who are entitled to receive the information.

In hiding the messages using steganography techniques, the several criteria that must be required [3], as follows:

a. Imperceptible means that every message inserted cannot be perceived by the human senses. If a message is inserted into a picture, then the picture that has been pasted by the message or information must be indistinguishable from the original image by the human sense of sight. For hiding the sound media, the ear must not find the difference between the original sound and the sound that has messages inserted. Similar to them, video as a combination of images and sound, the
eyes and ears must not be able to detect the difference between the original file and the file that has been pasted.
b. Fidelity means that the quality of the storage media (stego medium) does not change due to insertion.
c. Recovery means that confidential messages can be revealed again. The purpose of steganography is hiding information so that at any time, the confidential information can be retrieved and reused as needed.
d. Robustness means the data or information that hidden must be stand in the media cover manipulation. When the object processing operation is done, the hidden data will not be broken.

2.2.4. End of File algorithm
The end of File (EoF) technique is one of the steganography techniques. This technique is to utilize the attached data in the end of the file. This technique also able to insert a message that sized based on the need. Thus we will able to insert a message based on our need. However, we need to remember that the stego size transformation will increase significantly if the image/cover is bigger than 5 MB. This certainly needs attention to avoid damage in a stego image pixel. In this technique, the message also inserted in the end of the file.
On the digital image processing, histogram has an important role that able to use as a determinate indicator of a digital image quality. These are the roles of histogram in digital image processing:
1. Use to observe the spread of color intensity and can be used for decision making, for example in increasing the brightness or stretching the contrast of an image and color distribution.
2. Use to determine the boundaries in separating objects from their backgrounds.
3. Give a presentation of color composition and texture intensity for the purpose of identification of digital images.

2.2.5. System Usability Scale (SUS)
Usability is a qualitative analysis that determines how easily the user uses the interface application. An application product is categorized or called "usable" if it meets the criteria of useful, efficient, effective, satisfying, learnable, and accessible System Usability Scale (SUS) is an effective and reliable usability testing package for use in a various products and applications [11]. This SUS test has 10 questions using Likert scale of 1 to 5. Each question with an odd number (1,3,5,7,9) is a positive question, while an even-numbered question (2,4,6,8,10) is a negative question.

3. Finding and discussion

3.1. Designing the system
Modelling is the process of designing software before a coding. Creating a model of a complex system is very important because we can understand the system as a whole. Besides modelling can also be used as a visualization and construction in a software manufacturing process. In designing this software, the analysis is done by modelling the problem in the form of UML diagrams.
This application is designed to look as user friendly as possible, making it easier for users to use this application. Moreover, this application is not only targeted for users from the Informatics Engineering department, but for any kind of users. Therefore the appearance of this application is made by selecting the layout of the menu, the choice of color or text is made as attractive as possible.
3.2. Testing result

3.2.1. System output testing

- Imperceptible & fidelity criteria testing

This test is done in a subjective way, all digital images that will try to insert a message as much as 256 words with a file size of 1.62KB. This test as Table 1 is done by asking for the assessment of 5 respondents who of course have the most important criteria which is to have good vision as.

| Image | Original size | Stego image size | Result |
|-------|---------------|-----------------|--------|
| ![Image](image1.png) | 41KB | 1024KB | 5 respondents do not see the difference |
| ![Image](image2.png) | 21KB | 778KB | 5 respondents do not see the difference |
| ![Image](image3.png) | 17KB | 749KB | 5 respondents do not see the difference |
| ![Image](image4.png) | 8KB | 148KB | 5 respondents do not see the difference |
| ![Image](image5.png) | 14KB | 148KB | 5 respondents do not see the difference |

- Recovery criteria testing

A steganography can be categorized as good if all messages that have been inserted are able to be expressed again (recovery / extraction) because the purpose of steganography itself is to hide a message. Sometimes at a certain time, a message must be taken for a various purpose in the further use. Therefore, in testing this criterion, the inserted message has a different number of characters, ranging from 16 words, 64 words, 128 words, and 256 words. In this test as Table 2, the indicator used as a reference that the message recovery is successful if all the message characters are saved can be returned in full without any difference.

| Image | Words number | Extraction result | Result |
|-------|--------------|-------------------|--------|
| ![Image](image1.png) | 16 words | 16 words | Success |
| ![Image](image2.png) | 64 words | 64 words | Success |
| ![Image](image3.png) | 128 words | 128 words | Success |
| ![Image](image4.png) | 250 words | 250 words | Success |
| ![Image](image5.png) | 16 words | 16 words | Success |
| ![Image](image6.png) | 64 words | 64 words | Success |
| ![Image](image7.png) | 128 words | 128 words | Success |
| ![Image](image8.png) | 250 words | 250 words | Success |
Robustness criteria testing

Robustness criteria are information/messages that are hidden into digital images that are resistant to a form of manipulation performed on stego images (Table 3). To do this test various manipulations will be carried out on the stego image. Then, the message that has been inserted will be tried to be extracted again.

The testing step is done by making a stego image using the system, where the inserted message consists of 10 words. After successfully embedded/encoding, the stego image will be manipulated by changing it into a number of patterns, and then it will be extracted again. Of the 6 pieces that have been used, one will be chosen as a robustness test, namely the Lena.png image.

Table 3. Test of robustness criteria

| Stego image   | Manipulation Types          | Extraction Result                                                                 | Result                           |
|---------------|----------------------------|-----------------------------------------------------------------------------------|---------------------------------|
|               | 90° Rotation                | The message has been extracted but the cyphertext is broken                        | Cyphertext is failed to decrypted|
|               | 50% brightness addition     | The message has been extracted but the cyphertext is broken                        | Cyphertext is failed to decrypted|
|               | Mercury effect addition     | The message has been extracted but the cyphertext is broken                        | Cyphertext is failed to decrypted|
|               | Zeke effect addition        | The message has been extracted but the cyphertext is broken                        | Cyphertext is failed to decrypted|

System usability scale testing

This testing scenario is done by presenting how the system works, how to use the system to respondents, and asking respondents to try the system, the result as Table 4. The number of respondents used in this system test is 10 people from various background/study programs in order to know that this system can be used by all types of users.
Table 4. Respondent systems

| User          | SUS Score | Grade       |
|---------------|-----------|-------------|
| 10 user testing | 75.625    | Acceptable (B) |

3.3. **Respondents calculation result**

By looking at the SUS value obtained at 75.625 and rank with grade B, then this application is categorized acceptable or acceptable.

3.4. **Recapitulate all questions points**

In accordance with the explanation on the theory of the system usability scale (SUS) that questions that are odd numbered are positive-pitched questions, while questions that are even-numbered are negative. This means that the more agreed the respondent is with the odd numbered questions, the higher the positive score obtained. And conversely, the higher score even numbered the respondent disagrees with the question [11]. In accordance with the tests that have been done, the recapitulation of the results of all odd and even question values can be seen in Figure 3

![Figure 3. Respondents score diagram](image)

Referring to Figure 3, it can be seen that Question No. 10 has the highest score which means that most respondents gave a negative assessment of the aspects in question number 10. From this, it can be identified aspects that become the lack of this application. So for the development of this application which is the main focus aside from the main function of the application are aspects related to question number 10. If it is further examined questions that are positive in tone also still have less than the maximum value such as question no 1, then this needs to be deepened and needs to be improved in the future.

4. **Conclusion**

From all the research steps that have been carried out, the following conclusions can be drawn as follows: First, the application is made capable of implementing a combination of cryptographic blowfish and end of file steganography very well. Without any damage to the image which is used as a cover for hiding messages. Second, the performance of the EOF method (Output system testing) was tested using imperceptible, fidelity, recovery, robustness, and histogram criteria. It turns out that this EOF method successfully passed the imperceptible, fidelity and recovery tests but was unable to pass the robustness criteria test. While for the histogram, found a slight difference between the original image histogram with the stego image histogram. Finally, from the usability scale (SUS) system testing conducted on 10 respondents, this application has an acceptable of 72.5%, meaning that this application has a good usability level.
References

[1] Moreno F M M, Blanco F, Garaizar P and Matute H 2017 Fishing for phishers Improving Internet users’ sensitivity to visual deception cues to prevent electronic fraud 2017 Comput Human Behav vol 69 pp 421–36.

[2] Hamilton A 2019 Identity theft resource centre (ITRC): Third party data breach becomes focus a hacker [Online] Available: idtheftcenter.org/tag/cybercrime/. [Accessed: 26-Apr-2019].

[3] Chauhan S 2017 Multiple layer text security using variable block size cryptography and image steganography in 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT) pp 1–7.

[4] Mishra R and Bhanodiya P 2015 A review on steganography and cryptography in Conference Proceeding - 2015 International Conference on Advances in Computer Engineering and Applications, ICACEA 2015 pp 119–22.

[5] Hussain M, Wahab A W A, Bin Idris Y I, Ho A T S and Jung K H 2018 Image steganography in spatial domain: A survey Signal Process Image Commun vol 65 pp 46–66.

[6] Dwi G, Rumani and Nasrun M 2015 Implementation of cryptography and steganography in image media using the blowfish algorithm and the least significant bit method e-Proceeding Eng vol 2 no 2 pp 3762–69.

[7] Abbas T, Beiji Z and Abdullah M Y 2011 Information security technique in frequency domain Int. J Digit Content Technol its Appl vol 5 no 12 pp 279–89.

[8] Rwabutaza A, Yang M and Bourbakis N 2012 A comparative survey on cryptology-based methodologies Int J Inf Secur Priv vol 6 no 3 pp 1–37.

[9] Abdul N and Mustafa A 2014 The effect of the smoothing filter on an image encrypted by the blowfish algorithm then hiding it in a BMP image J Next Gener Inf Technol vol 5 no 4 pp 83–92.

[10] Joshi A, Wazid M and Goudar R H 2015 An efficient cryptographic scheme for text message protection against brute force and cryptanalytic attacks Procedia Comput Sci vol 48 no Iccc pp 360–66.

[11] Martins A I, Rosa A F, Queirós A, Silva A and Rocha N P 2015 European portuguese validation of the System Usability Scale (SUS) Procedia Comput Sci vol 67 no Dsai pp 293–300.