

Steganalysis of LSB Matching in WAV Audio*

Jie Zhang, Feng Du

College of Information Science and Technology
Hainan University
Haikou, China
zhangxiaoyuanwork@163.com

Songbin Li

Haikou Laboratory
Institute of Acoustics, Chinese Academy of Sciences
Haikou, China
lisb@dsp.ac.cn

Abstract—To expose the existence of hidden message, this paper presents a steganalysis method for LSB (Least significant bit) matching steganography in wav audio based on multi-order Markov feature. Noise sequences are extracted respectively based on two methods of local correlation and 5/3 wavelet de-noising, and then the correlation of noise sequence is quantified as multi-order Markov feature vectors. Based on the extracted feature vectors, a steganography detector based on support vector machine (SVM) is trained to identify whether the wav audio file carries hidden data. Experimental results show that detection accuracy can reach more than 80% even when the embedding rate is only 10%, which is far better than other method.

Keywords—LSB matching; steganalysis; local correlation; wav-elet de-noising; support vector machine (SVM)

I. INTRODUCTION

Steganalysis of digital information is to judge whether the statistic properties of given files are changed, thereby exposing the existence of hidden message, or even to extract the hidden message. The most simple and classic audio steganography algorithm is the time domain of the LSB algorithm, which includes LSB substitute and LSB matching. However, LSB matching steganography can avoid the statistical asymmetry of value pair transition caused by LSB substitute steganography. And this phenomenon leads that the degree of difficulty for detecting increased significantly. Thus, how to achieve effectively detecting has aroused some scholars' attention. For detection of LSB matching steganography in audio, Liu proposed a steganalysis method based on temporal derivative-based spectrum and Mel-cepstrum [1]. Yang proposed a steganalysis method for LSB Matching in MIDI audio based on the rate of smoothness [2]. Cogranne studied a reliable detection method based on assumptions detection theory, and designed statistical model of medium to deal with nuisance parameters [3]. Up to now, there have been not too many reports about steganalysis of LSB matching steganography in wav audio. However, there are a lot of researches of steganalysis on other information hiding methods in audio. Such as, Kraetzer proposed a steganalysis

Priority Research Program of the Chinese Academy of Sciences (no. XDA06010302).

method of Mel-cepstrum for LSB matching steganography in VoIP [4-5]. Meng proposed a scheme for steganalysis of MP3Stego based on feature mining and pattern recognition techniques [6]. Li proposed a detection of QIM steganography in speech based on quantization index sequence analysis [7]. Wei proposed a blind steganalysis based on feature fusion [8]. Although these methods are not specific to LSB matching steganography, the thoughts and the methods have great reference significance.

Normally, audios hidden data at high embedding rate can be detected easily. But in practical application, audios with a low embedding rate can hide a lot of information because its length is unlimited. Therefore, the study of low embedding rate audio LSB matching steganography detection method is quite necessary. The key point of steganalysis is to extract some sensitive statistical properties of steganography. In order to extract it, and due to the fact that there is a correlation between the noise of audio sampling signals, this paper presents a method of extracting high-dimensional feature for steganalysis based on multi-order Markov model. In order to realize the detection of LSB matching steganography in the audio at low embedding rate, this paper constructs a detector combining extracted features with support vector machine.

This paper is organized as follows. Section II presents a LSB matching steganography method in audio. Section III introduces how to model the correlation of audio sampling noise and gives two extraction methods of high-dimensional feature. Section IV introduces the design of classifier. Then experimental results are presented and summarized in section V. Finally, section VI concludes the work.

II. STEGANOGRAPHY METHOD OF LSB MATCHING IN AUDIO

The principle of LSB matching is to plus or minus 1 on the least significant bit randomly. In this section, a LSB matching steganography method in audio is given. The formula is shown as follows:

$$P_s = \begin{cases} P_c + 1 & \text{if } b \neq \text{LSB}(P_c), (M_i = 1) \text{ or } (P_c = 0) \\ P_c - 1 & \text{if } b \neq \text{LSB}(P_c), (M_i = 0) \text{ or } (P_c = 255) \\ P_c & \text{if } b = \text{LSB}(P_c) \end{cases} \quad (1)$$

*This work is finished when the first author doing the internship in Haikou Laboratory, Institute of Acoustics, Chinese Academy of Sciences. This work is supported by the National High Technology Research and Development Program of China (no. 011AA01A102) and the Strategic

Where b denotes a hidden secret bit, P_c denote a signal carrier, P_s denotes a new signal carrier that hidden secret bit. M_i is a message sequence that consists of random variables 0 and 1, these variables obey uniform distribution. The sequence's length is the number of hidden secret information bits. Therefore, the least significant bits of the quantized sample values have equal probability to plus 1 or minus 1. That's to say, the change of value pair caused by LSB matching steganographic is basically symmetrical. For this reason, the LSB matching steganography cannot be detected effectively by methods of RS, SPA, etc.

III. HIGH-DIMENSIONAL FEATURE EXTRACTION

A. Model the correlation of sampling signal noise

The digitized voice signals are obtained from a continuous voice signals by steps of sampling, quantization, etc. As affected by the collected environment and the precision of acquisition equipment, the digitized voice signals contain a certain noise. Therefore, the voice sampling signal sequence (without hidden data) can be expressed as : $S(n) = \hat{S}(n) + \delta(n)$, where $\hat{S}(n)$ denotes the real voice sampling value, $\delta(n)$ denotes the noise value at moment n . In this paper, LSB matching steganography is regarded as additional noise. Therefore, the voice sampling signal sequence (with hidden data) can be expressed as : $S(n)^{LSB} = \hat{S}(n) + \delta(n) + \delta(n)^{LSB} = \hat{S}(n) + \hat{\delta}(n)$, where $\hat{\delta}(n)$ denotes the integrated noise.

Collected environment and acquisition equipment are relatively stable during a very short period of time. Hence, the adjacent voice samples collected from this time have a strong correlation (such as $\delta(n+1) = \delta(n)$). We model the correlation of sample noise by using Markov model, and extract feature vectors for detection of LSB matching steganography. Based on the assumption that the current noise signal is only related to the former noise signal, we can describe the correlation of noise with first-order Markov model:

$$M_{\lambda_1 \lambda_2} = \Pr(\delta(n+1) = \lambda_1 | \delta(n) = \lambda_2) = \frac{\Pr(\delta(n+1) = \lambda_1, \delta(n) = \lambda_2)}{\Pr(\delta(n) = \lambda_2)} \quad (2)$$

If it is assumed that the current noise signal is related to the former two noise signals, we can describe the correlation of noise with second-order Markov model:

$$M_{\lambda_1 \lambda_2 \lambda_3} = \Pr(\delta(n+2) = \lambda_3 | \delta(n+1) = \lambda_1, \delta(n) = \lambda_2) = \frac{\Pr(\delta(n+1) = \lambda_1, \delta(n) = \lambda_2, \delta(n+2) = \lambda_3)}{\Pr(\delta(n+1) = \lambda_1, \delta(n) = \lambda_2)} \quad (3)$$

Higher order Markov model can be defined for the correlation of more neighbor noise signals. However, since the correlation recedes gradually with the rise of order, it is

unnecessary to define higher than second-order Markov model to describe it.

B. Estimation of noise sequence

Assuming that there is a sampling value sequence of voice : $S(n) = \{t_1, t_2, \dots, t_n\}$. The noise sequence $\delta(n)$ of $S(n)$ can be estimated in two ways. The first is based on lifting wavelet transform, and the other is based on local correlation of voice sampling signals

a) Extract noise sequence based on wavelet de-noising

The wavelet transform has many characteristics, such as low entropy performance, de-correlation performance, multi-resolution performance, flexible base-chosen performance, etc. The main principle of that is to separate high and low frequency signals in the premise of without losing the important components of original signals, and reconstruct (inverse transformation of wavelet) the signals with proper wavelet coefficients in wavelet space. In this way, noise can be eliminated. This paper exploits 5/3 lifting wavelet transform that defined in JPEG2000. 5/3 lifting wavelet transform is a kind of reversible wavelet transform. And its calculation equations of decomposition and reconstruction process are shown as the following formulas:

Decomposition:

$$\begin{aligned} y(2n+1) &= x(2n+1) - \left\lfloor \frac{x(2n) + x(2n+2)}{2} \right\rfloor \\ y(2n) &= x(2n) + \left\lfloor \frac{y(2n-1) + y(2n+1) + 2}{4} \right\rfloor \end{aligned} \quad (4)$$

Reconstruction:

$$\begin{aligned} x(2n) &= y(2n) - \left\lfloor \frac{y(2n-1) + y(2n+1) + 2}{4} \right\rfloor \\ x(2n+1) &= y(2n+1) + \left\lfloor \frac{x(2n) + x(2n+2)}{2} \right\rfloor \end{aligned} \quad (5)$$

After multi-level wavelet decomposing of voice signals, we can obtain high-frequency and low-frequency signals from each level. Then, the high-frequency signals that contain noise are modified by threshold processing. Finally, the de-noised voice signal can be reconstructed according to (5). There are two threshold processing methods: hard threshold and soft threshold processing. However, hard threshold processing will make the signals oscillate in the vicinity of the singular point. Thus, in order to decrease the oscillation, we apply soft threshold processing, shown as the following formula:

$$w = \begin{cases} w - \lambda, & w \geq \lambda \\ 0, & w < \lambda \\ w + \lambda, & w < -\lambda \end{cases} \quad (6)$$

Based on the method above, we can obtain the estimated signals $\bar{S}(i)$ of original signal $S(i)$. Then the noise sequence can be extracted by formula (7):

$$\delta^1(i) = S(i) - \bar{S}(i), i = 1, 2, \dots, n-1 \quad (7)$$

The first and second-order correlations of noise is generated by using the formula $\Pr(\delta(n+1) = \lambda_1 | \delta(n) = \lambda_2)$ and $\Pr(\delta(n+2) = \lambda_3 | \delta(n+1) = \lambda_1, \delta(n) = \lambda_2)$. The results are shown in Figure 1. The noise data are extracted from 1600, 8-bit encoded audios. In Figure 1, the different colors and sizes of dots correspond to the degree of first and second-order correlation of noise.

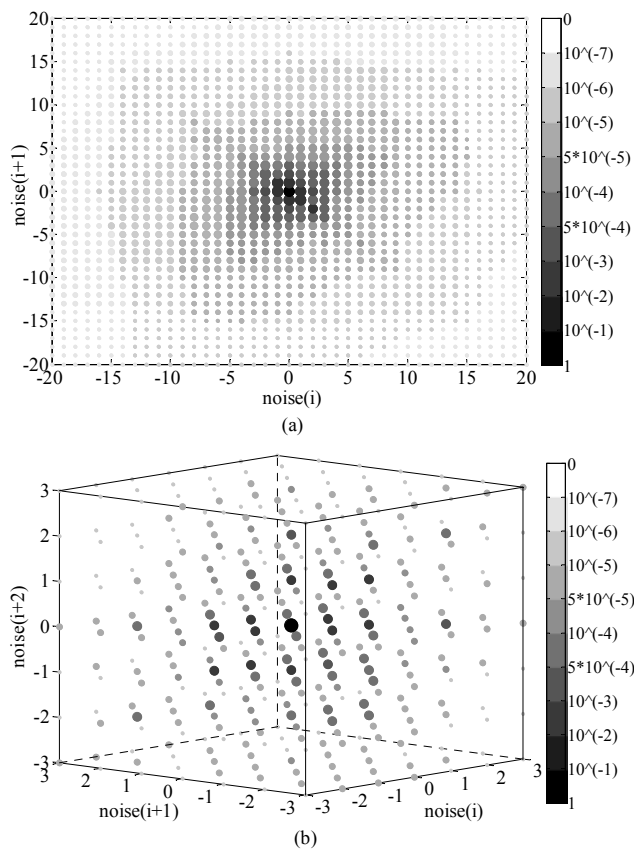


Figure 1. Correlation of the noise that extracted based on the method of wavelet de-noising (a) first-order correlation of the noise (b) second-order correlation of the noise

As seen in Fig.1, the distribution of occurrence probabilities is uneven, and occurrence probabilities reduce outward radially. In Fig.1(a), noise values in the range $[-3, 3]$ have bigger occurrence probabilities. That is mean that the noise values are more close to 0, the occurrence probabilities are bigger. In Fig.1(b), the three adjacent noises whose differences in the range $[-1, 1]$ have bigger occurrence probabilities and stronger correlation. Consequently, with the noise values and their difference values increasing, the

occurrence probability and correlation of which reduce rapidly.

b) Extract noise sequence based on local correlation

It is the fact that the correlation exists between voice signals that sampled in a short time. Therefore, the adjacent voice sampling signals t_n and t_{n+1} should be same at many moments in the premise of ignoring the noise. Since the noise affect voice sampling value randomly, we can assume that the noise is superimposed to t_n or t_{n+1} . Thus, the difference of adjacent voice sampling values ($t_n - t_{n+1}$) can be regarded as a noise value. According to that, we can extract the noise sequence by formula (8):

$$\delta^2(i) = t_i - t_{i+1}, i = 1, 2, \dots, n-1 \quad (8)$$

c) Multi-order Markov feature vector of noise sequence

Since the range of $\delta^1(i)$ and $\delta^2(i)$ is two times as large as that of voice sampling values, we will get too many conditional probabilities if the value filed is too large. Thereby, a threshold μ is indeed needed. We calculate the conditional probabilities only when continuous noise values meet the requirement: $|\delta(i)| \leq \mu$. In this way, the number of first-order conditional probabilities of $\delta^1(i)$ are $(2\mu+1)^2$, we assume that these conditional probabilities are defined as $M_{-\mu, -\mu}^1, M_{-\mu, -\mu+1}^1, \dots, M_{\mu, \mu}^1$; and the number of second-order conditional probabilities are $(2\mu+1)^3$, we assume that these conditional probabilities are defined as $M_{-\mu, -\mu, -\mu}^2, M_{-\mu, -\mu, -\mu+1}^2, \dots, M_{\mu, \mu, \mu}^2$. In the same way, we can get $4(2\mu+1)^2(\mu+1)$ conditional probabilities of $\delta^2(i)$ in total. Since these conditional probabilities effectively quantify the correlation of adjacent noise signals, a feature vector can be defined as: $F = \{M_{\lambda_1, \lambda_2}^1, M_{\lambda_1, \lambda_2, \lambda_3}^1, M_{\lambda_1, \lambda_2}^2, M_{\lambda_1, \lambda_2, \lambda_3}^2\}$, and F is called as multi-order Markov (MOM) feature vector. In this paper, we take a threshold value: $\mu = 3$.

IV. THE DESIGN OF CLASSIFIER

The design of classifier is another key point in steganalysis. Currently, since machine learning-based classification method is a mainstream, support vector machine (SVM) has got a lot of attention. As an effective classified tool, SVM can solve two-class problem effectively and avoid the curse of dimensionality to some extent.

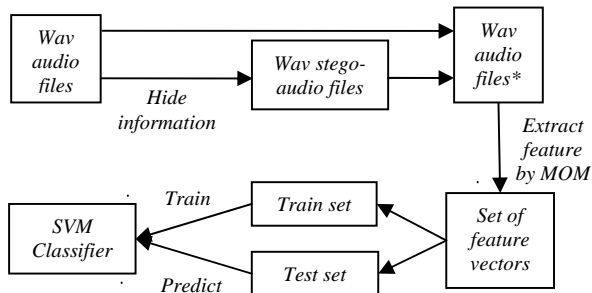


Figure 2. The steganalysis process based on SVM

The basic idea of the SVM is based on the theory structural risk minimization theory. That can be summarized as: mapping the sample to high dimensional space by nonlinear transformation, and then construct an optimal intersected hyperplane in the feature space to make the learning machine global optimization. In the experiments, LIBSVM is used as the classifier, and the kernel function of SVM is the RBF function. To sum up, based on SVM, the steganalysis process is summarized as Fig.2.

TABLE I. DETECTION ACCURACY OF MULTI-ORDER MARKOV FFEATURE (MOM) AND DERIVATIVE-BASED MEL-CEPSTRUM (D-MC) STEGANALYSIS METHODS. THE EMBEDDING RATE IS 1%,5%,10%,30%,50%,80%,100% RESPECTIVELY.

File length	1%		5%		10%		30%		50%		80%		100%	
	MOM	D-MC	MOM	D-MC	MOM	D-MC	MOM	D-MC	MOM	D-MC	MOM	D-MC	MOM	D-MC
1s	59.37%	57.50%	61.00%	61.62%	62.25%	63.75%	64.25%	69.50%	67.75%	79.00%	73.00%	86.87%	75.50%	88.50%
2s	61.62%	56.12%	64.50%	61.37%	65.37%	63.50%	68.37%	69.50%	74.50%	80.37%	72.00%	88.75%	84.75%	92.25%
3s	63.62%	56.25%	66.75%	61.12%	68.87%	65.75%	73.37%	70.25%	79.12%	79.37%	86.00%	91.00%	89.12%	94.75%
4s	64.62%	55.87%	68.87%	61.75%	71.62%	65.25%	75.62%	70.12%	82.62%	81.50%	88.75%	92.12%	92.12%	94.25%
5s	66.12%	54.75%	70.25%	60.25%	73.25%	63.87%	78.12%	69.50%	84.87%	80.75%	91.25%	91.37%	94.37%	95.00%
6s	69.37%	54.62%	73.12%	58.25%	74.87%	62.75%	90.62%	65.75%	86.37%	81.00%	93.00%	91.00%	95.12%	94.75%
7s	71.37%	55.12%	76.37%	59.50%	77.87%	64.62%	81.75%	69.12%	88.37%	91.00%	93.37%	93.75%	97.00%	94.50%
8s	73.50%	53.87%	77.25%	59.25%	79.25%	61.87%	84.00%	65.37%	90.25%	79.12%	95.62%	90.75%	97.62%	94.12%
9s	73.50%	55.12%	77.25%	60.12%	79.25%	62.12%	84.00%	68.75%	90.25%	82.25%	95.62%	92.87%	97.62%	94.75%
10s	74.75%	55.25%	78.37%	60.87%	81.37%	62.37%	84.50%	69.50%	90.62%	82.12%	95.50%	91.00%	98.50%	94.00%

VI. CONCLUSION

In this paper, the steganalysis method for LSB matching steganography in wav audio is studied. The experiments on a large number of wav audio files show that the performance of our method is superior to the method based on derivative-based Mel-cepstrum features. It is notable that the method proposed in [1] is a kind of universal steganalysis method for steganography in wav audio. However, our method is special for LSB matching steganography in wav audio. Since our method has stronger pertinence and extracts some sensitive features of steganography, we can attain better detection performance. In this paper, only 16-bit quantization PCM coded wav audios are considered, more work on 8-bit quantization wav audios will be further studied.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In order to prove the good universality and applicability of our method, we have experiments on wav audios that have different lengths and embedding rates. The experiment selects 8-kHz 16-bit quantization PCM coded wav audio files as test target. And the wav audio files consists of different lengths that range from 1 second to 10 seconds (each length file has 2000 segments). Then, the experimental results are shown in table 1. We still use the derivative-based Mel-cepstrum (D-MC) method proposed in [1] to test the audio files, and results are also shown in Table I.

As seen in Table I: 1) when the embedding rate is lower than 50%, classifying performance of our method is much superior to that in [1]. Especially, at the low embedding rate of 10%, detection accuracy of our method is as much higher as 10%; 2) when the embedding rate is more than 80%, there is not much difference between the classifying performance of two methods; 3) it is obviously that detection accuracy of our method is related to the length and the embedding rate of the detected wav audio files. Hence, our method can attain better classification performance if the file's length increases.

REFERENCES

- [1] Q. Liu, A.H.Sung, M. Qiao, "Temporal Derivative-Based Spectrum and Mel-Cepstrum Audio Steganaly-sis," IEEE Transactions On Information Forensics And Security, vol.4, Sept. 2009, pp.359-368, doi:10.1109/TIFS.2009. 2024718.
- [2] B.Yang, L. Guo, Y.J. Wang, C.P. Wang, "A Steganalysis Method of LSB Matching for MIDI Audio," Information Security and Communications , May.2010,vol.5, pp.87-89.
- [3] R. Cogranne, C. Zitzmann, F. Retraint, I. Nikiforov, L. Fillatre, P. Cornu, "Statistical Detection of LSB Mat-ching in the Presence of Nuisance Parameters," Proc. IEEE Statistical Signal Processing Workshop (SSP12), IEEE Press, Aug.2012, pp.912-915, doi: 10.1109/SSP.2012.6319857.
- [4] Christian Kraetzer, Jana Dittmann, "Mel-Cepstrum Based Steganalysis for VoIP-Steganography," Security, steganography, and watermarking of multimedia contents IX, Feb.2007, vol.6505, pp.650505.1-650505.12, doi:10.1117/12.704040.

- [5] Christian Kraetzer Jana Dittmann, "Impact of feature selection in classification for hidden channel detection on the example of audio data hiding," Proc. ACM workshop on Multi-media and security (MMSec12), IEEE Press, Sept.2008, pp.159-166, doi: 10.1145/1411328.1411356.
- [6] M.Y. Qiao, A. H. Sung, Q.Z. Liu, "Feature Mining and Intelligent Computing for MP3 Steganalysis," Proc. International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing (IJCBS09), IEEE Press, Aug.2009, pp.627-630, doi: 10.1109/IJCBS. 2009.119.
- [7] S.b. Li, H.Z. Tao, Y.F. Huang, "Detection of QIM Steganography in G.723.1 Bit Stream Based on Quantization Index Sequence Analysis," Journal Zhejiang University Science C, vol.13, Aug. 2012, pp.624-634, doi:10.1631/jzus.C1100374.
- [8] Y.F. Wei, L. Guo, Y.J. Wang, C.P. Wang, "A Blind Audio Steganalysis Based on Feature Fusion," Journal Of Electronics (CHINA), Vol.28, Feb.2011, pp.265-276, doi: 10.1007/ s11767-011-0567-z.