ISSN: 2688-8645

Research on Hidden Information Analysis Technology based on Image Data

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

[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

[a]Corresponding author e-mail: sl19890419@foxmail.com

Abstract

With the continuous development of the Internet and information technology, the government, enterprises and individuals rely more and more on image, voice and other multimedia forms for information acquisition and transmission, while steganography technology can embed ciphertext into ordinary information transmission, but it is not detected, which makes the public channel may become the channel of illegal information transmission, in order to consolidate the communication security foundation and block the secret communication The research on steganalysis of public channel is very important. At the same time, because JPEG image data and voice data are widely used in network transmission process, steganalysis technology for image information and voice information is of great significance in network security and communication security. In order to solve the potential communication security risks of steganography, this paper proposes a detection method for image information and speech information. With the in-depth study of steganalysis technology, the feature extracted by the algorithm is more and more complex, and its dimension is also higher and higher, which leads to the increasing computational complexity and the requirement of computational power.

Keywords

Steganalysis; Information Safty; JPEG Image.

1. Introduction

With the continuous advancement of network infrastructure, Internet-based Multimedia technology has been rapid development. At present, social information and data transmission is mainly completed through Internet channels, and most of the data and information are in form, such as video, video and audio [1]. For the management of these, on the one hand, it is necessary to ensure its safe transmission and prevent it from being eavesdropped, embezzled and tampered with; on the other hand, it should be fully supervised to prevent illegal information from spreading through public transmission channels. Based on these two management requirements, it is technically necessary to achieve steganography and steganalysis of information [2].

This traditional encryption system not only realizes secure communication, but also brings about an obvious problem, that is, the ciphertext generated by the encryption system is often some random code with no obvious significance. To a certain extent, this feature also helps the enemy locate the ciphertext more quickly [3]. At the same time, with the continuous

ISSN: 2688-8645

improvement of computer hardware computing power, especially the rapid development of parallel computing technology based on distributed network, the security risk of traditional encryption algorithm is increasing. In view of the defects of traditional encryption system, such as the ciphertext is easy to be identified, in recent years, a new idea of information security transmission has been widely discussed and studied by researchers. This information security transmission technology different from the traditional encryption system is information technology. Information Steganography technology refers to the transmission technology of hiding ciphertext in normal and unrelated ordinary data. Information Steganography technology has been used in ancient times, such as Tibetan poems, steganography ink and newspaper code. Nowadays, with the development of multimedia technology, this information steganography technology has a new vitality in the field of secure communication [4].

2. Research Status at Home and Abroad

Steganalysis process can be regarded as the decoding of steganalysis process, which is the opposite direction of steganalysis. The relationship between steganalysis and steganalysis is equivalent to the relationship of spear and shield. According to whether steganalysis algorithm has pertinence, steganalysis can be divided into two types: using steganalysis and using steganalysis. If the information steganalysis algorithm is known, it is called special steganalysis by analyzing the characteristics of the algorithm. For example, for the common steganalysis LSB algorithm [5], researchers have proposed such methods as WS steganalysis, spa steganalysis and RS steganalysis. For the F5 steganalysis algorithm, there are also a variety of steganalysis algorithms for efficient and accurate detection. The special steganalysis technology can detect the data efficiently and accurately because it has known the characteristics of steganalysis algorithm. However, since the steganalysis is only effective under some specific steganalysis methods, if the steganalysis algorithm is replaced, the detection accuracy rate will decrease obviously, so the application scope of special steganalysis technology is narrow. In recent years, with the rapid development of machine learning and deep learning, the universal steganalysis technology based on machine learning has become a hot topic in the field of steganalysis, and a large number of steganalysis algorithms with high recognition accuracy have emerged [6].

First thought that Markov random process can be used to represent the change process of adjacent pixels, then the Markov transition probability matrix can be obtained based on the change of adjacent pixels, and the matrix is taken as the key feature of steganalysis; in 2016, researchers such as other researchers proposed to use it for decision, and the feature was used for steganalysis. However, the dimension of the matrix is relatively low, and the number of features that can be used for decision is relatively small, which leads to the low accuracy of steganalysis. Therefore, the design of a feature vector that can reflect the key features of the image is the key content of the research. With the deepening of the research, in order to improve the detection accuracy, et al. Proposed [7]. The model constructs 106 sub models. By combining the sub models, the model dimension can reach 34671 dimensions. For the problem of classifier selection, because the key features proposed in the early stage of research are relatively simple and low, FLD and SVM decision makers can meet the requirements. In order to further improve the accuracy of steganalysis, the method of "high dimensional features + ensemble classifier" is widely used in current algorithms, but this method also has obvious shortcomings, which will significantly increase the computational complexity. In 2015, cogranne et al. Designed a linear classifier, which is a regularized Fisher linear discriminator. Using lslsmr linear classifier can reduce the computational complexity, but the detection accuracy is poor [8]. Because it can analyze data independently and extract the key features of data, it is widely used in steganalysis. After xunet and ynynet are proposed, the detection ability of steganalysis algorithm based on

ISSN: 2688-8645

deep learning is obviously better than SRM and massrm, but this kind of algorithm will increase the computational complexity obviously, which is also an unavoidable problem [9].

3. Steganography Classification

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

The replacement method is a common steganography method. The process is to replace the redundant or non key bits in the carrier media with the bits of secret information, so as to realize the embedding of ciphertext. For example, the least significant bit (LSB) of pixel color value (or gray value) in BMP file is replaced by the bit of secret message. Because the least significant bit (LSB) has little effect on image color or gray level, this replacement is difficult to be detected by human senses. Because of its simple process and good robustness, this method is widely used in the field of data steganography. But at present, the steganalysis of this method has been studied more and more, and it has been able to detect some steganography based on substitution, and even estimate the length of secret information [10].

The insertion method uses the places ignored by the application program of accessing carrier to store secret information to achieve hiding. For example, JPEG image files usually have an end tag of 0xffd9. When an application that reads and displays a JPEG encounters this tag, it considers the file has ended and will not access subsequent data. Therefore, the hidden information after this mark will not have adverse impact on the quality of the carrier, will not be detected by the application program accessing 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.

Both the replacement method and the insertion method need to open the carrier file to complete the secret information embedding, while the generation method does not need additional carrier files. The generation method can directly use secret information to control the generation of public carrier files. It is often used to create fractal images, and the related information of public files, such as color, angle and length, is used to represent secret information.

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

General steganalysis method is also the main research direction in steganalysis field. Key feature extraction and classifier design are also two key steps in current general analysis. For example, avcibas and other researchers take Heduo as a key feature for steganalysis, and extract them to judge the initial data and encrypted data [37]; Farid proposed to predict the variation characteristics of DWT coefficients based on the training of a group of dense images and their original carrier images [38], and analyze the statistical characteristics of the prediction error matrix. This method can be regarded as borrowing the idea of classifier in pattern recognition. The classifier is constructed by extracting the difference features between the appropriate dense image and the carrier image. A large number of normal and embedded ciphertext images (embedded with different steganography algorithms) are used to train the classifier. The relevant parameters are continuously adjusted and used for detection after the training. Experimental results show that the higher the embedding rate of the encrypted image, the better the detection effect. The detection effect is related to the training image, parameter setting, feature selection and embedding rate.

ISSN: 2688-8645

After a period of development, the general detection method has made some achievements, but it has the following defects:

1. Detection accuracy. Because of its generality, the feature of a steganalysis algorithm is not as accurate as that of the special detection algorithm, so the detection accuracy of a steganalysis algorithm is not as high as that of the special detection algorithm;

2. The embedding rate has great influence. When the encrypted image is higher, the detection rate is higher, and when the embedding rate is low, the detection accuracy decreases obviously;

3.The practicability is not high. Due to the fact that most general detection algorithms need multiple eigenvalues to detect, the computation is large, the complexity is high, and the detection speed is limited.

Information Steganography technology can not only encrypt the secret information that needs to be transmitted, but also embed the ciphertext into the common public media for transmission. It has strong deception and higher security compared with traditional cryptography technology, which directly leads to the development of Information Steganography technology. Information steganalysis technology not only brings encryption convenience, but also brings certain risks to communication security. Based on the needs of national strategic security, business secrets and personal privacy protection, steganalysis technology has been widely studied.

Aiming at the requirement of steganalysis of JPEG image, an image steganalysis technology based on least squares support vector machine algorithm is proposed. The analysis framework of information steganalysis is introduced in detail. At the same time, the feature selection and classifier design are also described. Finally, the validity and accuracy of the algorithm are simulated in MATLAB.

4. Features of Audio Steganography

Audio information hiding has several remarkable characteristics. They are as follows:

First, the frequency domain span of audio signal is wide, and the frequency band of different types of audio is different. For example, the frequency of ordinary narrow-band voice signal is usually between 300Hz and 4000Hz; for broadband voice, the upper limit of frequency can reach 8000Hz; for audio forms such as songs, its high-frequency frequency can reach about 20000 Hz, which is about five times the frequency of ordinary narrow-band voice information. This feature has a great impact on the effect of audio information hiding algorithm, which makes a good algorithm in one environment may not be suitable for another environment;

Secondly, the energy distribution of audio signal, especially speech, is very uneven. Signal characteristics, such as pronunciation or pause, voiced or unvoiced, spoken language, speech characteristics, and even the gender and age of the speaker, have a great impact on energy distribution. Therefore, it may be necessary to select different strategies for different periods of speech to achieve the best effect.

Thirdly, the environment of audio signal transmission and processing is quite different. There are not only lossless propagation in high-speed streaming media, but also congestion and packet loss in VoIP (voice over Internet Protocol), bit rate error (BER) caused by noise and other factors in transmission line, amplitude / phase change caused by analog / digital conversion (ADC) or reverse conversion (DAC), etc. On the one hand, the problems brought by transmission environment enhance the concealment effect of secret information; at the same time, it will greatly increase the complexity of ciphertext extraction algorithm.

ISSN: 2688-8645

5. Expectation

Information steganalysis is a very complex system engineering problem, which needs to rely on the comprehensive application of various cross domain knowledge such as pattern recognition, probability statistical image and language processing. In this paper, steganalysis of image and speech information is studied, but there are still many problems to be solved

The steganalysis of image information is still worth studying;

1. Better reference image. The reference image and the Fridrich reference image are complementary to each other, and the average of the two may have a subtle effect;

2. RGB color image processing of different color channels. The difference between the results of color image detection and gray image detection is very large, which shows that there is still a correlation between color channel and gray image. It is of great significance to detect the correlation and reflect it into the feature vector to improve the steganalysis ability of color JPEG images;

The steganalysis of audio information is still worth studying;

1. Extraction of common features of additive steganography. In speech steganalysis, we only use the least bit or the lowest two bit correction to achieve steganography, without considering other possible steganography methods. If additive steganography is generally considered, is there any method that can bring better effect, that is, less effect on perceptual features than modifying the lowest order? Are the extracted features the best choice for analyzing additive steganography? Is the nature of additive steganography adequately reflected? After all, the quality of speech is "heard" by people, not "calculated". Is there more deceptive steganography?

Analysis of non additive steganography. Additive steganography can be regarded as a non speech small signal floating outside the carrier, which has nothing to do with the content of the carrier itself. Can we design a steganography algorithm that can make full use of the content of the carrier to improve the concealment degree? For example, is steganography like a small person hiding in the shadow of a big man, blocking the ear's sense of steganography as if the eyes were blocked by the big man and could not see the little man? If we analyze such a speech information hiding system, we should add those considerations.

References

[1] Feng Deng-Guo, Zhang Min, Li Hao, et al. Big data security and privacy protection. Chinese Joural of Computers, 2014, 37(1):246-258

[2] Petitcolas F A P, Anderson R J, Kuhn M G. Information hiding survey [J]. Proceedings of the IEEE, 1999, 87(7): 1062-1078.

[3] Li B, He J, Huang J, et al. A survey on image steganography and steganalysis. Journal of Information Hiding and Multimedia Signal Processing, 2011, 2(2): 142-172.

[4] YANG Wanxia, TANG Shanyu, LI Miaoqi, CHENG Yongfeng, ZHOU Zhili. Steganalysis of Low Embedding Rates LSB Speech Based on Histogram Moments in Frequency Domain[J]. Chinese Journal of Electronics, 2017,26(06):1254-1260.

[5] Xianyi Chen, Guangyong Gao, Dandan Liu, Zhihua Xia. Steganalysis of LSB Matching Using Characteristic Function Moment of Pixel Differences[J]. Chinese Communication, 2016, 13(07): 66-73.

[6] Feng Ruan, Xing Zhang, Dawei Zhu, Zhanyang Xu, Shaohua Wan, Lianyong Qi. Deep learning for real-time image steganalysis: a survey[J]. Journal of Real-Time Image Processing, 2020, 17(1).

[7] Murugeswari Kandavel, Deisy Chelliah, Ganesan Govindan. An optimised approach to detect the identity of hidden information in grey scale and colour images[J]. Int. J. of Business Intelligence and Data Mining,2019,15(1).

ISSN: 2688-8645

[8] Tomás Denemark, Mehdi Boroumand, Jessica J. Fridrich. Steganalysis Features for Content-Adaptive JPEG Steganography. [J]. IEEE Trans. Information Forensics and Security, 2016, 11(8).

[9] Tong Qiao, Florent Retraint, Rémi Cogranne, Cathel Zitzmann. Steganalysis of JSteg algorithm using hypothesis testing theory[J]. EURASIP Journal on Information Security, 2015, 2015(1).

[10] Osman Hilmi Koçal, Emrah Yürüklü, Ismail Avcibas. Chaotic-Type Features for Speech Steganalysis. [J]. IEEE Trans. Information Forensics and Security, 2008, 3(4).