LOCATING INFORMATION-HIDING IN MP3 AUDIO

Mengyu Qiao, Andrew H. Sung
Department of Computer Science and Engineering, Institute for Complex Additive Systems Analysis
New Mexico Tech, Socorro, NM, U.S.A.

Qingzhong Liu
Department of Computer Science, Sam Houston State University, Huntsville, TX, U.S.A.

Bernardete M. Ribeiro
Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal

Keywords: Steganalysis, Steganography, MP3, Neural-fuzzy inference systems.

Abstract: Steganography provides a stealthy communication channel for malicious users, which jeopardizes traditional cyber security infrastructure. Due to the good quality and the small storage usage, compressed audio has been widely employed by online audio sharing, audio streaming broadcast, and voice over IP, etc. Several audio steganographic systems have been developed and published on Internet. Traditional blind steganalysis methods detect the existence of information hiding, but neglect the size and the location of hidden data. In this paper, we present a scheme to locate the modified segments in compressed audio streams based on signal analysis in MDCT transform domain. We create reference signals by reversing and repeating quantification process, and compare the statistical differences between source signals and reference signals. Dynamic evolving neural-fuzzy inference systems are applied to predict the number of modified frames. Finally, the frames of audio streams are ranked according to their modification density, and the top ranked frames are selected as candidate information-hiding locations.

1 INTRODUCTION

Steganography is the technique that embeds secret messages in innocent digital media, such as images, audios, videos and other files, without attracting listeners' or viewers' notice. The innocuous digital media or files are called carriers or covers, the media embedded with hidden data are called steganograms. The sender and receiver use the same steganographic tool to embed and extract the secret messages, and share cryptographic algorithm and key for encryption and decryption. Recent media reports (Web 1) and court documents released by the U.S. Department of Justice (Web 2, 3) have confirmed the use of steganography for espionage.

To detect the information-hiding in digital audio, linear prediction coding was presented to generate reference signal for error estimation (Ru et al., 2005). Mel-cepstrum based analysis was adopted to perform the detection of hidden data in uncompressed audio (Kraetzer et al., 2008). Temporal derivative based approach was employed to detect information-hiding of multiple steganographic systems (Liu et al., 2009a, b). Accumulative moment statistical features, and Generalized Gaussian Density (Qiao et al, 2009a, b) were introduced to detect information-hiding in MP3 audios.

Although several steganalysis methods were designed for detecting information-hiding in the past years, the locations of information-hiding in audio streams have been barely studied. As a qualitative measurement, blind steganalysis only provides a decision of the existence of information-hiding or not, but neglects the detailed analysis of the locations of hidden data in steganograms. In practice, extracting and decrypting hidden data are more useful to reveal covert communication for illegal purposes. Therefore, locating the embedding

Qiao M., Sung A., Liu Q. and Ribeiro B..
LOCATING INFORMATION-HIDING IN MP3 AUDIO.
DOI: 10.5220/0003296205040507
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 504-507
ISBN: 978-989-8425-40-9

positions will benefit the extraction and the destruction of covert communication.

In this paper, we design an approach for targeting embedding locations of MP3 compressed audio. We construct reference signals by reversing and repeating quantification process, and extract statistical features from covers and steganograms in MDCT transform domain. Dynamic evolving neural-fuzzy inference systems (DENFIS) are used to predict embedding strength of MP3 steganograms. By ranking and thresholding modification densities of audio frames, candidate frames with potential information-hiding are selected for information extraction and decryption.

The following parts are organized in this way: Section 2 briefly explain the embedding algorithm of MP3Stego and reference signal construction. Section 3 describes our method of feature extraction and selection information-hiding locations. Section 4 presents experiment design, followed by conclusion in section 5.

2 MP3 STEGANOGRAPHY AND REFERENCE SIGNAL

MP3Stego is one of the most widely used audio steganographic tools, which is implemented by combining the novel information-hiding algorithm with existing MP3 encoder. All the payloads are encrypted using 3DES, and then embedded in frames which are randomly selected by using SHA-1. The algorithm of MP3Stego exploits the audio degradation from lossy compression and embeds data by slightly expanding the distortion of the signal without attracting listeners' notice.

To overcome this effect, we design a reference based approach, which reduce the individual characteristics of signals by extracting the features from the errors between signal and its reference.

For long widow, there are 576 MDCT coefficients in one frame. For short window, three consecutive groups of 192 coefficients are combined into one frame. For the audio signal with N frames, we define the quantized MDCT coefficients as a matrix:

$$IX = \begin{pmatrix} ix_{0,0} & \cdots & ix_{0,575} \\ \vdots & \ddots & \vdots \\ ix_{N-1,0} & \cdots & ix_{N-1,575} \end{pmatrix} \tag{1}$$

To generate the reference signal, we first obtain the de-quantized MDCT coefficients by using equation (2).

$$\begin{gathered} xr'_{i,j} = sign(ix_{i,j}) \times ix_{i,j}^{4/3} \times \sqrt[4]{2}^{stepsize_i} \\ i = 0 \sim N-1, j = 0 \sim 575 \end{gathered} \tag{2}$$

where *stepsize* is the original scaling information in the MP3 file.

According to MP3 encoding standard, we calculate the suitable step-size for *xr'*, denoted as *stepsize'*. For covers, the new step-size should equivalent to the original one in the MP3 file. For steganograms, the original step-size differs from the new step-size, since it was tampered by the information-hiding behavior.

$$stepsize' = system\_const \times \ln sfm \tag{3}$$

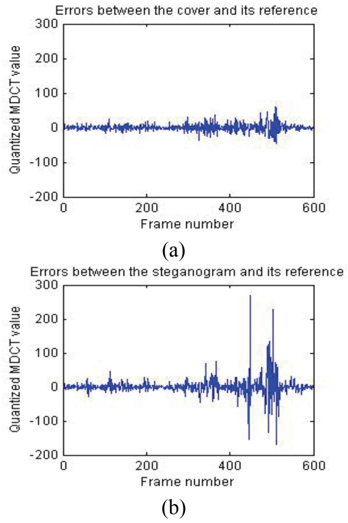
$$sfm_i = \frac{e^{\frac{1}{576}\left(\sum_{j=0}^{575} \ln xr'^{2}_{i,j}\right)}}{\frac{1}{576}\left(\sum_{j=0}^{575} \ln xr'^{2}_{i,j}\right)} \tag{4}$$

where *sfm* is the spectral flatness measure, and *system_const* is set to 8.0.

Then we repeat the quantization process to obtain quantized MDCT coefficients of the reference signal by using *stepsize'*. NINT is a function that returns the closest integer value.

$$ix'_{i,j} = NINT\left(\left(\frac{\left|xr'_{i,j}\right|}{\sqrt[4]{2}^{stepsize'}}\right)^{\frac{3}{4}} - 0.0946\right) \tag{5}$$

In the quantization step, the non-uniform quantization is caused by using power 0.75 in equation (5). If power 1 is used instead of 0.75, it will become uniform quantization.



(a)

(b)

Figure 1: Errors between the cover and its reference (a), errors between the steganogram and its reference (b).

3 PREDICTION OF EMBEDDING LOCATIONS

In order to accurately predict the strengths and the locations of information-hiding, we design our method as a two-step approach. In the first step, we extract statistical features from entire audio streams for embedding strength estimation. In the second step, same features are extracted from individual frames for embedding location prediction.

3.1 Feature Extraction

The stream features *S1-S5* are designed to extract the statistical features from entire audio streams. *ix* and *ix'* denote quantized MDCT coefficients from the original and its reference signal respectively. The value of *i* is the number of frames in the audio, and the value of *j* is the number of sub-band from 0 to 575.

$$S1=\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{575}\left|ix_{i,j}-ix'_{i,j}\right|}{\sum_{i=0}^{N-1}\sum_{j=0}^{575}\left|ix'_{i,j}\right|} \tag{6}$$

$$S2(j)=\begin{cases}\dfrac{\sum_{i=0}^{N-1}\left|ix_{i,j}-ix'_{i,j}\right|}{\sum_{i=0}^{N-1}\left|ix'_{i,j}\right|} & \text{if } \sum_{i=0}^{N-1}\left|ix'_{i,j}\right|\neq 0\\ 0 & \text{if } \sum_{i=0}^{N-1}\left|ix'_{i,j}\right|=0\end{cases} \tag{7}$$

$$d(i,j)=\begin{cases}\dfrac{\left|ix_{i,j}-ix'_{i,j}\right|}{\left|ix'_{i,j}\right|} & \text{if } \left|ix'_{i,j}\right|\neq 0\\ 0 & \text{if } \left|ix'_{i,j}\right|=0\end{cases} \tag{8}$$

$$S3(j)=\sum_{i=0}^{N-1}d(i,j) \tag{9}$$

Transition probabilities have been employed for image steganalysis (Shi et al., 2007). Since differences between the original signal and its reference signal reflect the distortion introduced by information-hiding, we design transition probability features to measure the similarity between them. δ = 1 if its arguments are satisfied, otherwise δ = 0; x and y are integers in the range of [-4, +4].

$$S4(x,y)=\sum_{j=0}^{575}\frac{\sum_{i=0}^{N}\delta\left(ix_{i,j}=x,ix'_{i,j}=y\right)}{\sum_{i=0}^{N}\delta\left(ix_{i,j}=x\right)} \tag{10}$$

$$S5(x,y)=\frac{\sum_{i=0}^{N}\sum_{j=0}^{575}\delta\left(ix_{i,j}=x,ix'_{i,j}=y\right)}{576N} \tag{11}$$

To location the information-hiding at frame level, we extract frame features *F1-F3* from individual frames of the original signal and its reference signal.

$$F1(i)=\frac{\sum_{j=0}^{575}\left|ix_{i,j}-ix'_{i,j}\right|}{\sum_{j=0}^{575}\left|ix'_{i,j}\right|} \tag{12}$$

$$F2(i)(x,y)=\frac{\sum_{j=0}^{575}\delta\left(ix_{i,j}=x,ix'_{i,j}=y\right)}{\sum_{j=0}^{575}\delta\left(ix_{i,j}=x\right)} \tag{13}$$

$$F3(i)(x,y)=\frac{\sum_{j=0}^{575}\delta\left(ix_{i,j}=x,ix'_{i,j}=y\right)}{576} \tag{14}$$

3.2 Embedding Location Searching

In order to locate the candidate embedding locations in audio streams, we need to estimate the embedding strength of the steganogram before ranking the potential locations. For estimation, we apply a dynamic evolving neural-fuzzy inference system (Kasabov et al., 2002) to stream features. The system will provide a ratio of modified coefficients to all coefficients denoted by *P%*, which also indicate the portion of frames affected by information-hiding. Another DENFIS system is employed to estimate the probability of information-hiding in individual frames. Then, frames are ranked according to their embedding probabilities, and the top *P%* frames of the rank are selected as candidate embedding locations.

4 EXPERIMENT SETTINGS

Our source audio data sets are 10000 44.1 kHz 16 bit PCM coded WAV audio files including digital speeches and songs in several languages, for instance, English, Chinese, Japanese, Korean, and several types of music (jazz, rock, blue), etc. The 10000 cover MP3 audios are compressed by using MP3Stego without any hidden date. We created 10000 steganograms by using MP3Stego with random selected embedding strength less than the maximum embedding capacity. The payloads include voice, video, image, text, executable codes, etc., and each steganogram carries a unique payload. All MP3 audio files are obtained in 128 kbps bit-rate and 44.1 kHz sample-rate. In the experiment, two DENFIS systems will be trained and tested separately, which provide embedding strength estimation for audio streams and embedding probabilities for individual frames. The prediction accuracy will be used for evaluation.

5 CONCLUSIONS

In this paper we propose a method to predict the embedding locations of MP3 steganography, which is a critical step towards extracting hidden data from steganogram. We design a reference based approach to extract stream and frame features of the quantized MDCT coefficients from audio streams and audio frames, and dynamic evolving neural-fuzzy inference systems are applied to the features for the prediction.

Since we have a feature set with large dimension, non-discriminative features will compromise the prediction efficiency and accuracy. Therefore, elaborate feature selection methods will contribute to the enhancement of prediction.

ACKNOWLEDGEMENTS

We gratefully appreciate the Institute for Complex Additive Systems Analysis of New Mexico Tech for supporting this study.

REFERENCES

http://www.msnbc.msn.com/id/38028696/ns/technology_and_science-science/

http://www.justice.gov/opa/documents/062810complaint1.pdf

http://www.justice.gov/opa/documents/062810complaint2.pdf

Ru, X., Zhang, H. and Huang, X. (2005). Steganalysis of Audio: Attaching the Steghide. *Proc. the Fourth International Conference on Machine Learning and Cybernetics*, 3937-3942.

Kraetzer, C. and Dittmann, J. (2008). “Pros and Cons of Mel-cepstrum Based Audio Steganalysis Using SVM Classification. *Lecture Notes in Computer Science*, 4567, 359-377.

Liu, Q., Sung, A., and Qiao, M. (2009a). Temporal Derivative-based Spectrum and Mel-cepstrum Audio Steganalysis. *IEEE Trans. on Info. Forensics and Security*, 4(3), 359-368.

Liu, Q., Sung, A., and Qiao, M. (2009b). Novel Stream Mining for Audio Steganalysis. *Proc. of 17th ACM International Conference on Multimedia*, 95-104.

Qiao, M., Sung, A., and Liu, Q. (2009a). Feature Mining and Intelligent Computing for MP3 Steganalysis. *Proc. of International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*, 627-630.

Qiao, M., Sung, A., and Liu, Q. (2009b). Steganalysis of MP3Stego. *Proc. of 22nd International Joint Conference on Neural Networks*, 2566-2571.

Shi, Y., Chen, C., and Chen, W. (2007). A Markov Process Based Approach to Effective Attacking JPEG Steganography. *Lecture Notes in Computer Sciences*, 437, 249-264.

Kasabov, N. and Song, Q. (2002). DENFIS: Dynamic Evolving Neural-fuzzy Inference System and Its Application for Time-series. *IEEE Trans. Fuzzy Systems*, 10(2), 144-154.