Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

Research on Steganography Technology and Principles of Steganalysis Based on Digital Graphics

Lei Shi [1, 2, 3, 4,*]

[1] Shaanxi Provincial Land Engineering Construction Group Co., Ltd, China;

[2] Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710075, China;

[3] Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Land and Resources, Xi'an 710075, China;

[4] Shanxi Provincial Land Engineering Construction Group Co., Ltd, Xi'an 710075, China.

*Corresponding author e-mail: sl19890419@foxmail.com

Abstract: Steganography usually uses the perception and statistical redundant information of multimedia files such as images, audio, video and text to complete the embedding of secret messages, and then complete the information transmission through public channels such as the Internet. Since information steganography is done through redundant information, the quality of multimedia files will not change before and after steganography. This method usually does not attract the attention of the other party, so it has strong concealment. Therefore, it is completed by information steganography technology. Secret communication is also more secure.
Keywords: Steganography Model; General steganography; Steganography algorithm; JPEG Image.

---

1. STEGANOGRAPHY MODEL

Data steganography usually requires two parts: steganography and parsing. The plaintext is the original data that needs to be transmitted. The ciphertext is generated after encryption by key 1. This step is no different from the traditional encryption system. The difference is that the ciphertext is embedded in carrier media such as images, audio and video through steganographic embedding algorithms to generate secret media. The secret media is transmitted through public channels such as the Internet. In the data decryption part, the steganographic extraction algorithm is used to extract the ciphertext in the ciphered media, and then the decryption algorithm is used to decrypt the ciphertext to generate the original plaintext. So far, the secret communication of information is completed [1].

After the secret message is encrypted, it is embedded in the carrier file with a steganography algorithm to form a secret (secret, also called secret) medium, which is transmitted in a public channel. After receiving the encrypted media, the receiver uses the steganographic extraction algorithm to extract the ciphertext, and then decrypts the plaintext with the decryption algorithm [2]. Since the perception of secret media is no different from ordinary media, steganography is difficult to detect. The reason why the secret message to be embedded is encrypted before steganography can provide double insurance on the one hand. On the other hand, encryption tends to randomize the bit stream of the secret message to make the steganography more secure [3].

2. STEGANOGRAPHY CLASSIFICATION

The methods of hiding secret information in the carrier mainly include content replacement, insertion, generation, spread spectrum, change of statistical characteristics, deformation, etc.

The replacement method is the more commonly used steganography method in steganography technology. The process is to replace redundant or non-critical bits in the carrier medium with bits of secret information to realize the embedding of ciphertext. For example, the bit of the secret message is used to replace the least significant bit of the pixel color value (or grayscale value) in the BMP file. Since the least significant bit (LSB) has little effect on the image color or grayscale, this replacement is difficult Perceived by human senses. Due to its simple process and good robustness, this alternative method is widely used in the current data steganography field. However, there are many researches on steganalysis for this method, and it has been able to more accurately detect some steganography realized by substitution-based steganography methods, and even estimate the length of secret information [4].

The insertion method uses the place that is ignored by the application program that accesses the carrier to store secret information to realize hiding. For example, JPEG image files usually use 0xFFD9 as the end of the file. When the application that reads and displays the JPEG encounters this mark, it thinks that the file has ended and will not access the subsequent data. Therefore, hiding information after this mark will neither adversely affect the quality of the carrier, nor will it be detected by the application program that accesses the carrier file, and the capacity will not be limited. However, this method will make the size of the carrier file larger, which will cause suspicion and be easily detected [5].

Both the replacement method and the insertion method require a public carrier file to complete the secret information embedding, while the generation method does not require an additional carrier file. The generation method can directly use the secret information to control the generation of a public carrier file. Commonly used, such as creating fractal images, etc., using public file related information such as color, angle and length to represent secret information.

Referring to the idea of spread spectrum communication, steganography also has a spread spectrum method, which embeds secret information into the spectrum by expanding the spectrum of the public carrier.The deformation method saves information through signal deformation, and measures the deviation from the original carrier when decoding.At present, the LSB replacement method is most widely used in various mainstream image steganography software or tools [6].

According to whether the steganography process requires a key, steganography technology can also be divided into keyless steganography technology and key steganography technology. The key steganography technology can be divided into symmetric and asymmetric key steganography according to whether the key used is symmetric.

Keyless steganography means that the steganography process does not require the participation of a key. The steganography algorithm is directly used to embed the information plaintext into the carrier media, and the extraction algorithm can be used to extract the information plaintext directly after transmission through the public channel. No key is involved in the whole process, so steganography and extraction algorithms need to be kept secret. This steganography method has certain risks. The encryption process also violates the Kerckoffs criterion. The security of the data depends only on the cryptographic algorithm and does not rely on the choice of the key [7].

The key steganography technology means that the data plaintext is encrypted by the key before steganographic transmission. This process meets the Kerchoffs criterion. For symmetric steganography, the encryption and decryption parties hold the key k at the same time. Before steganography, use the key k to convert the information plaintext into ciphertext, and then transmit it through the public channel. After extracting the ciphertext, the receiver uses the key k decrypts to obtain the information plaintext. Symmetric key steganography requires that the sender and receiver share the key, and there is a key exchange. At the same time, the key can also be used for steganographic transmission of information.

For asymmetric key steganography, the keys held by the sender and receiver are different, and two keys are required: a public key and a private key. The sender encrypts the plaintext of the information and the public key it holds and then embeds it into the carrier through steganography. The receiver first uses the parsing algorithm to obtain the ciphertext, and then uses the private key to decode the ciphertext into plaintext [8].

3. STEGANALYSIS TECHNOLOGY CLASSIFICATION

According to the analysis methods used, it can be divided into sensory, statistical and characteristic steganalysis.

Sensory steganalysis, as the name suggests, requires the use of various senses to detect such as sight and hearing. This method requires proper conversion of the measured data first, and then detection and judgment through hearing and vision. For example, Westfeld et al. visually detect the lowest plane of steganographic images. The disadvantage of sensory analysis is its poor reliability and manual interpretation [9].

Statistical steganalysis is a method for judging the difference in the statistical characteristics of a sentence carrier. First, obtain the theoretical statistical characteristics of the original carrier, and then compare the statistical characteristics of the test object with the initial statistical characteristics to compare the differences. The premise for statistical steganalysis to take effect is to obtain the statistical characteristics of the initial carrier, which is often difficult to obtain.

Feature steganalysis is an analysis method that extracts key features of confidential data and performs detection based on these key features. This is also the most used analysis method currently.

Key features can be divided into two types: identification features and statistical features. Corresponding to the particularity of some steganographic algorithms, after steganographic processing of ordinary data, it will inevitably change the characteristics of the original data itself, leaving traces of encryption, which can be checked to determine whether it is secret. Different steganography tools produce different identification features, which can be used as a basis for detection. Obviously, a detailed analysis of different steganographic tools is needed to obtain the identification features, which makes the identification features only suitable for detecting the steganography of known steganographic tools. In contrast, statistical characteristics have nothing to do with the specific steganography algorithm. As long as the steganography occurs, more or less the statistical characteristics of the carrier will change. By analyzing these statistical characteristics, it can be judged whether the steganography exists. If the secret information is steganographically written in the JPEG image, the histogram of the discrete cosine transform coefficients of the image will change. The DWT coefficients of unencrypted image information generally satisfy the Laplace distribution, and the DWT coefficients generally change after the steganography process. Therefore, in order to realize the steganalysis and detection of unknown steganography algorithms, the use of statistical features for steganalysis is an effective method [10].

4. STEGANALYSIS OF JPEG IMAGE

The JPEG image format can compress the image size as much as possible on the premise of ensuring the image quality to achieve the purpose of reducing storage overhead, and supports the compression ratio custom function. Due to the huge advantages of the JPEG format, it has become the most widely used image storage format. Most of the research on steganography for image data is based on the JPEG format.

The JPEG image format is the most widely used image thick lip format on the Internet. It steganizes the JPEG image. The encrypted JPEG image will be submerged in a large number of JPEG images, which increases the difficulty of steganalysis. At the same time, the steganalysis of JPEG images will bring certain image quality degradation. As an image compression format, the JPEG format itself will bring certain image quality degradation, so the image quality degradation caused by compression is certain. To a certain extent, steganography will be hidden. Based on the many advantages of JPEG format steganography, researchers have also conducted a lot of research, and a number of excellent steganography algorithms have emerged, such as Jsteg, Jphide, F5, OutGuess, MB, etc.

The JPEG compression process for true color images first converts the RGB format to YCbCr format, and then samples the Y, Cb and Cr channels separately. The sampling ratio is 4:1:1. After sampling, the smallest coding unit MCU is obtained. The dimensions of the MCU are 16×16, its composition is shown in Figure 2.3. The original image is divided into blocks according to the smallest coding unit, and a two-dimensional forward discrete cosine transform (2D-FDCT) is performed with the MCU as the unit.

Transform the image data represented in the spatial domain to the frequency domain, and use different discrete cosine transform coefficients for and respectively. Based on this table, the image brightness and color quantization values can be adjusted without causing significant quality deterioration. The

adjusted quality factor is 50, and the linear adjustment of the quantization value can dynamically adjust the image compression ratio.

Decoding JPEG images will get a set of discrete cosine transform coefficients, and the steganography algorithm for JPEG images is done in the frequency domain. The information embedding is completed by modifying the quantization step size of the high-frequency components in the image frequency domain. Since the medium and high frequencies have a small impact on the image quality, the modification and embedding does not significantly change the image quality. However, this modification will change the relative relationship between the middle and high frequency step length and the low frequency step length, which can be easily recognized by the steganalysis algorithm. Therefore, the security of this method is not high. With the development of steganography technology, the new JPEG image steganography algorithm no longer modifies the relative step length of the middle and high frequencies, but directly loads the ciphertext on the discrete cosine transform coefficients according to certain rules. Common image steganography algorithms such as Jsteg, Jphide, F5, and Outguess are all designed according to this idea. The four algorithms will be described in detail below.

5. EXPECTATION

Information steganography technology can not only encrypt the secret information that needs to be transmitted, but also embed the ciphertext into common public media for transmission. It is highly deceptive and has higher security than traditional cryptography, which directly leads to information hiding. The development of writing technology. While information steganography technology brings encryption convenience, it also brings certain risks to communication security. Based on the requirements of national strategic security, trade secrets and personal privacy protection, steganalysis technology has also been extensively studied.

Aiming at the demand of JPEG image steganalysis, an image steganalysis technology based on least squares support vector machine algorithm is proposed. The analysis framework of information steganography is introduced in detail, and the feature selection and classifier design are also explained. Finally, the effectiveness and accuracy of the algorithm analysis are simulated in the MATLAB environment.

For the demand for steganalysis of voice information, this article analyzes the current main methods of using audio for information hiding, selects narrowband voice as the carrier, and compares and analyzes the changes in voice quality and related parameters before and after the steganography based on the changes in audio characteristics after time domain information is embedded , Designed and implemented a blind detection algorithm for voice information hiding.

REFERENCES

[1] Jicang Lu, Fenlin Liu, Xiangyang Luo. Recognizing F5-like stego images from multi-class JPEG stego images[J]. KSII Transactions on Internet and Information Systems(TIIS),2014,8(11).

[2] Xiangyang Luo, Fenlin Liu, Chunfang Yang, Shiguo Lian, Daoshun Wang. On F5 Steganography in Images[J]. The Computer Journal,2012,55(4).

[3] Qian Zhang,Yuan Liu,Yu Nan, Tao Zhao,Fenlin Liu. Classification Algorithm of Jsteg and F5 Stego-images Based on Histogram Difference[J]. Energy Procedia,2011,13.

[4] Mathematics; Fuzhou University Researchers Describe New Findings in Mathematics (Deep-Learning Steganalysis for Removing Document Images on the Basis of Geometric Median Pruning)[J]. Journal of Mathematics,2020.
[5] Weixiang Ren,Liming Zhai,Ju Jia,Lina Wang,Lefei Zhang. Learning selection channels for image steganalysis in spatial domain[J]. Neurocomputing,2020,401.
[6] Ying Zou,Ge Zhang,Leian Liu. Research on image steganography analysis based on deep learning[J]. Journal of Visual Communication and Image Representation,2019,60.
[7] Kenneth Sullivan,Upamanyu Madhow,Shivkumar Chandrasekaran,B. S. Manjunath. Steganalysis for Markov cover data with applications to images.[J]. IEEE Trans. Information Forensics and Security, 2006,1(2).
[8] Mehdi Boroumand,Mo Chen,Jessica Fridrich. Deep Residual Network for Steganalysis of Digital Images[J]. IEEE Transactions on Information Forensics and Security,2019,14(5).
[9] Yu Yang, Yuwei Chen, Yuling Chen and Wei Bi. A Novel Universal Steganalysis Algorithm Based on the IQM and the SRM[J]. CMC: Computers, Materials & Continua,2018,56(2).
[10] Fridrich J. and Kodovsky, J. Rich Models for Steganalysis of Digital Images. IEEE Transactions on Information Forensics and Security. 2017, 868-882.