ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse240942020.pdf
https://doi.org/10.30534/ijatcse/2020/240942020

Detecting Audio Steganography using Machine Learning

Lai Van Duong[1], Tisenko Victor Nikolaevich[2],Nguyen QuocHoang [3], Pham ThiThuong[4], Dong Xuan Anh[5]
[1,3,4,5]Information Assurance dept. FPT University, Hanoi, Vietnam, duonglvse05009@fpt.edu.vn, HoangNQSE06012@fpt.edu.vn, ThuongPTSE05856@fpt.edu.vn, AnhDXSE06086@fpt.edu.vn,
[2]Department Quality Systems, Peter the Great St. Petersburg Polytechnic University, Russia, St.Petersburg, Polytechnicheskaya, 29, v_tisenko@mail.ru

ABSTRACT

In recent years, steganography techniques are rapidly developing. In addition to the outstanding advantages of the ability to hide and transmit secret information, it has a huge disadvantage that is being easily exploited by hackers. This poses increasing and serious threats and challenges to cyber security. Audio steganography is one of the most difficult techniques to detect today. Traditional methods of detecting steganography can only detect individual audio steganography techniques. In this paper, we propose a method to detect many audio steganography techniques using machine learning.

Key words: Audio steganography, machine learning, audio steganography detection, abnormal behavior, Random Forest, SVM, Feature selection.

1. INTRODUCTION

STEGANOGRAPHY is a technique of hiding important information into digital data without compromising the intuition and the original quality of digital data. Audio steganography is a technique of hiding or embedding information in a contain environment which are audio files. In particular, the audio that doesn't contain hidden information is called the original audio or cover audio, and the audio that contains hidden information is called stego-audio. This technique hides information into the gaps of audio such as frequency, wavelength, cycle and amplitude, propagation speed, etc. Audio steganography includes the main techniques: LSB, spread spectrum, phase, echo, steghide. In addition to the outstanding advantages of the ability to hide and transmit secret information, it has a huge disadvantage that is being easily exploited by hackers. Therefore, the detection of audio steganography is very necessary.

Detecting audio steganography is a technique to detect the existence of hidden information in audio. The purpose is to detect a medium that carries information, try to retrieve that confidential information, or lost its integrity. If its embedding method and statistical models are known, the optimum detector can be built. However, this information is not often available. Therefore, the steganography detection system is usually built based on machine learning techniques. Due to the flow nature, popularity, and widespread use of audio signals, they become good targets for steganography. Like the cryptographic technique, if steganography in multimedia in general, in digital audio, in particular, is a complicated issue, the detection of audio steganography is more difficult and complicated. However, no matter how sophisticated the steganography algorithm is, it still reveals weaknesses. Those weaknesses are the basis for building detection algorithms. When a certain amount of information is hidden in the audio, they all leave certain traces and changes. These traces may not be perceived by the human ear but they can be detected by modern mathematical techniques. Based on these traces, we can conclude whether the audio contains hidden data or not. The detection is facilitated by the conventional statistical hypothesis testing method if given a pair of original audio files (cover files) and audio files containing corresponding hidden information (stego files). In another special case, if given arbitrary audio and steganography algorithm, determine whether the audio contains hidden information or not. With this case, the available detection algorithm is relatively effective (with an accuracy of over 85%).

In fact, we often encounter the following problem: Given an audio file, determine whether the audio file contains hidden information or not. By extension, given N arbitrary audio files, determine how many audio files contain hidden information. Thus, the problem is to build an algorithm to classify N audio files into two classes: one class contains audio files containing hidden information (stego files), the other class contains the original audio files (cover files). To solve the above problem, in this paper, we will select and extract the features of audio files in different domains including frequency domain, time domain, and perceptual domain. Then, we will use the appropriate machine learning algorithm to classify it, thus detecting the stego files.

2. RELATED WORKS

In the document [1], X.-M. Ru et al. proposed using linear prediction code (LPC) and SVM to detect the Steghide steganography technique. LPC is used to extract the correlation between neighbor samples. S. Rekik et al. [2] used the autoregressive time-delay neural network (AR-TDNN) to detect audio steganography for both LSB and DWT algorithms. The articles [3], [4] proposed using Mel-frequency cepstrum coefficients (MFCC) as features, namely ordinary MFCC, wavelet-based MFC Cs, and derivative-based MFCC. Document [5] combines MFCCs, different moments of spectral, audio quality metrics, LPC residue and SVM algorithms. Ozer et al. [6] used wavelet-thresholding to reduce noise and estimate the cover, then used the AQM metrics to evaluate. Finally, those AQM metrics will be used as features to put into machine learning models to detect audio steganography. S. Geetha et al. [7] proposed combining the Hausdorff distance and Decision Tree algorithm. Z. Kexin et al. [8] used the Gaussian Mixture Model and the Generalized Gaussian Distribution to directly compare the wavelet coefficients distribution of cover files and stego files. To detect Phase and Echo steganography techniques, a common method is combining the SVM algorithm with the analysis of the information of frequency domain [9], [10].

In summary, most of the previous studies proposed combining the SVM algorithm with the feature group of each individual domain to detect audio steganography. In this paper, we propose using the Random Forest algorithm and features of all 3 domains: frequency domain, time domain, and perceptual domain. Accordingly, the list of abnormal features in the audio file structure in 3 domains is described in section 3.2 of the paper. The Random Forest algorithm for classifying cover and stego files is presented in section 3.3. The experiment section evaluates the effectiveness of each classification method described in section 4 of the paper

3. A method of detecting audio steganography using machine learning

3.1 *Proposed Model*

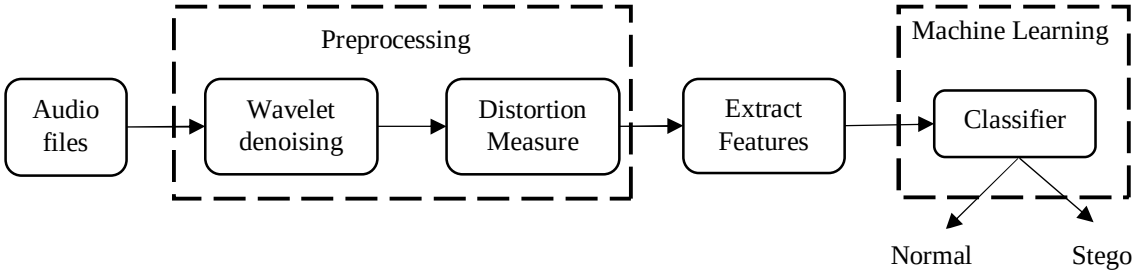

Figure 1: Model of detecting audio steganography using machine learning

Figure 1 presents the proposed audio steganography detection system using machine learning. The model consists of the following components:

1. Audio files: includes cover audio files and stego audio files.
2. Preprocessing: including Noise Reduction and Distortion Measure. Noise Reduction restores the characteristics of the original audio file and removes as much noise as possible. In this paper, the wavelet transform is used to reduce noise. With Distortion Measure, Hausdorff distance measurement will be applied to measure the distortion or reduction of the original audio signal.
3. Extract Features:
4. The features are extracted from each audio file. Details of these features will be presented in the next section of this paper.
5. Training: Random Forest algorithm will be used to classify from those features

3.2. Feature selection

Each audio file is extracted to 23 features, with the cover signal $x(i), i = 1,...N$ and the stego signal $y(i), i = 1,...N$.

In this section, features are listed by 3 groups consisting (in table 1) of feature group on the time domain, frequency domain, and perceptual domain.

Table 1: List of features

Feature Group	No.	Feature Name	Description
Feature group on	1	LLR	Log-Likelihood Ratio: $LLR = log\frac{a_x^T R_x a_x}{a_y^T R_y a_y}$. Where: a_x is LPC coefficient vector for the cover signal $x[n]$, a_y is corresponding vector for stego

frequency domain			signaly$[n]$ with the corresponding variance matrix, R_x and R_y.
	2	LAR	Log Area Ratio: is another LPC-based technique, using PartialCorrelation (PARCOR) coefficients. PARCOR coefficients constitute a set of parameters derived from the short-term LPC representation of the tested speech signal.
	3	ISD	Itakura-Saito distance: is the difference between the power spectrum of the stego signal Y and the cover signal$X$:$IS = \int_{-\pi}^{\pi} \left(log \frac{Y(w)}{X(w)} + \frac{X(w)}{Y(w)} - 1 \right) \frac{dw}{2\pi}$.
	4	ID	ItakuraDistance: is a variation of ISD.
	5	COSH	COSH Distance: is a symmetric measurement of the Itakura-Saito distance. Here, the overall measurement is calculated by averaging COSH values across segments: $$COSH = \int_{-\pi}^{\pi} \left[\frac{1}{2} \left(\frac{Y(w)}{X(w)} + \frac{X(w)}{Y(w)} \right) - 1 \right] \frac{dw}{2\pi}$$
	6	CD	Ceptral Distance: is a distance, defined by the cepstral coefficients of the cover signal X and stego signal Y: $$CD = \frac{\sum_{m=1}^{M} w(m) d(c_x, c_y, m)}{\sum_{m=1}^{M} w(m)}$$ Where: M is the total number of frames and $w(m)$ is the weight relative to the m-th frame. Weight is the power in the reference frame.
	7	CDM	Cepstral Distance Measure coefficients can be calculated using the LPC parameters. A measure of audio quality based on cepstralcoefficients$c_x(k)$ và $c_y(k)$ of the cover andstego signals, can be calculated: $$d(c_x, c_y, m) = \left[\left[c_x(0) - c_y(0) \right]^2 + 2 \sum_{k=1}^{L} \left[c_x(k) - c_y(k) \right]^2 \right]^2$$
	8	STFT	Short-Time Fourier-Radon Transform Measure: determines the mean-square distance of Radon transforms of the STFT of two signals.
	9	SP	Spectral Phase Distortions: The phase deviation and the spectral level were observed: $$SP = \frac{1}{N} \sum_{w=1}^{N} \left| \theta_x(w) - \theta_y(w) \right|^2$$ Where: $\theta_x(w)$ is the phase spectrum of the cover signal X, $\theta_y(w)$ is the phase of the stego signal Y.
	10	SPM	Spectral Phase-Magnitude Distortionswith $\lambda$is the option to attach corresponding weights to the phase and magnitude terms: $$SPM = \frac{1}{N} \left(\lambda * \sum_{w=1}^{N} \left| \theta_x(w) - \theta_y(w) \right|^2 + (1 - \lambda) * \sum_{w=1}^{N} \left| |X(w)| - |Y(w)| \right|^2 \right)$$ Where: $\theta_x(w)$ is the phase spectrum of the cover signal X, $\theta_y(w)$ is the phase of the stego signal Y, $X(w)$ is the magnitude spectrum of the cover signal, and $Y(w)$ is the magnitude spectrum of the stego signal.

Feature group on time domain	11	SNR	Signal-to-Noise Ratio:$SNR = 10\log_{10}\frac{\sum_{i=1}^{N}x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2}$ Where:$x(i)$is the cover audio signal, $y(i)$ is the stegoaudio signal.
	12	SNRseg	Segmental Signal-to-Noise Ratio: is defined as the average of SNR values over short segments: $$SNRseg = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\sum_{i=Nm}^{Nm+N-1}\left(\frac{x^2(i)}{(x(i)-y(i))^2}\right)$$
	13	CZD	Czenakowski Distance: directly compare the sample vectors in the time domain.$C = \frac{1}{N}\sum_{i=1}^{N}1-\left(\frac{2*\min(x(i),y(i))}{x(i)+y(i)}\right)$
Feature group on perceptual domain	14	BSD	Bark Spectral Distortion: $BSD = \sum_{i=1}^{K}\left[S_x(\mathrm{i}) - S_y(\mathrm{i})\right]^2$ Where: *K* is the number of critical bands, and $S_x(\mathrm{i})$ and $S_y(\mathrm{i})$ are the Bark spectrum in the critical band*i*.
	15	MBD	Modified Bark Spectral Distortion is a modified version of BSD, incorporating noise maskingthreshold to distinguish between audio distortions: $$MBSD = \sum_{i=1}^{K}M(i)D_{xy}(i)$$ Where:$M(i)$ and $D_{xy}(i)$ represent the indicator of perceptible distortionand the volume difference in the *i-th* critical band, *K* is the number of critical bands.
	16	SPD	Spectral PhaseDistortion feature will receive phase noise due to embedding information
	17	EMD	Enhanced Modified Bark Spectral Distortion: is a variation of MBSD spectrum, used to calculate the volume difference. $$MBSD = \sum_{i=1}^{15}M(i)D_{xy}(i)$$
	18	PAM	Perceptual Audio Quality Measure: Optimized for the auditory system of humans.
	19	PSM	Perceptual Speech Quality Measure: is a version of PAQM. Optimized for the human auditory system for speech.
	20	MN1	Measuring Normalizing Blocks1: Based on the perceptual module to estimate speech error (calculated through different time-frequency structures).
	21	MN2	Measuring Normalizing Blocks2

	22	WSD	Weighted Slope Spectral Distance Measure: with $\{X(k), Y(k)\}$ are the spectra in decibels: $$WSSD = \sum_{k=1}^{36} w(k)\{[X(k+1) - X(k)] - [Y(k+1) - Y(k)]\}^2$$
	23	MNB	MNB describes the important role of cognitive modules to estimate speech quality.

3.3. Algorithm selection

In the field of detection of malicious applications on android can use machine learning algorithms such as [11]: Decision trees, Naive Bayes, Support Vector Machine (SVM)... To detect audio steganography, in this paper, we use two Random Forest and SVM algorithms. Random Forest is an ensemble classification method [12]. This algorithm is the result of the ensemble of classifiers, which normally are Decision Trees to make the final prediction. The theoretical foundation of this algorithm is based on Jensen's inequality [13]. Accordingly, in the classification problems, the combination of many models may produce less error rate than that of each individual model. The studies [14], have proven the Random Forest algorithm has many advantages than other machine learning algorithms. In this paper, we use the Random Forest algorithm with the number of decision trees of 10 in order to classify. SVM is the supervised learning algorithm used for data classification [15]. SVM algorithm constructs a hyperplane or a set of hyperplanes in a multi-dimensional or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, for the best classification, the hyperplane is as far away from the data points of all classes (called the margin function), because the larger the margin the lower the generalization error of the classification algorithm [15].The SVM method requires data to be expressed as vectors of real numbers. Thus, if the input is not a number, it is necessary to find a way to convert them into the numeric format of SVM.

4. EXPERIMENTS AND EVALUATION

4.1 The experimental dataset and scenarios

The experimental dataset includes 7000 audio files, including 2000 cover files and 5000 stego files. In particular, stego files consisting of 1000 stego files using Echo technique; 1000 stego files using LSB technique; 1000 stego files using Phase technique; 1000 stego files using Spread Spectrum technique; 1000 stego files using Steghide technique.

All audio files are extracted 24 features which describe in table 1. The cover audio files are labeled '0' and the stego audio files are labeled '1'.

The above dataset is divided into 5 datasets by combining the cover files with each type of stego file. Specifically, the experimental scenario is as follows:

1. Dataset A: consists of 1000 cover audio files and 1000 stego files using LSB technique.
2. Dataset B: consists of 1000 cover audio files and 1000 stego files using Echo technique.
3. Dataset C: consists of 1000 cover audio files and 1000 stego files using Phase technique.
4. Dataset D: consists of 1000 cover audio files and 1000 stego files using Spread Spectrum technique.
5. Dataset E: consists of 1000 cover audio files and 1000 stego files using Steghide technique.

4.2 Evaluation criteria

The accuracy of a classification model (ACC) is calculated by the following formula:

Accuracy: the percentage of correct decisions among all testing samples

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\ \%$$

Where: TP (True positive) is the number of stego files that are correctly classified; FN (False negative) is the number of stego files that are incorrectly classified; TN (True negative) is the number of cover files that are correctly classified; FP (False positive) is the number of cover files that are incorrectly classified.

5.3 Experimental Results

Table 2 below describes the results of audio steganography detection using the SVM and Random Forest algorithms.

Through experimental results, the Random Forest algorithm gives higher accuracy than the SVM algorithm. With the Random Forest algorithm, the highest accuracy is 94% with dataset A. Meanwhile, the lowest accuracy is 86.4% when using the SVM algorithm.

From those experimental results, can see that the steganography method using the LSB technique often changes some structure and information distribution ratio of images, making them easier to detect. In addition, the steganography method using the Spread Spectrum technique gives the lowest results. This demonstrates that the method of detecting audio steganography is very difficult to analyze stego audio files that use the Spread Spectrum method.

Table 2:. Experimental results of audio steganography detection

Algorithm	Accuracy (%)				
	Dataset A	Dataset B	Dataset C	Dataset D	Dataset E
SVM algorithm	0.8640	0.8687	0.8660	0.8693	0.8767
Random Forest algorithm	**0.9400**	0.9380	0.9193	0.8993	0.9347

5. CONCLUSION AND FUTURE DIRECTION

Detecting audio steganography quite difficult and complex. In this paper, based on a combination of features from different domains and the Random Forest machine learning algorithm, we have successfully built a model for detecting audio steganography with high accuracy. In the future, we will improve the feature to detect stego audio files even when these files are compressed data. We will also perform clustering by audio content before analysis to be more effective.

REFERENCES

[1] X.-M. Ru, H.-J. Zhang, and X. Huang. **Steganalysis of Audio: Attacking The Steghide**.*Machine Learning and Cybernetics*, Proceedings of 2005 International Conference, pp. 3937-3942, 2005.

[2] S. Rekik. **An Autoregressive Time Delay Neural Network for Speech Steganalysis**. *Information Science, Signal Processing and their Applications (ISSPA)2012 11th International Conference*, pp. 54-58, 2012.

[3] C. Kraetzer. **Mel-Cepstrum-Based Steganalysis for Voip Steganography**.*Electronic Imaging 2007, J. Dittmann*, pp. 505-512-12, 2007.

[4] Q. Liu, A. H. Sung, and M. Qiao. **Temporal Derivative-Based Spectrum and Mel-Cepstrum Audio Steganalysis**.*IEEE Trans. Information Forensics and Security*, vol. 4, pp. 359-368, 2009.

[5] Y. Wei, L. Guo, Y. Wang, and C. Wang. **A Blind Audio Steganalysis Based on Feature Fusion.***Journal of Electronics*, vol. 28, pp. 265-276, 2011.

[6] H. Özer, B. Sankur, N. Memon, and İ. Avcıbaş. **Detection of Audio Covert Channels Using Statistical Footprints of Hidden Messages**.*Digital Signal Processing*, vol. 16, pp. 389-401, 2006. https://doi.org/10.1016/j.dsp.2005.12.001

[7] S. Geetha, N. Ishwarya, and N. Kamaraj. **Audio Steganalysis with Hausdorff Distance Higher Order Statistics Using a Rule Based Decision Tree Paradigm**. *Expert Systems with Applications*, vol. 37, pp. 7469- 7482, 2010.

[8] Z. Kexin. **Audio Steganalysis of Spread Spectrum Hiding Based on Statistical Moment**. *Signal Processing Systems (ICSPS 2010)*, pp. V3-381-V3-384, 2010.

[9] W. Zeng, H. Ai, and R. Hu. **A Novel Steganalysis Algorithm of Phase Coding in Audio Signal**.*Advanced Language Processing and Web Information Technology (ALPIT 2007)*, pp. 261-264, 2007.

[10] W. Zeng, H. Ai, and R. Hu. **An Algorithm of Echo Steganalysis Based on Power Cepstrum and Pattern Classification**.*Audio, Language and Image Processing (ICALIP 2008)*, pp. 1344-1348, 2008. https://doi.org/10.1109/ICALIP.2008.4590036

[11] Smola, A.; Vishwanathan, S.V.N. **Introduction to Machine Learning**.*Cambridge University Press*: Cambridge, UK, 2008.

[12] Leo Breiman. **Random Forests**.*Machine Learning*, vol. 45, no. 1, pp. 5- 32, 2001.

[13] Thomas G. Dietterich. **Ensemble Methods in Machine Learning**. *Proceedings of the International Workshop on Multiple Classifier Systems, (MCS 2000)*, pp 1-15, 2000.

[14] Cho Do Xuan, HoaDinh Nguyen and Tisenko Victor Nikolaevich, "Malicious Url Detection Based on Machine Learning," International Journal of Advanced Computer Science and Applications(IJACSA), 11(1), 2020, http://dx.doi.org/10.14569/IJACSA.2020.0110119.

[15] C.J.C. Burges Chris J.C. **Burges. A Tutorial on Support Vector Machines for Pattern Recognition**,*Data Mining and Knowledge Discovery*, vol. 2, pp 121-167, 1998. https://doi.org/10.1023/A:1009715923555