RESEARCH ARTICLE OPEN ACCESS

Using Steganography within WEB Security and Its Application in User Login System

Toghrul VALIBAYLI*, Ahmet GÜRHANLI **
**(Department of Computer Engineering, Istanbul Aydin University, Turkey*
***(Department of Computer Engineering, Istanbul Aydin University, Turkey*

ABSTRACT
Currently, the modern information society is increasingly demanding research and development in the field of steganography, which is associated with the numerous uses of digital multimedia formats. However, at the same time, there are problems of resource management and copyright for digital files. This raises the urgent task of hiding information within the framework of the network communication infrastructure of Internet participants in the media space. At the initial stage of the research, a state analysis was carried out in the field of steganographic algorithms designed to hide information. Because of this analysis, it was concluded that it is necessary to develop a steganography algorithm that hides large amounts of data in still images of widely used graphic formats. The purpose of the research is to evaluate several steganographic methods and algorithms that embed and hide large amounts of information.

Key Words: Steganography, Web Security, User Login System

Date of Submission: 17-01-2020 Date Of Acceptance: 05-02-2020

I. INTRODUCTION

One of the most important factors in the social development of the second half of the 20th century, the importance of which will increase in the 21st century, is the ongoing scientific and technological revolution in the field of computer technology and telecommunications. The consequence of this was a significant acceleration of the globalization of the processes of economic and political development of human society, increasing the importance of its "informational dimension." A growing number of states are beginning to realize the presence in this area of significant common interests, the need to combine the efforts of the international community in combating threats arising in the information sphere. Since 1998, UN resolutions have increasingly expressed concern that information technology may adversely affect the security of states in both the civilian and military spheres, and calls for the promotion of multilateral consideration of existing and potential threats to information security (IS), as well as possible measures to neutralize these threats.

Currently, the modern information society is increasingly demanding research and development in the field of steganography, which is associated with the numerous use of digital multimedia formats. However, at the same time, there are problems of resource management and copyright for digital files. This raises the urgent task of hiding information within the framework of the network communication infrastructure of Internet participants in the media space.

At the initial stage of the research, a state analysis was carried out in the field of steganographic algorithms designed to hide information. Because of this analysis, it was concluded that it is necessary to develop a steganography algorithm that hides large amounts of data in still images of widely used graphic formats.

The purpose of the research is to evaluate several steganographic methods and algorithms that embed and hide large amounts of information.

To achieve this goal during the dissertation research, it is necessary to solve the following tasks:

1. To analyze the classes of steganographic algorithms.
2. To develop a steganographic algorithm performing the operations of embedding a large amount of information in a graphic digital image on the transmitting side and extracting embedded information on the receiving side.
3. To implement the reliable functioning of the developed steganography algorithm in case of loss of bits, performing inter-format conversions.
4. Test the developed steganography algorithm and similar systems available, comparing the speed and volume of embedded data.
5. Identify the most promising areas of application of the developed algorithm.

The object of the study is steganography methods and algorithms, based on which it is possible to create a system for converted transmission of a large

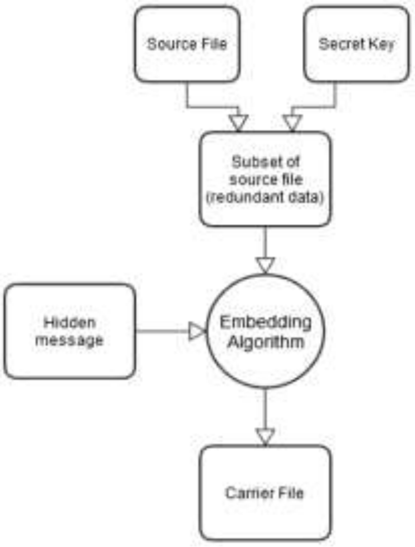
amount of data in the absence of embedded visualization artifacts.
The solution of these problems was carried out using methods of the theory of information security, probability theory and cryptography, steganography, steganography algorithms and file conversion.

II. WHAT IS STEGANOGRAPHY?

Steganography is the art of concealing hidden information in a visible environment. Steganography is literally a method of hiding information in Greek, which is derived from stego-confidential and grahy-writing meanings, which is covered writing (Petitcolas, Anderson, & Kuhn 1999:1062). The distinctive feature of steganography from cryptography is that although confidential information is considered, it is not noticeable. Because letter cryptography or systematic arrangement of numbers in cryptography can be, remarkable and major crypto-analysis can be applied for solution (Mandal & Das 2012:17). Information security is directly proportional to how well the data to be sent is hidden. In this case, the information to be sent in the steganography cannot be noticed by anyone else, if he knows which channel and how to send it to the other party.

In the data storage procedure of the steganography system in Figure 1 below, the secret information is stored in the information carrier (voice, video, text files, IP packets, etc.) with the data-embedding algorithm. The data medium to be stored may vary according to confidential information. Only the sender and receiver should also know the data-embedding key (stagey). Coding can also be used to protect information privacy.

Figure 1. Data embedding procedure in steganography



Source: https://peerj.com/articles/cs-7/, 19.01.2020

Stego, the cover data, is made more reliable with the help of the key that the receiver also has and if encryption is applied, it is encrypted. The environment can also be listened to by the active or passive guards (spies in the communication channel, spying in the non-communication channel) on the cover channel. Confidential information is obtained by data extraction (Petitcolas, Anderson, & Kuhn 1999:1063).

According to studies, the first steganography technique dates back to Ancient Greece. Steganography methods perhaps wanted to inform Sparta that Dermetaus Xerxes was pending for the occupation during the war between Sparta and Xerxes (440 BC). He wrote the confidential information he wanted to send on the board and covered it with wax. The guards were not suspected of this situation and Dermetaus successfully performed the steganographic procedure. In addition, there were many people who carried out this process in Greek history. In the researches, data covered with a lot of wax was found. Another old example dating back to BC is the slave's hair scrapped and the slave is kept on the scalp until the hair grows, and the information is sent to the place where it has to deliver the information. The slave's hair cut on the opposite side caused the information to convey confidential information without any attention. (Chen, 2005:59)

The Germans also used steganography techniques in the First and Second World Wars. One of these is Morse codes. Morse codes provided information to be sent using short and long signs and the corresponding sound or lights. It was used in 1837 after it was discovered by Samuel Morse. In addition, using invisible ink at that time are some of the steganography methods. (Chapman, Davida & Rennhard 2001:156)

Today, information privacy is in a more important place with the inclusion of the internet in our lives. The fact that the internet is in our home and even wherever we are, everything we share by social media actually gives clues about our safety. Information on any innocent pictures or advertisements that we notice on our websites is hidden. The files that appear harmless in the communications we make over the internet (audio, video, text, picture, etc.) are suitable patterns for steganography. It is located in the areas where communication steganography performed in wired or wireless environment in a local network will be used. Steganalysis methods are recommended, especially if the size of the carrier and data to be embedded is known. If the carrier data is fixed in the computer environment and it still has confidential information, it will also be exposed to more attacks and risk. In the researches conducted after the September 11 attack, it was revealed that terrorists communicate with each other using steganographic methods. (Garg, 2011:130)

III. TYPES OF STEGANOGRAPHY ALGORITHMS

Making the concept of steganography first by storing written texts in printed publications has opened new fields of study on digital data. The simplest method of storing in text is to hide the data to be stored by hiding the first letter of each word in the text. After the Theater All Clients Keep A Tab Down At Wesleys Nook (ATTACKATDAWN), the message to be hidden is indicated as such. Text hiding methods in many similar texts have been used in communication for centuries. (Por et all., 2008:78)

If there is data to be stored on image files, basically the idea of storing it in the description field of the file comes to mind. The limited size of description field may prevent this. Areas that are not used on image files are used to hide data. This method, which is known and uncovered by most tools, is very easy to recognize and solve. Many methods are used to hide information on image data. Spatial (Image Domain) and Frequency (Transform Domain) techniques are among the methods used to embed the data in images. (Douglas, Mandy, et al, 2017:5)

The most important feature of this method is that data is embedded directly into the image file. The part that represents the pixel values during this process is the data stack values that hide the information. LSB (Least Significant Bit Insertion) method is an example of spatial domain technique (Albahar, 2017:55). In the Frequency Domain method, data storage is performed over the percentages of change on the outside. An example of this is data storage methods for image files saved in JPG format. Information embedding is done over DCT (Global Discrete Cosine) coefficient. Other common methods are; LSB (Least Significant Bit Insertion), JPEG Algorithm (DCT), BPCS (Bit-Plane Complexity Segmentation) Method, Masking and Filtering.

LSB

In LSB method, the process is performed by placing the bits of the data to be hidden in order, instead of the lowest bit of each byte of the segments of the data to be used for data storage. The resulting new stego data is packaged imperceptibly. As it is a practical method, it has high awareness and high security weakness. [44] It is easy to reach the information in the new data created when steganalysis is performed.

For example, the red color RGB equivalent is FF-01-01. It is 11111111-00000001-00000001 for Red: FF (hex), Green: 01 (hex), Blue: 01 (hex) and binary number system. Here, a new data stack is obtained with the data hidden in its most meaningless bits (111111111-000000011-000000011). It is not possible to make this digital added data directly understood by the user. When the same operation is performed on each color data, numbers with a different letter equivalent are obtained. An image data of 1024x768 size contains 819200 pixels. With this method, when a 3-bit data is hidden, the new pixel value is 2457600, which indicates that the remaining data area for the data to be hidden is quite high.

It is seen that the rate of change in the data increased to 50% with the change made to the last bits. In this way, the last two bits can be added instead of adding to the last bits. Thus, the amount of data to be stored doubles. With a 24-bit image, this data can be stored unnoticed successfully. This situation shows us that LSB will be more successful to apply on the high grayscale of 24 and 8 bit image files. The disadvantage here is that the image data takes up a large area and makes sharing distribution difficult.

DCT

In DCT JPEG Algorithm, the data is transported with little loss on the quality of high quality image files. Here, primarily RGB data is converted to luminance chrominance data. Filters can be applied in the form of brightness and interference of image data.

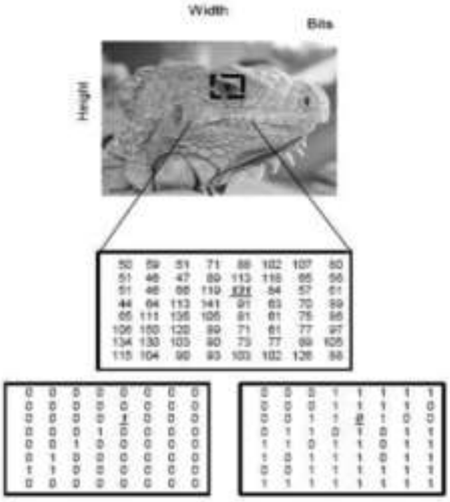
$$F(u) = \left(\frac{2}{N}\right) \sum_{i=0}^{N-1} A(i)\, cos \left[\frac{\Pi\mu}{2N}(2,+1)\right] \mathfrak{f}(i) \quad (1)$$

The new pixels obtained are transformed into blocks, subjected to DCT processing and coefficients are obtained. The file is created by adding a title to the data reaching high frequencies. The values used in embedding the data are DCT multipliers. The fact that the values used are a coefficient complicates the corruption in the file. Due to the nature of the JPEG files, the data to be hidden should not be of a high size. Because the increase in the size of the data to be hidden may cause distortions due to the corruption on the carrier file due to the coefficients. Considering the general DCT equation, the coefficient values are calculated with a simple 2D (m image from n) seen above.

BPCS

It is one of the newest methods among steganography methods. When evaluated in terms of architecture, it can be accepted within the spatial domain technique (Chawla, Kumar, 2017:85). The main difference and advantage is data hiding mass capacity. While the bulk storage capacity in stego data is 10% -17% in conventional classical methods, this area goes up to 40-60% in BPCS method. When the development stages are examined, it is seen that it is ideal for 24-bit BMP format. It is also implemented on 8 bit GIF format.

Figure 2: BPCS Algorithm



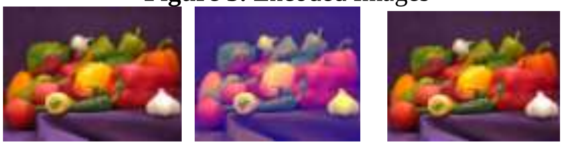
Source: Lopez-Hernandez, 2017, 3

Basically, first stage file is analyzed by BPCS algorithm. During analysis, the bits that can be used to hide data are detected. The purpose of this determination is how much image data has shape information. Thus, the image is turned into a monochrome image after analysis. For example, let's say image information with n bits. Here, an image is created with the first bits of each pixel and another image data is created using the second bits. (Lopez-Hernandez, 2017:5)

IV. EXPERIMENTAL SETUP

Standard login systems use the username and password to check the database and access the system. To make this system more secure, an Django / Python login system has been developed. In order to access this logon using Steganography, in addition to username and password, a file (image) to be generated using advanced algorithms must also be downloaded. The image can be encrypted either during registration or manually. During encryption, we get the hash version using the user-defined password sha256 algorithm and encode the image using one of the steganography algorithms (LSB, DCT, BPCS), extracting the first 30 characters of the hash we receive. The system generates a password hidden inside the image (Figure 3).

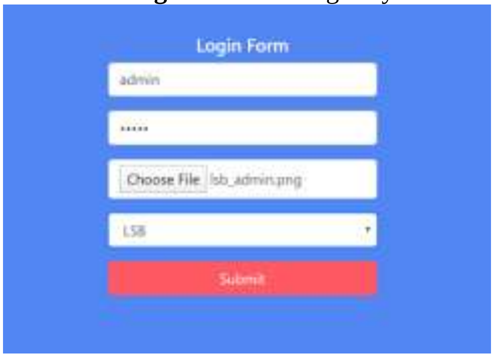
Figure 3. Encoded Images



BPCS **DCT** **LSB**

To access the system, the username and the password that was previously added to the database should be entered, and the image that was previously encrypted should be uploaded. (Figure 4).

Figure 4. User Login System



Login Form

admin

•••••

Choose File lsb_admin.png

LSB

Submit

In this case, the image is first decoded by the selected algorithm, and the presence or absence of hidden information is checked. Access to the system is not allowed if the data is empty and the image is not encoded. If information is obtained after decoding, this is saved in a single change. The user-entered password is again retrieved by the sha256 algorithm and stored in another variable, in addition to the first 30 characters, the hashing is added to another variable. In the next process, the encrypted data from the image is met with the 30-character part that the user enters and assigns. If the data is the same, it is then accessed in the database, the user is searched, and if there is such a username, the hash form of the user's input and the hash version available in the database are met. Access is allowed if it is equal. Access to the system is not allowed if the user is not found or the password is not equal.

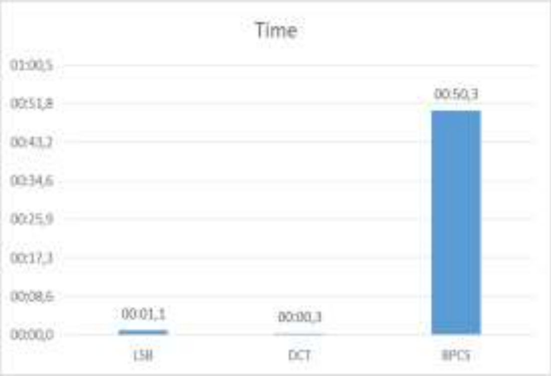
The main purpose of using steganography here is to prevent brute force attacks and minimize database queries. SQL also significantly reduces the risk of injection due to the fact that the password attached to the picture is verified.

For each of the three algorithms used in the system, speed and PSNR value measurement tests were applied. The time values of PSNR with the python data library were obtained by comparing the image with pre- and post-encoding with the numpy library.

V. EXPERIMENTAL RESULTS

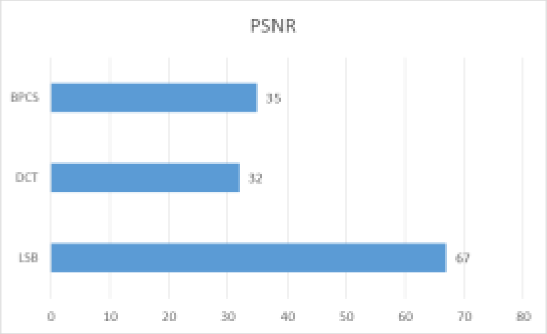
As a result, the developed algorithms were evaluated for speed and PSNR values. Speed covers from the moment one clicks on the login button to the successful processing of the login steps. By the time values, DCT is the most successful algorithm and it completes the whole process in 0.346306 seconds. The LSB completed the process in more than 1 second and BPCS takes for more than 50 seconds (Chart 1).

Chart 1: Steganography algorithms time comparison



One of the main issues in this application is to maintain the originality of the picture so that third parties cannot easily discover the steganography aims. The value of the image originality is measured by the Peak signal-to-noise ratio - PSNR. The higher the value of PSNR, the higher the quality. (Mehra, 2016:172) (Chart 2)

Chart 2. PSNR value comparison



DCT: The picture is clearly deformed and the PSNR value is the lowest with 32.
BPCS: PSNR value is much lower than LSB; the deformation in the picture is not easily discoverable when compared with DCT.
LSB: PSNR comes out very high with a value of 67, and the picture deformation is very hard to discover by human eye.

VI. CONCLUSION

Speed is one of the key factors when logging in and DCT is the fastest algorithm that completes the whole process in 0.346306 seconds. However, this algorithm produces more deformed pictures and makes a noticeable difference, which is not good for security concerns. The BPCS algorithm produces acceptable pictures but works very slowly and is not compatible with this system. The LSB algorithm is the most suitable one to maintain image quality with the highest PSNR value of 67. Even if it runs slightly slower than DCT, LSB method can be regarded as the optimal choice among the three algorithms.

REFERENCES

[1] Albahar, Marwan Ali, Et Al. (2017)"A Novel Method For Bluetooth Pairing Using Steganography." International Journal On Information Technologies And Security 9.1 : 53-66

[2] Anley, C., (2002) “Advanced SQL Injection In SQL Server Applications”, Next Generation Security Software Publication, Surrey, 2002

[3] Chawla, Rashmeet Kaur, and Sunil Kumar Muttoo.(2017) "Steganography using Bit Plane Complexity Segmentation and Artificial Neural Network." International Journal 8.5.

[4] Chapman, M. Davida, G. I. and Rennhard, M. (2001) ‘A Practical and Effective Approach to Large-Scale Automated Linguistic Steganography’, Proceedings of the 4th International Conference on Information Security, pp. 156–165.

[5] Chen, K. (2005) ‘Information Hiding , Digital Watermarking and Steganography’, Encyclopedia of Multimedia Technology and Networking, p. 8. doi: 10.4018/978-1 59140-561-0.ch055

[6] Douglas, Mandy, et al. (2017) "An overview of steganography techniques applied to the protection of biometric data." Multimedia Tools and Applications: 1-41.

[7] Lopez-Hernandez, J., MartinezNoriega, R., Nakano-Miyatake, M., & Yamaguchi, K. (2008). Detection of BPCS steganography using SMWCF steganalysis and SVM. In Information Theory and Its Applications, 2008. ISITA 2008. International Symposium on (pp. 1-5). IEEE

[8] Mandal, J. K., &Das, D. (2012). Colour Image Steganography Based on Pixel Value Differencing in Spatial Domain. IERI Procedia, 2(4), 17-24.https://doi.org/10.5121/ijist.2012.2408

[9] Mehra R.P., (2016). Estimation of the Image Quality under Different Distortions. International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 5 Issues 7 July 2016, Page No. 17291-17296

[10] Petitcolas, F. A. P., Anderson, R. J. and Kuhn, M. G. (1999) ‘Information hiding- a survey’, Proceedings of the IEEE, 87(7), pp. 1062–1078. doi: 10.1109/5.771065.

[11] Por, Lip Yee; Delina, B. (2008) Information hiding: A new approach in text steganography. In: WSEAS International Conference. Proceedings. Mathematics and Computers in

Science and Engineering. World Scientific and Engineering Academy and Society.

[12] Vural Y., (2007) "Kurumsal Bilgi Güvenliği ve Sızma Testleri" Yüksek Lisans Tezi, Bilgisayar Mühendisliği, Gazi Üniversitesi, ANKARA

[13] http://www.gokselcuryan.com/haberler/3-teknolojiyenilikleri/100-sql-injection-nedir.html, Erişim Tarihi: 20.01.2020

Toghrul VALIBAYLI. "Using Steganography within WEB Security and Its Application in User Login System." *International Journal of Engineering Research and Applications (IJERA*), vol.10 (02), 2020, pp 06-11.