INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 1685-1688

e-ISSN : 2583-1062

Impact Factor: 5.725

DATA HIDING USING STEGANOGRAPHY TOOLS

Dr. Asha S R Ph[1], Adithya R P[2], Gowtham Raj K P[3], Govind Shewale P[4], Kuber Raj T P[5]

[1]Professor of Dept of CSE, Sambhram Institute of Technology, Bengaluru Urban, Karnataka, India.

[2,3,4,5]Final Year Student Dept.Of CSE, Sambhram Institute of Technology, Bengaluru Urban, Karnataka, India.

DOIhttps://www.doi.org/10.58257/IJPREMS33607

ABSTRACT

The staggering growth in communication technology and usage of public domain channels (i.e. Internet) has greatly facilitated transfer of data. However, such open communication channels have greater vulnerability to security threats causing unauthorized information access. Traditionally, encryption is used to realize the communication security. However, important information is not protected once decoded. Steganography is the art and science of communicating in a way which hides the existence of the communication. Important information is firstly hidden in a host data, such as digital image, text, video or audio, etc, and then transmitted secretly to the receiver. Steganalysis is another important topic in information hiding which is the art of detecting the presence of steganography. This paper provides a critical review of steganography as well as to analyze the characteristics of various cover media namely image, text, a u dio and video in respects of the fundamental concepts, the progress of steganographic methods and the development of the corresponding steganalysis schemes.

1. INTRODUCTION

Steganography is the art and science of hiding information by embedding messages within other, seemingly harmless messages. Steganography means “covered writing” in Greek. As the goal of steganography is to hide the presence of a message and to create a covert channel, it can be seen as the complement of cryptography, whose goal is to hide the content of a message. Another form of information hiding is digital watermarking, which is the process that embeds data called a watermark, tag or label into a multimedia object such that watermark can be detected or extracted later to make an assertion about the object. The object may be an image, audio, video or text only. A famous illustration of steganography is Simmons’ Prisoners’ Problem [1].An assumption can be made based on this model is that if both the sender and receiver share some common secret information then the corresponding steganography protocol is known as then the secret key steganography where as pure steganography means that there is none prior information shared by sender and receiver. If the public key of the receiver is known to the sender, the steganographic protocol is called public key steganography [4], [9] and [8].For a more thorough knowledge of steganography methodology the reader may see [9], [24].Some Steganographic model with high security features has been presented in [28-33]. Almost all digital file formats can be used for steganography, but the image and audio files are more suitable because of their high degree of redundancy [24].

2. METHODOLOGY

AUDIO STEGANOGRAPHY METHODOLOGY

In audio steganography, secret message is embedded into digitized audio signal which result slight alteration of binary sequence of the corresponding audio file. Moreover, audio signals have a characteristic redundancy and unpredictable nature that make them ideal to be used as a cover for covert communications to hide secret messages [150].

Audio Steganography Algorithms: In this section, the four major audio steganography algorithms: Low-bit encoding, Phase encoding, Spread spectrum coding and Echo data hiding are described.

Low-bit Encoding: In Low-bit encoding (e.g., [113]), the binary version of the secret data message is substituted with the least significant bit (LSB) of each sample of the audio cover file. Though this method is simple and can be used to embed larger messages, the method cannot protect the hidden message from small modifications that can arise as a result of format.

Audio Steganalysis for High Complexity Audio Signals More recently: Liu et. al [129] propose the use of stream data mining for steganalysis of audio signals of high complexity. Their approach extracts the second order derivative based Markov transition probabilities and high frequency spectrum statistics as the features of the audio streams. The variations in the second order derivative based features are explored to distinguish between the cover and stego audio signals. This approach also uses the Mel-frequency cepstral coefficients [119], widely used in speech recognition, for audio steganalysis.

3. MODELING AND ANALYSIS

This method [149] presents an information-theoretic method for performing steganography and steganalysis using a statistical model of the cover medium. The methodology is general, and can be applied to virtually any type of media. It provides answers for some fundamental questions which have not been fully addressed by previous steganographic methods, such as how large a message can be hidden without risking detection by certain statistical methods, and how to achieve this maximum capacity. Current steganographic methods have been shown to be insecure against fairly simple statistical attacks. Using the model-based methodology, an example steganography method is proposed for JPEG images which achieves a higher embedding efficiency and message capacity than previous methods while remaining secure against first order statistical attacks.

DWT based Data Hiding: Wavelet-based steganography [55-60] is a new idea in the application of wavelets. However, the standard technique of storing in the least significant bits (LSB) of a pixel still applies. The only difference is that the information is stored in the wavelet coefficients of an image, instead of changing bits of the actual pixels. The idea is that storing in the least important coefficients of each 4 x 4 Haar transformed block will not perceptually degrade the image. While this thought process is inherent in most steganographic techniques, the difference here is that by storing information in the wavelet coefficients, the change in the intensities in images will be imperceptible.

IMAGE BASED STEGANALYSIS

Steganalysis is the science of detecting hidden information. The main objective of Steganalysis is to break steganography and the detection of stego image is the goal of steganalysis. Almost all steganalysis algorithms rely on the Steganographic algorithms introducing statistical differences between cover and stego image. Steganalysis deals with three important categories: (a) Visual attacks: In these types of attacks with a assistance of a computer or through inspection with a naked eye it reveal the presence of hidden information, which helps to separate the image into bit planes for further more analysis. (b) Statistical attacks: These types of attacks are more powerful and successful, because they reveal the smallest alterations in an images statistical behavior. Statistical attacks can be further divided into (i) Passive attack and (ii) Active attack. Passive attacks involves with identifying presence or absence of a covert message or embedding algorithm used etc. Mean while active attacks is used to investigate embedded message length or hidden message location or secret key used in embedding. (c) Structural attacks: The format of the data files changes as the data to be hidden is embedded; identifying this characteristic structure changes can help us to find the presence of image.

TEXT BASED STEGANALYSIS

The usage of text media, as a cover channel for secret communication, has drawn more attention [95]. This attention in turn creates increasing concerns on text steganalysis. At present, it is harder to find secret messages in texts compared with other types of multimedia files, such as image, video and audio [96-101]. In general, text steganalysis exploits the fact that embedding information usually changes some statistical properties of stego texts; therefore it is vital to perceive the modifications of stego texts. Previous work on text steganalysis could be roughly classified into three categories: format- based [102, 103], invisible character-based [104-106] and linguistics, respectively. Different from the former two categories, linguistic steganalysis attempts to detect covert messages in natural language texts. In the case of linguistic steganography, lexical, syntactic, or semantic properties of texts are manipulated to conceal information while their meanings are preserved as much as possible[109].Due to the diversity of syntax and the polysemia of semantics in natural language, it is difficult to observe the alterations in stego texts. So far, many linguistic steganalysis methods have been proposed. In these methods, special features are designed to extend semantic or syntactical changes of stego texts. For example , Z.L. Chen[108] et al. designed the N- window mutual information matrix as the detection feature to detect semantic steganagraphy algorithms. Furthermore, they used the word entropy and the change of the word location as the semantic features [109,110], which improved the detection rates of their methods. Similarly, C.M. Taskiran et al [111] used the probabilistic context-free grammar to design the special features in order to attack on syntax steganography algorithms. In the work mentioned above, designed features strongly affect the final performances and they can merely reveal local properties of texts. Consequently, when the size of a text is large enough, differences between Natural texts (NTs) and Stego texts (STs) are evident, thus the detection performances of the mentioned methods are acceptable. Whereas, when the sizes of texts become small, the detection rates decrease dramatically and can not be satisfied for applications. In addition, some steganographic tools have been improved in the aspects of semantic and syntax for better camouflage [112]. Therefore, linguistic steganalysis still needs further research to resolve these problems. Some more work on Text Steganalysis has been discussed below.

4. USE OF STASTICAL DISTANCE MEASURES FOR AUDIO STEGANALYSIS

H. Ozer et. al [122] calculated the distribution of various statistical distance measures on cover audio signals and stego audio signals vis--vis their versions without noise and observed them to be statistically different. The authors employed audio quality metrics to capture the anomalies in the signal introduced by the embedded data. They designed an audio steganalyzer that relied on the choice of audio quality measures, which were tested depending on their perceptual or non- perceptual nature.

The selection of the proper features and quality measures was conducted using the (i) ANOVA test [123] to determine whether there are any statistically significant differences between available conditions and the (ii) SFS (Sequential Floating Search) algorithm that considers the inter-correlation between the test features in ensemble [124]. Subsequently, two classifiers, one based on linear regression and another based on support vector machines were used and also simultaneously evaluated for their capability to detect stego messages embedded in the audio signals. The features selected using the SFS test and evaluated using the support vector machines produced the best outcome. The perceptual- domain measures considered in [122] are: Bark Spectral Distortion, Modified Bark Spectral Distortion, Enhanced Modified Bark Spectral Distortion, Perceptual Speech Quality Measure and Perceptual Audio Quality Measure. The non-perceptual time-domain measures considered are: Signal-to-Noise Ratio, Segmental Signal-to-Noise Ratio and Czenakowski Distance. The non-perceptual frequency- domain measures considered are: Log-Likelihood Ratio, Log-Area Ratio, Itakura- Satio Distance, Cepstral Distance, Short Time Fourier Random Transform Distance, Spectral Phase Distortion and Spectral Phase Magnitude Distortion.

A. Audio Steganalysis based on Hausdorff Distance

The audio steganalysis algorithm proposed by Liu et. Al [125] uses the Hausdorff distance measure [126] to measure the distortion between a cover audio signal and a stego audio signal. The algorithm takes as input a potentially stego audio signal x and its de-noised version x as an estimate of the cover signal. Both x and x are then subjected to appropriate segmentation and wavelet decomposition to generate wavelet coefficients

[129] at different levels of resolution. The Haus- dorff distance values between the wavelet coefficients of the audio signals and their de-noised versions are measured. The statistical moments of the Hausdorff distance measures are used to train a classifier on the difference between cover audio signals and stego audio signals with different content loadings. However, the above approach of creating a reference signal via its own de- noised version causes content-dependent distortion.

This can lead to a situation where the variations in the signal content itself can eclipse the classifier from detecting the distortions induced during data hiding. In [128], Avcibas proposed an audio steganalysis technique based on content- independent distortion measures. The technique uses a single reference signal that is common to all the signals to be tested.

B. Audio Steganalysis for High Complexity Audio Signals More recently, Liu et. al [129] propose the use of stream data mining for steganalysis of audio signals of high complexity. Their approach extracts the second order derivative based Markov transition probabilities and high frequency spectrum statistics as the features of the audio streams. The variations in the second order derivative based features are explored to distinguish between the cover and stego audio signals. This approach also uses the Mel-frequency cepstral coefficients [119], widely used in speech recognition, for audio steganalysis.

5. CONCLUSION

In conclusion, the utilization of steganography tools for data hiding presents a multifaceted realm with significant implications across various domains. Through the amalgamation of sophisticated algorithms and innovative simulation software, researchers and practitioners have unlocked avenues for concealing sensitive information within innocuous cover objects. The journey towards effective data hiding involves algorithm development, rigorous testing, and continuous refinement facilitated by simulation software.

This software serves as a crucible wherein steganographic techniques are forged, evaluated, and optimized for diverse applications. Moreover, the efficacy of steganography in safeguarding data hinges upon its ability to balance concealment with imperceptibility, robustness, and computational efficiency. Simulation software acts as a conduit for exploring this delicate equilibrium, enabling researchers to navigate the intricate landscape of trade-offs inherent in data hiding. In essence, the symbiotic relationship between steganography tools and simulation software underscores their pivotal role in the realm of data hiding. As technology continues to evolve, so too will the capabilities of these tools, propelling the field of steganography towards ever-greater heights of innovation and sophistication.

6. REFERENCES

[1] Gustavus J. Simmons, "The Prisoners' Problem and the Subliminal Channel", in Proceedings of CRYPTO '83, pp 51- 69. Plenum Press (1984).

[2] P. Wayner, “Strong Theoretical Steganography”, Cryptologia, XIX(3), July 1995, pp. 285-299.

[3] J.T. Brassil, S. Low, N.F. Maxemchuk, and L. O’Gorman, "Electronic Marking and Identification Techniques to Discourage Document Copying", IEEE Journal on Selected Areas in Communications, vol. 13, Issue. 8, October 1995, pp. 1495-1504.

[4] “Stretching the Limits of Steganography", RJ Anderson, in Information Hiding, Springer Lecture Notes in Computer Science v 1194 (1996) pp 39-48.

[5] Kahn, The Codebreakers - the comprehensive history of secret communication from ancient times to the Internet, Scribner, New York (1996).

[6] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding", IBM Systems Journal, vol. 35, Issues 3&4, 1996, pp. 313-336.

[7] Scott Craver, "On Public-key Steganography in the Presence of an Active Warden," in Proceedings of 2nd International Workshop on Information Hiding, April 1998, Portland, Oregon, USA. pp. 355 - 368.

[8] Ross J. Anderson and Fabien A.P. Petitcolas, "On the limits of steganography," IEEE Journal on Selected Areas in Communications (J-SAC), Special Issue on Copyright & Privacy Protection, vol. 16 no. 4, pp 494-481, May 1998.

[9] N. F. Johnson and S. Jajodia, "Steganography: seeing the unseen," IEEE Computer.,Feb., 26-34 (1998).

[10] L. M. Marvel, C. G. Boncelet, Jr. and C. T. Retter, "Spread spectrum image steganography," IEEE Trans. on Image Processing, 8(8), 1095-1083 (1999).

[11] Digital Watermarking :A Tutorial Review S.P.Mohanty ,1999.

[12] J. Shi and J. Malik, "Normalized cuts and image segmentation.,"IEEE Trans. PAMI, vol. 22, no. 8, pp. 888-905,2000.

[13] Y. K. Lee and L. H. Chen, "High capacity image steganographic model," IEE Proc.-Vision, Image and Signal Processing, 149(3), 288-294 (2000).

[14] Analysis of LSB Based Image Steganography Techniques ,R. Chandramouli, Nasir Memon, Proc. IEEE ICIP, 2001.

[15] M.Chapman, G. Davida, and M. Rennhard, “A Practical and Effective Approach to Large-Scale Automated Linguistic Steganography”, Proceedings of the Information Security Conference, October 2001, pp. 156-165.

[16] An Evaluation of Image Based Steganography Methods,Kevin Curran, Kran Bailey, International Journal of Digital Evidence,Fall 2003.

[17] G. Doërr and J.L. Dugelay, "A Guide Tour of Video Watermarking", Signal Processing: Image Communication, vol. 18, Issue 4, 2003, pp. 263-282.

[18] K. Gopalan, "Audio steganography using bit modification", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03),