스테가노그래피 방법과 응용에 관한 연구

A Study on Steganographic Methods and Its Applications

아미루자만, 김 형 중*
(Amiruzzaman Md. and Hyoung Joong Kim*)

Abstract: In this paper a detail study of existing steganographic methods are presented. An example is given of LSB substitution with uncompressed domain i.e., BMP image. In case of compressed domain JPEG image steganography is presented. Almost all popular steganographic algorithms, such as JPEG JSteg, F3, F4 and Selective Block Steganography (SBS) are described. The applications of steganographic methods are also presented briefly.

Keywords: Steganography, JPEG image, coefficient.

I. Introduction

Steganography, the art and science of invisible communication, aims to transmit information that is embedded invisibly into carrier data. Different from cryptography it hides the very existence of the secret. Its main requirement is undetectability, that is, no method should be able to detect a hidden message in carrier data. This also differentiates steganography from watermarking where the secrecy of hidden data is not required. Watermarking serves in some way the carrier, while in steganography, the carrier serves as a decoy for the hidden message. The theoretical foundations of steganography and detection theory have been advanced rapidly, resulting in improved steganographic algorithms as well as more accurate models of their capacity and weaknesses. However, the field of steganography still faces many challenges. Recent research in steganography and steganalysis has far-reaching connections to machine learning, coding theory, and signal processing. There are powerful blind (or universal) detection methods, which are not fine-tuned to a particular embedding method, but detect steganographic changes using a classifier that is trained with features from known media. Coding theory facilitates increased embedding efficiency and adaptiveness to carrier data, both of which will increase the security of steganographic algorithms.

II. Literature Review

Least significant bit (LSB) modification method is considered as a pioneer work of steganography. LSB modification and LSB matching have two different application areas. LSB modification is popular for uncompressed domain, while LSB matching is popular for compressed domain. It is found that detection processes of these techniques are also different. Nowadays, steganographic techniques are getting more secure against statistical attacks and undetectable by other different attacks. Many innovative steganographic algorithms are developed within last decade. Among them, [4, 5, 6, 7, 8] are most popular. However, many researchers are also having interest to break steganographic schemes. There are several steganalysis methods invented within last decade [3, 9]. Among them statistical attack [9] is one of the most popular and effective attacks in steganographic world. Another famous attack is the calibrated statistics attack [1, 2].

Data hiding methods have to be designed to make them secure from statistical attack because this attack is relatively easy to combat. Simple solution against this attack is keeping the same or similar stego image histogram to the original image histogram. However, keeping the same shape of a magnitude histogram is not easy to achieve as long as the coefficient magnitudes are altered.

Note that one branch of past history of steganography was inventing methods to preserve the original histogram perfectly. LSB overwriting methods including OutGuess [4] can preserve the original histogram almost perfect (in fact, not absolutely perfect). This method modifies half of the nonzero coefficients and corrects the distorted histogram by adjusting with the rest of unused coefficients. In general, perfect preservation is not possible because data pattern is not ideal, but random.

The most popular and revolutionary method is F5 by Westfeld [9]. F5 method [9] also tries to narrow the gap between original and modified histograms by decrementing nonzero JPEG coefficients to 0 and applying matrix embedding and permutative straddling.

III. Detail of Existing Methods

In the digital imaging world, steganography begins with uncompressed domain. In uncompressed domain data hiding was made by changing the LSB of image pixel value. LSB modification or substitution method is quite successful. LSB modification method is nearly impossible to detect.

When attackers start believing that LSB modification method is difficult to detect then they start to destroy all uncompressed images whenever they got chance. Any kind of uncompressed image is suspicious for attacker (i.e., Steganalysis). As the technology is upgrading new techniques are also coming from good researcher. Since, uncompressed domain became suspicious for Steganalysis researchers moved to compressed domain. In uncompressed domain JPEG JSteg, F3, F4, and F5 method is popular. Most of the methods developed based on them (i.e., JSteg, F3, F4, and F5).

1. Uncompressed domain

The LSB modification/substitution is the popular and pioneer work. This method is very simple and difficult to detect. The LSB modification method is undetectable because after data hiding, it does not make significant change in image histogram.

In LSB substitution method every pixel value is used to hide one bit

* Hyoung Joong Kim (Corresponding Author)
논문접수 : 2008. 08. 12., 채택확정 : 2008. x. xx.
아미루자만, 김 형 중 : 고려대학교 정보보호대학원
(amir@korea.ac.kr, khj-@korea.ac.kr)
※본 연구는 ITRC(Information Technology Research Center)와 MIC(Ministry of Information and Communication, Korea)의 지원을 받아 연구되었음.

of data. If the pixel value is odd and hidden bit is 0 then the pixel value is increase or decreased by one. Similarly, is the pixel value is even and the hidden bit is 1 then also that particular pixel value is increase or decreased by one (i.e., to hide 1). More specifically, the odd pixel values are representing hidden bit 1, and even pixel values are representing hidden bit 0 (see Fig 1).

20	21	24	27	31	34	35
0	1	0	1	1	0	1

Fig 1. A typical example of pixels and bit representation by the pixel values, odd pixels ore representing 1, and even pixels are 0.

From the given example it is clear that as pixel value 20 is even, thus pixel value 20 is representing 0. On the other hand pixel value 21 is odd, representing 1. If any method would like to hide 1 in all given pixels then even pixels need to change by one, while odd pixel values will remain same (see Fig 2).

21	21	25	27	31	35	35
1	1	1	1	1	1	1

Fig 2. Even pixel values are increased by one to represent 1, and odd pixels are remaining as it was.

2. Compressed domain

Compressed domain means working with coefficients instead of pixel values. In case of compressed domain the hiding capacity is lower than uncompressed domain.

JSteg:

Positive and negative AC coefficient values are used to hide data. Even positive AC coefficient values are representing 0 and odd positive AC coefficients are 1. In case of negative AC coefficients, the situation is little different. Odd negative AC coefficients are representing 1, while evens are 0. Remarkable point is during modification 0 and 1 was skipped (see Fig 3).

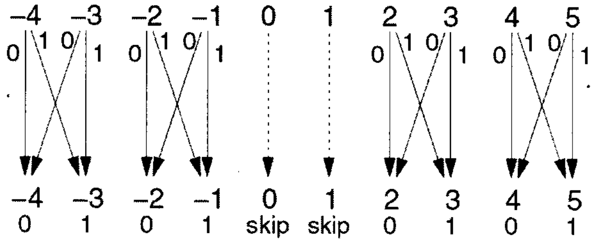


Fig 3. Data hiding technique by JSteg algorithm.

F3:

Positive and negative AC coefficient values are used to hide data. Even positive AC coefficient values are representing 0 and odd positive AC coefficients are 1. Similarly negative odd AC coefficients are representing 1, and evens are representing 0. Remarkable point is during modification only 0 was skipped, -1 and 1 shrinkage to 0. in case of hiding 0, AC coefficient value -1 shrinkage to 0, and similarly in case of hiding 0, AC coefficient value 1 shrinkage to 0 (see Fig 4).

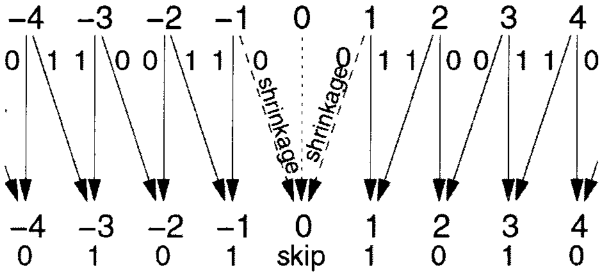


Fig 4. Data hiding technique by F3 algorithm.

F4:

Positive and negative AC coefficient values are used to hide data. Even positive AC coefficient values are representing 0 and odd positive AC coefficients are 1. At the same time negative even AC coefficients are representing 1, and odds are representing 0. Remarkable point is during modification only 0 was skipped, in case of hiding 1, AC coefficient value -1shrinkage to 0, and in case of hiding 0, AC coefficient value 1shrinkage to 0 (see Fig 5).

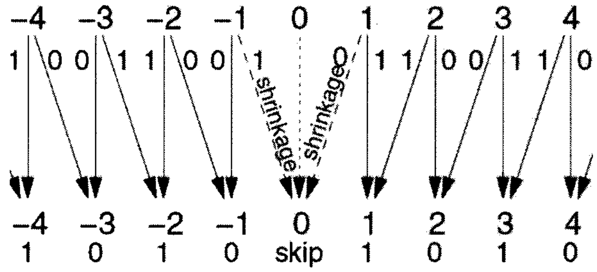


Fig 5. Data hiding technique by F4 algorithm.

F5:

The algorithm F5 is quite different with previous method i.e., JSteg, F3, F4. In F5, matrix encoding technique used, a 7-3 technique is widely used to embed hidden bits. In that case one coefficient is changed out of 7 coefficients stream and represents 3 bit.

7-3 matrix embedding is similar with 3-2 matrix embedding (see Fig XHere, a_i is position of coefficient and x_i is hidden bit. 3-2 matrix encoding is following (see Eqn. 1, and Eqn. 2)

$$\begin{array}{c|ccc|} & a_1 & a_2 & a_3 \\ \hline x_1 & 1 & 0 & 1 \\ x_2 & 0 & 1 & 1 \end{array}$$

$$\begin{aligned} x_1 &= a_1 \text{ xor } a_3 \\ x_2 &= a_2 \text{ xor } a_3 \end{aligned} \qquad (1)$$

$$\begin{array}{c|ccccccc|}
 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \\
\hline
x_1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
x_2 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
x_3 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{array}$$

$$\begin{aligned}
x_1 &= a_1 \text{ xor } a_3 \text{ xor } a_5 \text{ xor } a_7 \\
x_2 &= a_2 \text{ xor } a_3 \text{ xor } a_5 \text{ xor } a_7 \\
x_3 &= a_4 \text{ xor } a_5 \text{ xor } a_6 \text{ xor } a_7
\end{aligned} \qquad (2)$$

■

SBS:

Selective Block Steganography (SBS) is a completely new method. Where, the authors of SBS proposed the data hiding without changing the nonzero AC coefficient values. They proposed the data hiding technique by changing the number of zeros. In between two nonzero AC coefficients are used to embed their hidden bits (see Fig 6). First they separate some 8x8 JPEG blocks, which they named as a selective block (SB) and later they fix a size of shifting zeros (either to left or to right). The size of selection is determined by a threshold value (see Fig 7).

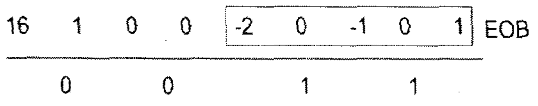


Fig 6. Data hiding technique by SBS algorithm.

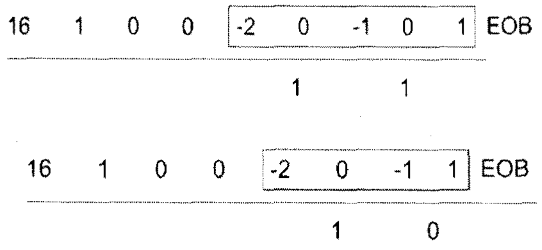


Fig 7. Data hiding technique by SBS algorithm.

SBS is the first method which is not changing any nonzero AC coefficients and can preserve exactly same histogram as original image [10].

IV. Applications of Steganography

Steganographic methods have several application areas. Such as, fingerprinting, forensics security, watermarking, and secret data transmission.

Fingerprinting, to maintain criminal database and authenticate them, steganographic methods may use. Steganographic images or texts are invisible not accessible by other people.

Forensics security is an emerging area for security and authenticity. Steganography also can be use in forensics security.

Watermarking, is kind of steganography. The basic difference is in case of watermarking the existence may be known to all. Steganography conceal the existence of hidden information. All the application areas can be use as a steganographic application area.

Secret transmission, it is seen that steganography can be very useful for secret communication. Before the 911 tragedy, the terrorist group communicated with each other by using steganographic images. Instead of mail, they choose steganographic images and texts. Top secret government communication can be done by steganography; also it can be helpful in battle-field.

V. Conclusions

This paper described few existing method of steganography, and also gave a brief about applications. This paper can be very helpful for the people who are working in the area of steganography and watermarking. This paper also described the SBS method, which is noticeable work and can be improved in future.

Acknowledgement

This work was in part supported by Information Technology Research Center (ITRC), by the Ministry of Information and Communication, Korea.

References

[1] J. Fridrich, M. Goljan, H. Hogea, "Attacking the OutGuess," *Proc. of the ACM Workshop on Multimedia and Security*, pp. 967-982, 2002.

[2] J. Fridrich, M. Goljan, H. Hogea, "Steganalysis of JPEG image: Breaking the F5 algorithm," *In Lecture Notes in Computer Science*, vol. 2578, pp. 310-323, 2003.

[3] J. Fridrich, "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," *In Lecture Notes in Computer Science*, vol. 3200, pp. 67-81, 2004.

[4] N. Provos, "Defending against statistical steganalysis," *Proc. of the 10th USENIX Security Symposium*, pp. 323-335, 2001.

[5] P. Sallee, "Model-based steganography," *In Lecture Notes in Computer Science*, vol. 2939, pp. 154-167, 2004.

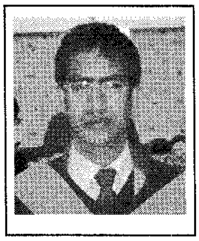
[6] P. Sallee, "Model-based methods for steganography and steganalysis," *In International Journal of Image and Graphics*, vol. 5, no. 1, pp. 167-190, 2005.

[7] K. Solanki, A. Sarkar, B. S. Manjunath, "YASS: Yet another steganographic scheme that resists blind steganalysis," *Proc. of the 9th International Workshop on Information Hiding*, Saint Malo, Brittany, France, pp.16-31, 2007.

[8] A. Westfeld, A. Pfitzmann, "Attacks on steganographic systems," *In Lecture Notes in Computer Science*, vol. 1768, pp. 61-75, 2000

[9] A. Westfeld, "F5: A steganographic algorithm: High capacity despite better steganalysis," *In Lecture Notes in Computer Science*, vol. 2137, pp. 289-302, 2001.

[10] Amiruzzaman Md., and H. J. Kim, "Selective Block Stegangraphy," *Proc. of the 3rd International Joint Workshop on Information Security and Applications*, pp. 123-133, 2008.



Amiruzzaman Md. received the B.S degree in Computer Science and M.S. degree in Computer Science and Engineering from National University, Bangladesh in 2002 and Sejong University, Korea (ROK) in 2006 respectively. Now he is a PhD student in the Graduate School of Information Management and Security, Korea University, Korea (ROK). His research interests Steganography, Watermarking, and Information Theory.

Hyoung Joong Kim received the B.S degree in Electrical Engineering from Seoul National University, Korea (ROK) in 1978. He also received M.S, and PhD in Control and Instrumentation Engineering from Seoul National University, Korea (ROK) in 1986, and in 1989 respectively. Currently he is a full professor in the Graduate School of Information Management and Security, Korea University, Korea (ROK). His research interests Multimedia Computing, Multimedia Security, Semantic Analysis, and e-Learning applications.