A Stenography Application for Hiding Student Information into an Image

Yıldıray YİĞİT
Department of Computer Programming
Bitlis Eren University
Bitlis, Turkey
yygit@beu.edu.tr

Murat KARABATAK
Department of Software Engineering
Firat University
Elazığ, Turkey
mkarabatak@firat.edu.tr

Abstract— **Information security is a major problem today. Different approaches and methods are introduced every day for data protection. One of them is steganography. The word steganography combines the Greek words steganos (στεγανός), meaning "covered, concealed, or protected," and graphein (γράφειν) meaning "writing". The purpose of steganography is to construct the stego object by placing important information invisible into the ordinary cover object (image, sound, video, text, etc.) and to transmit it to the recipient. In this study, it is aimed to strengthen the LSB technique which is one of the steganography methods by suggesting the use of mask which will provide the least change on the image while hiding the data into a digital image. In the proposed method, the data is also compressed by the LZW algorithm, thus allowing more data to be hidden.**

Keywords: Steganography, LSB, Image, LZW, compression, data security.

I. INTRODUCTION

With the rapid development of technology and widespread use of the Internet, data sharing and storage on the internet has become inevitable. Personal data are social media sites, e-government applications, universities and so on. It is recorded in the database of many institutions. The recorded personal data can be used for many purposes, such as marketing techniques, advertising, intelligence, providing services to government agencies. Storage of data in digital media provides many advantages in interpreting and processing data. These process steps, which make our lives easier, come with very serious security vulnerabilities. Today, due to these deficits, information security is an important problem that needs to be handled carefully. Personal data can be easily obtained by third parties and personal rights can be violated. New methods and software are emerging every day to prevent this and various mechanisms are created. There are two important techniques used in the confidentiality of communication and the secure transmission of the message [1]. The first is cryptography. Cryptography is the transmission of the data to be transmitted to the receiver by means of a key, which is encrypted with a determined algorithm. If the receiver has the key, it receives and decodes the incoming encrypted message. If the receiver has no key, it will not be able to solve the message. However, with new methods developed, passwords are easily broken and important information is obtained. The second method for the secure transmission of information in computer environment is steganography. The word steganography combines the Greek words steganos (στεγανός), meaning, "covered, concealed, or protected," and graphein (γράφειν) meaning "writing" [2]. The purpose of steganography is to construct the stego object by placing important information invisible into the ordinary cover object (image, sound, video, text, etc.) and to transmit it to the recipient. The data hidden will not be seen by unauthorized persons because the data is not visible to the naked eye. Steganography; linguistics is divided into steganography and technical steganography. Linguistics steganography is text-based steganography. Technical steganography is the steganography applied to pictures, sounds and videos. For unannounced viewers, the stego object exhibits the contents of the overlay object. The stego object for recipients hides important information. Information hiding technique has been used for many years and still continues to be used. Nowadays, with the development of technology, information hiding methods are developed and new methods are introduced. Information hiding techniques today medical images, audio files, text documents and so on. often used in many places.

The reliability of a steganographic system is evaluated according to multiple criteria. These are the amount of data obfuscation (cover object) replacement, the amount of information that can be stored in the environment, and how much durability it has. The endurance power is measured by how little impacts are affected by visual and statistical attacks on the environment. With developing technology, many data hiding methods have emerged. The increase in the diversity of data hiding methods has necessitated the development of many steganaliz methods. Because each data-hiding method uses its own way, the ways to detect them are different. The steganalysis method developed for a steganographic method does not work for another steganographic method [3]. Several studies have been conducted in the field of steganography. The literature review of these are given below.

In his work in 1996, Bender described data storage in picture, sound and text in detail. In this study, data storage with low bit coding, phase coding, spread spectrum and echo data storage methods were investigated. In addition, the use of spaces, the structure of the spoken language and the methods of storing data in the text using synonym words are examined in detail in this study [4, 5].

Natthawut, in his study in 2009, gave a detailed explanation of how information is hidden from mobile phones (SMS) [6]. In the study, information was tried to hide information in SMS by using the Thai language. The reason for the use of the Thai language as a message written in this

978-1-7281-2827-6/19/$31.00 ©2019 IEEE

language has been suggested to contain more bits. In this way, he concludes that information can be hidden within the short message.

In 2008, Alwan and his colleagues (2008) performed some applications on how to select pixels in image steganography [7]. In this study, edge finding and edge pixel correlation are used. The selection of the pixel to hide the information using this correlation is described in detail.
Fridrich and his colleagues discussed and detailed the detection of steganography performed with the LSB (Least Significant Bits) in gray level and color images in their articles [8]. In this study, RS (Rescaled Range Analysis) tried to obtain information that was hidden from the data hidden by programs such as Cover image, Steganos and Hide4PGP.

In 2010, Cheddad and his colleagues investigated steganography methods and steganographic software [2]. In the study, they explained the methods and advantages and disadvantages of steganographic software while hiding data.

Sui and his colleagues (2004), in the text on the web page, instead of adding gaps, changing the position of the tags (HTML tags) has made a study that stores data [9]. By looking at the results they obtained from this study, they reported that the method they applied at the moment is still weak in terms of information security systems, although they are still weak in terms of data retention methods in picture and sound [9].

In 1998, Anderson and his colleagues (10) reported how much information could be concealed in files used to hide information. In the study, they tried to determine the limit of information in theory and practice while hiding information.

In 2013, Rahul Joshi and his colleagues discussed the LSB method in image steganography [11]. As a result of this study, the information is hidden from the LSB method by creating a key.

In 1999, Marvel and his colleagues Published a stigma technique for spreading images [12]. In this study, they stated that the message can be hidden into pictures, music, pictures or any digital signal. As a result of this study, they reported that this technique was more resistant to attacks and that the original picture was not needed when obtaining the hidden information.

Karakış and her colleagues performed steganography for medical Dicom images in their studies in 2014 [13]. In the same year, they also conducted a fuzzy logic-based steganography study [14]. As a result of these studies, it has been reported that personal information of the patients and the diagnosis and treatment of the doctor can be hidden within the image of the patient and the necessary confidentiality can be provided.

Güvenoğlu and his colleagues have tried to improve the LSB method with the SUFFLE algorithm in their articles [15]. In this study, we have used the suffle algorithm and key method. As a result, they stated that the key is required to obtain the information by the method used and that there is no noticeable difference in the image hidden in the message and there is no pixel overlap.

Ketizmen and his colleagues reported that in 2007, personal data could be easily obtained in e-government applications [16]. They pointed out that personal data could not be adequately protected.

Karabatak and his colleagues tried to improve their LSB method using a mask in color images in their studies in 2018 [17]. In their study, they reported that they provide less distortion on the image.

II. Material And Method

Steganography is hiding data into other data. The LSB method is the most commonly used method in image steganography. In the LSB method, each bit of the message to be hidden is written to the last bit of a byte of the data that creates the image file. The process of hiding data with the LSB method is as shown in Figure 1.

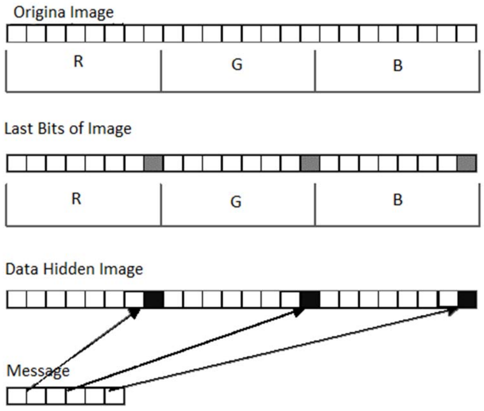

Fig. 1. Data hiding with LSB method

There is no requirement to follow a linear sorting sequence starting from the first byte to the last byte when hiding data. The order of the bytes to be used in data retrieval is regulated by a switch called stego-key and data can be hidden in a completely mixed order. In the LSB method, adding to the last two bits instead of adding to the last bit doubles the amount of data that can be hidden. While the last two bits of data hiding are performed, the change in the cover object is still invisible, while the probability of being determined by steganaliz methods is increased. This study was developed based on the study of Karabatak M. and Yigit Y. [17].

In this study, it is aimed to hide more data by compressing the data that will be hidden in addition to the baseline study. The block diagram of the developed method is shown in Figure 2.

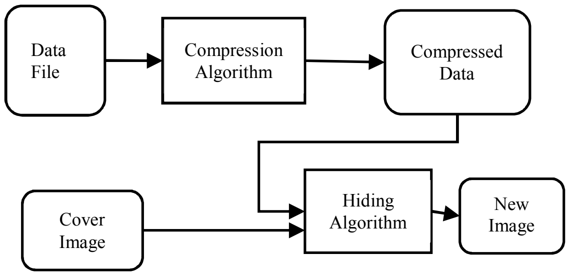

Fig. 2. Data Hiding Algorithm Block Diagram

In order to recover the data, the data hidden image file is separated into the color channels (R, G, B). The length and mask value of the data hidden from the separated color channels are obtained. As a second step, the data hidden from the color channels is obtained as a bit sequence. The final step is to obtain the compressed data by subjecting the resulting bit sequence to the reverse XOR process using mask values. In the next step, the original data is obtained using the compression algorithm. The block diagram of the data acquisition algorithm is shown in Figure 3.

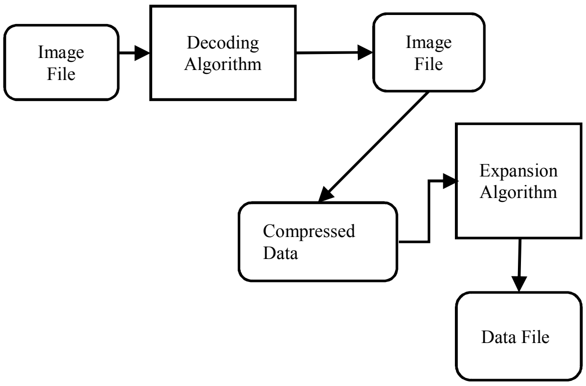

Fig. 3. Data Acquisition Algorithm

In the experimental study, by using the PSNR (Peak signal-to-noise ratio) and MSE (Mean Squared Error) values to compare the results obtained, the amount of differences between the original image and the data hidden image is calculated. The higher the PSNR value, the lower the difference between the two images. For higher PSNR, the MSE value should be close to zero. This means that there is no difference between the original version of the image and the data hidden. The PSNR value is calculated by the formula written in Eq. 1.

$$MSE = \frac{1}{m*n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \qquad (1)$$

$$PSNR = 10 * \log_{10}\left(\frac{{MAX_I}^2}{MSE}\right)$$

Since the pictures used are 24-bit color images, the PSNR value should be calculated for color images. In color images, the PSNR value is calculated separately for each color channel and the arithmetic mean of these three values is taken. For color image files, the PSNR value is calculated through the formulas in Eq. 2.

$$PSNR\ r = 20 * \log_{10}\left(\frac{255}{MSE\ \ r}\right)$$

$$PSNR\ g = 20 * log_{10}\left(\frac{255}{MSE\ \ g}\right) \qquad (2)$$

$$PSNR\ b = 20 * \log_{10}\left(\frac{255}{MSE\ \ b}\right)$$

$$PSNR = \frac{PSNR\ r + PSNR\ g + PSNR\ b}{3}$$

III. Results And Discussions

Student information and BMP file types are used in the study. The BMP file type is selected because this file format is an uncompressed 24-bit color image file. BMP Since the file type is not compressed, there is no change in the size of the image when the data is hidden. The size of the pattern file does not change and reduces the noticeability.

In the scenario applied to compare the experimental study, there are 10 student photographs and personal information of these students. Student photos are in 24 bit BMP format. In order to achieve the purpose of the study, students' personal information is hidden and recorded in their own photographs. The image files and information used are shown in Table 1. After hiding personal information, hidden photos of data are recorded in the database.

Table 1. Image Files Used

ID	Photo	Student Information
1		15042829014 KADİR CAN AKGÜN Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 mecbure13@outlook.com selçuklu mah. Hal binası sk. No: 104 Ahlat
2		16042829010 OĞUZ BAŞKARA Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 oguz.poyrazz@gmail.com vanyolu mah. örnekler 2 sk. Konak apt. no: 8/20 Erciş
3		16042829020 YUSUF CAN BİLGİÇ Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 yusuf_bilgic_13@hotmail.com erkizan mah. 2 nolu sanayi sokak no: 1 Ahlat
4		16042829001 ABDULMUTTALİP ÇİÇEK Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 talipcicek@gmail.com yankıtepe mah. Morgedik sk. Erciş
5		16042829005 EMİN GÖZEL Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 emin.gozel@hotmail.com erkizan mah. balcılar sk. no: 17 Ahlat
6		15042829029 MUHAMMED İŞLER Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 kaka_108_13@hotmail.com yeniköprü köyü yolun solu mevkii no: 81 Ahlat
7		13042829011 GÖKSEL KIRMACI Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 gokselkirmaci@msn.com selçuklu mahallesi Ahlat
8		14042829022 ÖZKAN ONUK Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 bjk_kartal72@hotmail.com zahit mah. ay yıldız cad. no: 42 / 4
9		16042829025 CİVAN ENES ÖZLEK Ahlat Meslek Yüksekokulu Bilgisayar Programcılığı 2 enesozlek@gmail.com salihiye mah. 122. taşlar sk. no: 10/1 Erciş
10		182137004 YILDIRAY YİĞİT Fırat Üniversitesi Fen Bilimleri Enstitüsü Yazılım Mühendisliği doktora özel öğrenci yyigit@beu.edu erkizan mah. Mazlum yegül cad. Taha 2 sitesi no: 309/9

The PSNR values obtained in the study are shown in Table 2. As shown in Table 2, the results of LZW+Proposed method have better performance than other methods for all images.

Table 2. PSNR Values

	LSB	LZW +LSB	Proposed Method	LZW + Proposed Method
1	37,7535	43,7656	61,1353	65,2219
2	38,3401	45,5412	60,6864	64,9349
3	36,6383	41,4850	61,2089	65,3892
4	38,5868	45,3664	60,8633	64,8999
5	36,1897	42,5502	60,7995	64,9211
6	35,5330	40,9431	60,3537	64,2992
7	38,4923	42,7814	60,8208	64,9364
8	37,3330	44,1884	60,9535	65,2691
9	39,5388	41,1429	59,8065	64,0945
10	35,1177	39,7674	60,3859	64,5246

IV. CONCLUSSION

The results of the study showed that the proposed method produces better results than the LSB algorithm. In addition, when the results obtained in Table 2 are examined, the compression of the data to be concealed has an important share in the distortion of the picture. In this study, the amount of data to be hidden is increased while the distortion on the image is minimized. The proposed method conceals personal information in photographs, allowing the data to be stored securely. The attacker will not be able to obtain personal data even if he has captured the database and photos.

REFERENCES

[1] C. Koçak, "Kriptografi ve stenografi yöntemlerini birlikte kullanarak yüksek güvenlikli veri gizleme," University of Erciyes, Journal of Naturel and Applied Science Institution, vol. 31, no. 2, pp. 115-123, 2015.

[2] A. Cheddad, J. Condell, P. M. Kevitt and K. Curran, "Digital Image Steganography: Survey and Analysis of Current Methods", Signal Processing, vol. 90, no. 3, pp. 727-752, 2010.

[3] A. Şahin, "Görüntü Steganografide Kullanılan Yeni Metotlar ve Bu Metodların Güvenirlikleri", PhD Thesis, Edirne, 2007.

[4] R. J. Anderson and F. A. P. Petitcolas, "On the Limits of Steganography", IEEE Journal on Selected Areas In Communications, vol. 16, no. 4, pp. 474-481, 1998.

[5] M. A. Atıcı, "Steganografik Yaklaşimlarin Incelenmesi, Tasarimi ve Geliştirilmesi", Master Thesis, Ankara, 2007.

[6] W. Bender, D. Gruhl and N. Morimoto, "Techniques for data hiding", IBM systems journal, vol. 3, no. 4, pp. 313-336, 1996.

[7] N. Samphaiboon, "Steganography via running short text messages", Multimed Tools Appl, no. 52, pp. 569-596, 12 2009.

[8] R. H. Alwan, F. J. Kadhim and A. T. Al-Taani, "Data Embedding Based on Better Use of Bits in Image Pixels", International Journal of signal processing, vol. 2, no. 1, pp. 104-107, 2005.

[9] J. Fridrich, M. Goljan and R. Du, "Detecting LSB Steganography in Color and GrayScale Images", IEEE MultiMedia, vol. 8, no. 4, pp. 22-28, 2001.

[10] S. Xin-Giiang and Luo Hui, "A new steganography method based on hypertext," in Radio Science Conference, 2004. Proceedings. 2004 Asia-Pacific, Qingdao, China, 2004.

[11] R. Joshi, L. Gagnani and S. Pandey, "Image Steganography With LSB", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 2, no. 1, pp. 228-229, 2013.

[12] L. M. Marvel, C. G. Boncelet and C. T. Retter, "Spread spectrum image steganography", IEEE Transactions On Image Processing, vol. 8, no. 8, pp. 1075-1083, 08 1999.

[13] R. Karakış and İ. Güler, "Medikal Dicom Görüntüler için Steganografi Uygulaması", 7th International Data Security and Cryptography Conferences, İstanbul, 2014.

[14] R. Karakış and İ. Güler, "Bulanik Mantik Tabanli Görüntü Steganografi Uygulamasi", Signal Processing and Communications Applications Conference (SIU), 2014 22nd. IEEE, Trabzon, 2014.

[15] E. M. Esin and E. Güvenoğlu, "Resim İçine Yazı Gizlenmesi Amacıyla Kullanılan LSB Ekleme Yönteminin Shuffle Algoritmasıyla İyileştirilmesi", Türkiye Bilişim Vakfi Bilgisayar Bilimleri ve Mühendisliği Dergisi, vol. 2, no. 2, 2006.

[16] K. Murat and Y. Yıldıray, "Developing LSB method using mask in colored images", 2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, 2018.

[17] K. Muammer and Ç. Ülküderner, "E-devlet uygulamalarında kişisel verilerin korun (ma) ması.", XII.“Türkiye’de İnternet” Konferansı, 2007.