Multimed Tools Appl (2018) 77:15431–15455
DOI 10.1007/s11042-017-5123-x

A new audio steganalysis method based on linear prediction

Chunling Han[1] · Rui Xue[1] · Rui Zhang[1] · Xueqing Wang[1]

Received: 7 February 2017 / Revised: 30 June 2017 / Accepted: 16 August 2017 /
Published online: 1 September 2017

Abstract Steganography and Steganalysis have attracted a lot of attention in decades. Recently, voice communication has been more and more popular, which provides ways to covert communication. However, the existing audio steganalysis methods can only gain good detection accuracies when the hidden ratio is high. Besides, majority of the audio steganalysis methods can not provide a general evaluation, only provide the detection accuracies according to several high hidden ratios. In this paper, we proposed a new method for audio steganalysis by introducing linear prediction method, a technique from signal coding and speaker identification filed, into audio steganalysis, which can bring significant differences between covers and stegos. The linear prediction based features are utilized as the classification features loaded in a support vector machine for detection. In our work we used hidden message to cover ratio to replace the concept of hidden ratio, providing a uniform criterion to compare the performance among steganalysis methods. Furthermore, we exploited a general dataset, in which the hidden message size ranges from several bits to the maximum hiding capacity for a general evaluation on steganalysis methods. Experiment results show that our method delivers a better performance than previous two prestigious methods and brings above 96% accuracy. In general evaluation, our method gains a higher score than the other two methods. Steganalysis is a challenging work, this linear prediction based method maybe an approach to bring improvement to this filed and provide inspiration for other form of media steganalysis.

Keywords Linear prediction · Audio steganalysis · Steganography · Support vector machine

✉ Rui Xue
xuerui@iie.ac.cn

[1] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

1 Introduction

1.1 Related Work

In recent years, steganography and steganalysis have attracted increasing attention. Steganagraphy provides a way to accomplish covert communication, using images, audios and videos as common carriers. In contrast, steganalysis is used to detect the suspicious files, to detect whether the carrier has been hidden secret message, sometimes it disposes the suspicious carrier or extracts the secret message when it is possible. There have been many methods for image, audio and video steganalysis, especially in image steganalysis [3, 17, 19, 30, 31], the improvement of image steganography [10, 35] and image processing have given a lot of inspiration to this field, the steganalysis of image [31] has been more and more mature. Recently, voice communication has been popular among social networks, such as WhatsApp, Facebook, Wechat, LINE and other social platforms. The popularity of voice communication makes audio signal become a common carrier to hide information. There have been many practical steganography tools to hide secret massage in audio files, including S-Tools, Hide4PGP for wave format audio files, MP3Stego for MP3 files and other methods that have not been developed into applications or softwares. However, there are less steganalysis methods for audio files, as a consequence, audio steganalysis becomes more and more important.

Recently, audio steganalysis has been investigated by many research groups. For WAV format files, Liu et al. proposed a derivative-based Fourier spectrum and Mel-cepstrum audio steganalysis method [11], their experiments received a good performance when the hidden ratio is high. Later, Liu et al. used Mel-cepstrum coefficients and Markov transition features from the second-order derivative of the audio signal [12] and obtained a better performance.

Djebbar et al. used lossless data compression ratios as the classification features [4], their method can efficiently detect suspicious WAV files embedded by S-tools4, Steghide and Hide4PGP. Tint et al. proposed a steganalysis method for WAV files by extracting Mel-frequency ceptral coefficients, zero crossing rate, spectral flux and short time energy features of audio files [27], their method performs well in steganalysis of stegos produced by Hide4PGP, Stegowav and S-tools4.

Yavanoglu et al. presented a new intelligent steganalysis method and developed a steganalysis software named HITIT [33], they extracted FFT features and applied efficient and complicated artificial neural networks to undertake the classification task, they received 75% accuracy. Ghasemzadeh et al. introduced a new Reversed-Mel scale, they used this new model to construct a virtual ear which is highly sensitive to the variation in high frequency region [7]. Their work received a higher accuracy rate than previous MFCC based methods [9]. Later, they proposed a new psychoacoustic model of human hearing model [8] to improved the previous 2D-MM based method [12].

As for MP3 audio files, Yu et al. used the main_data_begin in side information as the feature [34] and applied a recompression calibration to detect suspicious MP3 files generated by MP3Stego. Their result shows that their method is effective when the hidden ratio is high. Qiao et al. proposed a comprehensive steganalysis method for MP3 audio files [20], attacking the steganography software MP3Stego, the features derived from the combination of quantized MDCT coefficients, frequency-based subband moment statistical features, accumulative Markov transition features and accumulative neighboring joint density features on

second order derivatives. They applied different feature selection algorithms to improve the detection accuracy. Yan et al. presented a method to attack MP3Stego using standard deviation of the second-order differential sequence from quantization step [32]. Their experiment result suggests that their method is effective and has achieved better performance than other steganalysis methods for MP3Stego.

For other audio signal forms, Ren et al. proposed an AMR steganalysis method based on the probability of same pulse position [21] to attack the existing steganography methods proposed by Geiser et al. [6] and Miao et al. [15], when the hidden ratio reaches 60%, their steganalysis accuracy is above 99%. Tian et al. presented a statistical method using statistical features of pulse pairs [26] to detect AMR audio files processed by Geiser's and Miao's steganography methods, their approach delivered a good performance when the hidden ratio is above 80%. Ren et al. used calibrated Markov model of adjacent codebook to detect AAC files [22], attacking Huffman codebook based steganography algorithms proposed in [25, 37], they obtained almost 100% accuracy in some cases.

Though there have been a number of audio steganalysis methods for different types of audio files. The existing audio steganalysis methods can introduce high false positive rate and have limited application range. Some methods [12, 21, 26] can only get a high detection accuracy when the carrier has been hidden in a large ratio of message, for example, over 60% of the maximum hiding capacity.

However, in practical use, the hidden message size is usually relatively small, sometimes it will be only several bits. When the hidden ratio is low, these steganalysis methods will bring low detect accuracies. The poor performance in detecting low ratio hidden message making these steganalysis methods not practical in reality. Remarkably, the evaluation of steganalysis methods is based on the hidden ratio. Nevertheless, how can the detector know the hidden ratio before he extracts the hidden message? To some extent, this kind of evaluation does not bring a practical evaluation of a steganalysis method.

1.2 Contributions

To overcome these drawbacks, in this paper, we proposed an audio steganalysis method using linear prediction based features, our scheme received a good performance. Furthermore, we proposed a new general evaluation method for audio steganalysis algorithms to close the gap between experiment and practice. Additionally, our steganalysis method can effectively attack the steganography tools Hide4PGP, S-Tools and two newly emerged steganography tools StegoMagic and Xiao Steganography. The contributions of this paper are as follows.

Firstly, we proposed an audio steganalysis method using linear prediction. Linear prediction is usually used as a tool in signal coding and speaker identification, we introduced this technique into audio steganalysis, which can bring significant differences between the cover and the stego. We extracted linear prediction coefficients, linear prediction residual, linear prediction spectrum and linear prediction cepstrum coefficients as features. In our experiment, we noticed that the linear prediction based feature differences between the cover and the stego are dramatic, in some cases, they are magnitude order differences. Support vector machine (SVM) with a Gaussian radial basis function (RBF) kernel was employed as the classifier, a k-fold cross validation method was used to find the best parameters and to improve the detection accuracy.

Secondly, We defined the "hidden message to cover ratio (HCR)" to replace the concept of hidden ratio. The hidden ratio is calculated by using hidden message size over maximum hiding capacity, it can not reflect the relation between hidden message and the cover. In addition, it is not suitable to compare the hiding strength among steganography algorithms. Therefore, we used the concept of HCR by calculating the hidden message size over cover size, reflecting the relation between the hidden message and the cover. It also provides a way to evaluate the hiding strength and creates a uniform criterion compared among steganography algorithms. When using the same HCR, the lower detection accuracy means that the associated steganography algorithm has a better resistance. We also proposed a new general evaluation method, given the fact that, the hidden ratio will be relatively low in practice. To evaluate the practical performance of a steganalysis method, we collected a test dataset, in which the hidden message size ranges from the maximum hiding capacity according to different hiding algorithms to only several bits. We also tested the other two schemes [11, 12], they both got lower accuracies than our work. This general practical performance evaluation to a great extent closes the gap between an experimental evaluation and a practical one.

Thirdly, we employed our method on four steganography algorithms, including Hide4PGP, S-Tools and two newly emerged steganography algorithms StegoMagic and Xiao Steganography. In these four hiding algorithms, our method delivered better detection accuracies than previous D-MC and 2D-Mel methods.

In addition, we employed linear prediction on wavelet coefficients, in some particular cases, it delivered higher accuracy, but some cases did not. We tested two different models to solve linear prediction, autocorrelation model and geometric mean lattice model, these two methods show little difference on obtained linear prediction features. As an exploration, we tested our method on MP4 files, experiments show that our method can also analyze MP4 files attacking Huffman codebook based steganography schemes. We also tried to use Artificial Neural Network (ANN) as an optional classifier and presented its performance. These contents are attached in the discussion part.

The rest of the paper is organized as follows: Section 2 introduces the linear prediction used in audio signal analysis, linear prediction coefficients and its derived characteristics, which will be used as the classification features. Section 3 presents the proposed linear prediction based audio steganalysis scheme. Then the experiment of the method is shown in Section 4, followed by discussion in Section 5, future work in in Section 6 and conclusion in Section 7.

2 Linear prediction in audio signal analysis

2.1 Linear prediction

Linear prediction analysis [14] is an advanced technology and has been used in many fields. As an important tool utilized in audio signal process, it has been one of the most popular and effective methods, especially in evaluating the basic speech features such as fundamental tone, formant, spectrum, lower-speed transmissions, voice storage, etc. It can precisely evaluate the features of audio signal, and can correctly represent the audio signal's time domain, frequency domain features with less parameters. In this paper, we introduced this technique into audio steganalysis, extracted the linear prediction coefficients, linear prediction residual, linear prediction spectrum and linear prediction cepstrum coefficients as features to reflect the differences between covers and stegos.

2.1.1 Linear prediction equation

In a signal model, the input signal is $u(n)$ and the output signal is denoted as $x(n)$. An all-pole model system with transfer function $H(z)$ is shown in (1).

$$H(z) = \frac{G}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{1}$$

Notice that the coefficient a_i and the gain G are both parameters of the model. Then followed by a difference equation expressing the relation between $u(n)$ and $x(n)$

$$x(n) = \sum_{i=1}^{p} a_i x(n-i) + Gu(n) \tag{2}$$

Notice that, when the cover represent the signal $u(n)$, the stego can be indicated as $s(n) = u(n) + er(n)$. Then we obtain different linear prediction coefficients a_i from cover signal $u(n)$ and stego signal $s(n)$. Which can be seen from (2) and (3) So as the following coefficients derived from linear prediction coefficients.

$$x(n) = \sum_{i=1}^{p} b_i x(n-i) + G(u(n) + er(n)) \tag{3}$$

The system described by the (4) is called a linear predictor. $\hat{x}(n)$ is a prediction of $x(n)$, calculated by a linear combination of the past p values of $x(n)$.

$$\hat{x}(n) = \sum_{i=1}^{p} a_i x(n-i) \tag{4}$$

In (4), a_i is the Linear Prediction Coefficients (LPC) and p is the order of linear prediction analysis. The difference between the signal $x(n)$ and the predicted value $\hat{x}(n)$, defined as $e(n)$, is called linear prediction residual.

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{i=1}^{p} a_i x(n-i) \tag{5}$$

The basic problem of LP is finding a set of a_i values to minimize the LP residual according to the least mean square error criterion. The mean square error is given by:

$$E = \sum_{n} e^2(n) = \sum_{n} \left[x(n) - \sum_{i=1}^{p} a_i x(n-i) \right]^2 \tag{6}$$

E is minimized by setting the coefficients a_i to make the partial derivative $\frac{\partial E}{\partial a_i}$ to zero:

$$\frac{\partial E}{\partial a_i} = 0 \qquad (1 \le i \le p) \tag{7}$$

There has been many methods to calculate the LPC a_i, such as autocorrelation method, auto covariance method, lattice based method, etc. In the autocorrelation method, a Toeplitz matrix is used to express the mean square error equation, called Yule-Walker equation. Then Levinson-Durbin recursion algorithm is used to calculate the a_i.

2.2 Derived parameters

2.2.1 Linear prediction spectrum

When we get a frame of an audio signal, the p-order linear prediction model can be represented by an all-pole signal model, shown in (8)

$$H(z) = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{8}$$

when we set $z = e^{jw}$, we can get the linear prediction spectrum:

$$H(e^{jw}) = \frac{1}{1 - \sum_{i=1}^{p} a_i e^{-jwi}} \tag{9}$$

2.2.2 Linear Prediction Cepstrum Coefficients (LPCC)

As we know that the cepstrum of an audio signal can be obtained by using fourier transform, getting the modulus of the logarithm, then using fourier transform again. As for the model $H(z)$, when an inverse fourier transform is done on $log|H(e^{jw})|$, we get the cepstrum coefficients. The function of the synthesis filter derived from linear prediction is represented in (10), its impulse response is denoted as $h(n)$, and the cepstrum of $h(n)$ is $\hat{h}(n)$. According to homomorphic signal processing, we get:

$$\hat{H}(z) = log H(z) \tag{10}$$

As for $H(z)$ is analyzed in the unit circle, so it can be expressed by a series form, such that

$$\hat{H}(z) = \sum_{n=1}^{+\infty} \hat{h}(n) z^{-n} \tag{11}$$

we set $\hat{h}(0) = 0$, then take the partial derivative of (10) with respect to z^{-1}

$$\frac{\partial}{\partial z^{-1}} log \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} = \frac{\partial}{\partial z^{-1}} \sum_{n=1}^{+\infty} \hat{h}(n) z^{-n} \tag{12}$$

then we have

$$(1 - \sum_{i=1}^{p} a_i z^{-i}) \sum_{n=1}^{+\infty} n\hat{h}(n) z^{-n+1} = \sum_{i=1}^{+\infty} i a_i z^{-i+1} \tag{13}$$

when we set the coefficients of all power of z identical, we can obtain the recurrence relation between $\hat{h}(n)$ and a_i, and the linear prediction cepstrum is given as follows:

$$\hat{h}(n) = \begin{cases} a_1 & n = 1 \quad (14) \\ a_n + \sum_{i=1}^{n-1} (1 - \frac{i}{n}) a_i \hat{h}(n-i) & n \epsilon (1, p] \quad (15) \\ \sum_{i=1}^{p} (1 - \frac{i}{n}) a_i \hat{h}(n-i) & n > p \quad (16) \end{cases}$$

As we can see that when the cover is represented as signal $u(n)$, and the stego is indicated as $s(n) = u(n) + e(n)$, we obtained the different linear prediction coefficients, linear prediction residuals, linear prediction spectrums and linear prediction cepstrum coefficients from cover and stego.

3 Proposed method

Linear prediction coefficients are excellent features reflecting the characteristics of the voice, which effectively and correctly present the characteristics of time domain and frequency domain of an audio signal. We extracted linear prediction coefficients, linear prediction residual, linear prediction spectrum and linear prediction cepstrum coefficients as features to detect the changes brought by steganography algorithms to audio files in time and frequency domain.

As a clear presentation, Fig. 1 illustrates the detailed process of feature extraction. In our experiment, the samples are mono 16bit quantization WAV files with rate of 44.1kHz and each audio has a duration of 10 seconds.

WAV format audio files store two-channel audio sampled 44,100 times per second with 16 bits per sample, so WAV format files are 16 bit quantized and have rate of 44.1 kHz. Two-channel is an optional parameter, sometimes it can be mono (single-channel) and sometimes it can be stereo (two-channel). In our experiment, we use mono WAV files for simplicity. The length of audio files is 10 second, providing proper and enough space to hide the massage producing the stego audio files and being easily processed in our experiment.

In audio steganalysis, the audio signal sample will act as $\hat{x}(n)$, which will be predicted by the past p samples in linear prediction, shown in (4) (5) (9) (14) (15) (16), then the LPC, LP residual, LP spectrum and LPCC will be obtained.

To extract features, we need to split the audio files into about 20 ms frames. That is because consecutive samples of audio signal correlate only within short duration which is usually 10-30 ms, therefore in majority of methods of audio analysis [7, 16, 18, 23, 24], the frame time length is set between 10 ms and 30 ms. In our experiment we choose the intermediate value about 20 ms. According to the relation between time (T) and frequency (f): $T = 1/f$, in our experiment, the frequency is 44.1 kHz, we can set the frame size as 256, 512, 1024, 2048, etc. as empirical options, of course it can be customized according to specific setting. When we set the frame length as 1024, according to the relation between time and frequency, we get: $t = 1024/44100 = 23.2$ ms (i.e. about 20 ms). We did not use any window structure on frames.

Next, the processing procedures are listed to get the features. The steps are as follows.

1) Split the audio file into about 20 ms frames.
2) Select one frame then calculate the LP coefficients (12th order) and the gain G.

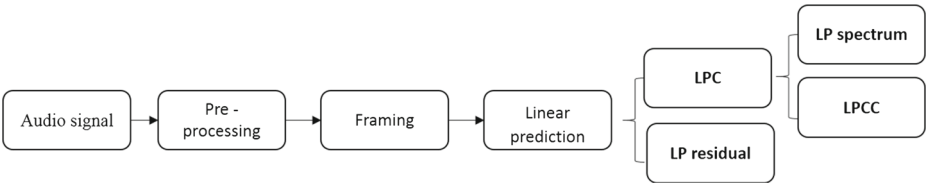


Fig. 1 The procedure of feature extraction

3) Calculate the LP residual, given in (5).
4) Calculate the LP spectrum using rfft.
5) Obtain the LPCC, shown in (14) (15) (16).

The best prediction order p typically lies at 8 to 20. In practice, we often consider the tradeoff between linear prediction performance and computation load, 12 becomes an empirical choice, we followed this empirical setting. According to the paper titled "Linear Prediction: A Tutorial Review [14]", the optimal value p is a turning point when the prediction error is the smallest, we set p at this point as p_0. When $p > p_0$, the curve slopes upward (the prediction error increases), but very gently with slight variations. Therefore, when p is at this range (from p_0 to 20), the prediction error varies slightly and will bring little effect to linear prediction. However, there has not been an accepted criterion to determine the optimal value p_0 yet, since it is affected by some parameters used on the Internet and characteristics of audio signal input. So the experimenter could feel free to adjust the choice to suit one's application.

We used Levinson-Durbin recursion algorithm to calculate the LPC, LP residual, LP spectrum and LPCC in our experiment. Additionally, we also tried to use a lattice based linear prediction to compute the features, which is illustrated in section 5 for discussion.

We selected a piece of audio signal as a cover, embedded the secret data by using Stego-Magic hiding algorithm to obtain the stego. The cover signal is a 10 second, 44.1kHz 16 bit 861KB mono WAV file. The hidden data is a 66KB BMP file.

Figure 2 presents the linear prediction based features of the cover and the stego, including LPC, LP residual, LP spectrum and LPCC. It clearly shows that the embedded information has brought differences to linear prediction based features. The linear prediction coefficients of the cover and the stego shown in (a) and (b) are different in positive and negative values. It well worth noticing that the difference of LP residual shown in (c) and (d) is a distinction on the orders of magnitude, which is a very significant difference between the cover and the stego. Similarly, the LP spectrums of the cover and the stego given in (e) and (f) show some differences in maximum values. As shown in (g) and (h), the distinction in LPCC between the cover and the stego is also obvious.

We calculated the maximum values of the four features to show the differences between the cover and the stego. As we can see from Fig. 2, there are significant differences between the cover and the stego in linear prediction features, especially in peaks and valleys. So we can use these differences as our analysis features. Maximum value is a simple and excellent choice. Take LP residual and LP spectrum for example, as shown in Fig. 2, the maximum values of LP residual and LP spectrum show significant distinction between the cover and the stego. We can also use mean value or minimum value as options, but they show smaller difference between the cover and the stego than maximum values. So maximum value is more favorable and simple for our steganalysis method.

To illustrate the difference between the cover and the stego, a detailed comparison of linear prediction based features between the cover and the stego is given in Table 1. From Table 1 we can see that the hidden message has brought significant differences to the stego compared with the cover.

In our experiment, we calculated the maximum value of LPC, LP residual, LP spectrum and LPCC as features. A support vector machine (SVM) with a Gaussian radial basis function (RBF) kernel is utilized to be the classifier. A feature extraction function is provided to be loaded in SVM model to obtain the features and then train the classifier.

To improve the detection accuracy, we need to adjust the parameters of the SVM. We employed a 4-fold cross validation method to find the best parameters. Then used the best

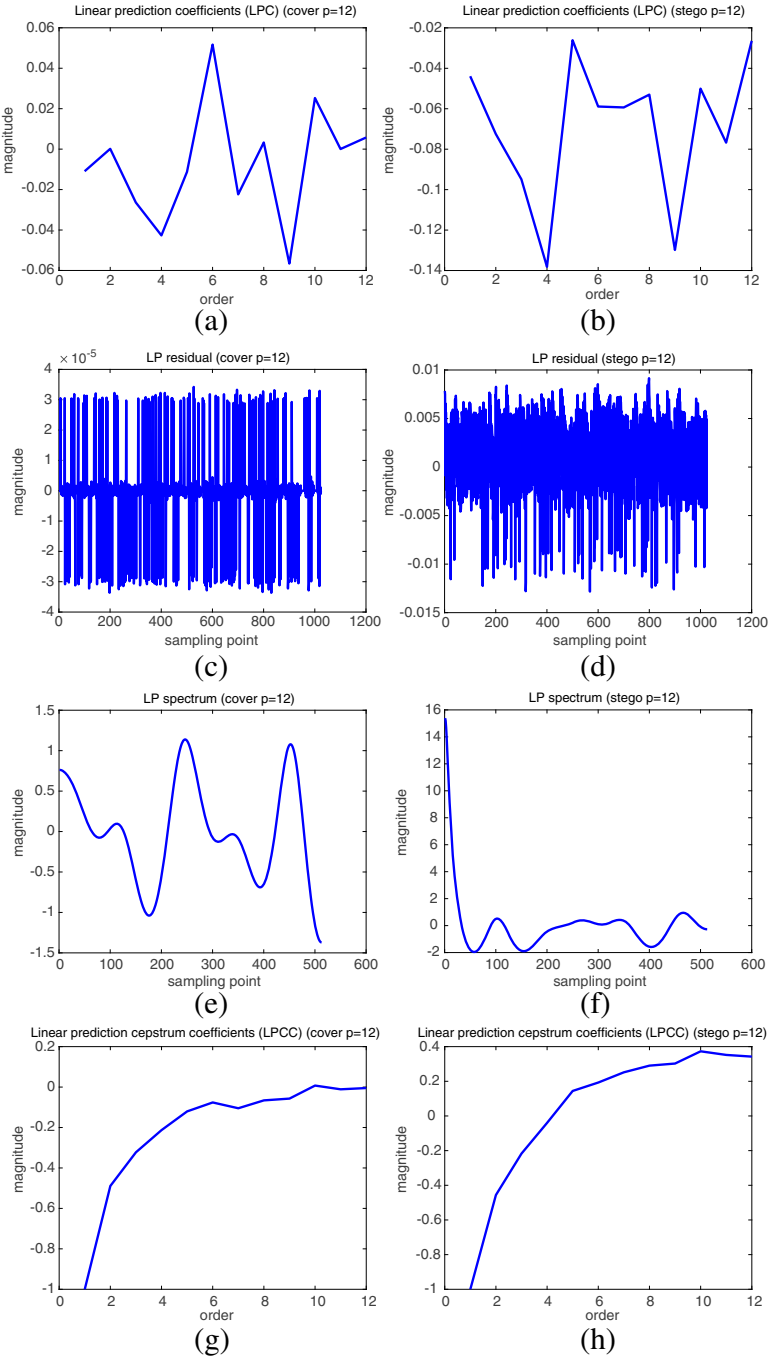


Fig. 2 The differences of linear prediction based features between a cover audio signal (left) and its asociated stego audio signal (right), including the comparison of LPC shown in (**a**) and (**b**), the difference of LP residual given in (**c**) and (**d**), the difference of LP spectrum shown in (**e**) and (**f**), and distinction in LPCC shown in (**g**) and (**h**)

Table 1 The distinction of features between the cover and the stego

	LPC	LP residual	LP spectrum	LPCC
Cover	0.0517	3.4224e-05	1.1386	0.0074
Stego	−0.0261	0.0092	15.3864	0.3731

parameters to retrain the classifier and classify the test samples. Figure 3 illustrates the detailed processing phrases of this machine model. It works as following steps.

1) Feature Extraction. When the datasets have been collected, firstly, features are extracted from data files using feature extraction functions. In our experiment, we applied the maximum value of each feature as the classification feature, these are the maximum value of LPC, LP residual, LP spectrum and LPCC.

2) k-fold cross validation. A 4-fold cross validation is applied to train the machine, using three different quarters of the dataset as the training set and finding the best parameters.

3) Decision. When the training phrase is completed, a decision phrase is conducted on the test data (the rest quarter of the dataset) using the best parameters and the detection result is shown at the end.

4) The training and decision phrases are repeated k times and obtain the mean value of the decision result.

4 Experiment

In this Section we conducted several experiments to test the accuracy of our linear prediction based audio steganalysis method. Four features were extracted by using feature extraction functions, a SVM with RBF kernel was applied as the classifier.

4.1 Setup

We collected 8000 mono 16bit quantization WAV audio files with 44.1kHz, including songs, speeches in different languages, such as English, Chinese and French. Each file has a duration of 10 seconds. We used Hide4PGP, S-Tools, StegoMagic and Xiao Steganography to

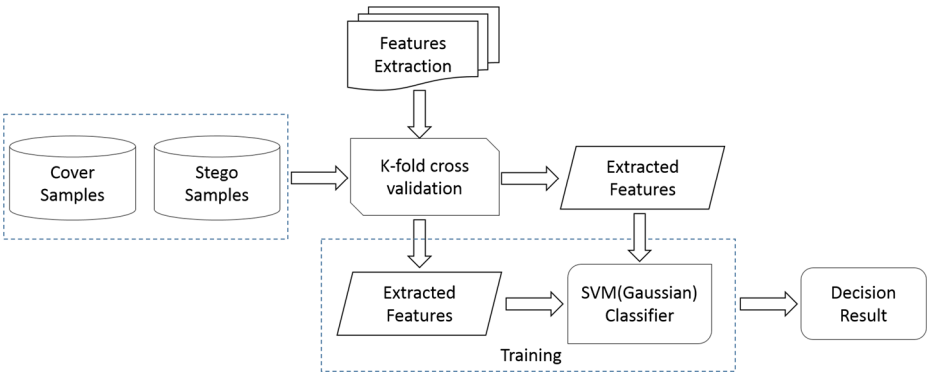


Fig. 3 The processing phrases of support vector machine

embed the hidden data into the covers respectively. The hidden data includes image, text, doc, executable file, voice, keyboard input messages, etc. They are of different size from several bits to the maximum hiding capacity. Therefore, the hidden ratio has a great range. S-Tools, StegoMagic and Xiao Steganography provide keys for hiding, while Hide4PGP not.

Since different steganography algorithm has different maximum hiding capacity (the maximum size of hidden message). The hidden ratio or embedded ratio means the hidden rate to the maximum capacity (hidden message size over maximum hidden message size), this concept does not accurately reflect the relation between the hidden message size and the cover size. Most of the audio steganalysis methods use hidden ratio to present the hiding strength. However, we think this can not reflect the relation between the hidden message and the cover, which is not a universal measurement among different steganography algorithms. From this point of view, we use the concept of hidden message to cover ratio (HCR), the ratio is calculated by using hidden message size over cover size, given in (17), regardless of the maximum hidden message size for each steganography algorithm.

$$HCR = \frac{hidden\_size}{cover\_size} \tag{17}$$

To measure the efficiency of our proposed method, we conducted our experiments using different datasets. For each hiding algorithm detection, we processed the samples to get the stegos. Four group of features are extracted from covers and stegos. A SVM with 4-fold cross validation was employed as the classifier, every test result is the mean value of 4 times decision results according to the 4-fold cross validation. In our experiment, we used accuracy (ACC) to describe our decision result. The accuracy is calculated according to (18).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

TP: True Positive, the occurrence that a stego is classified as a stego.
TN: True Negative, the occurrence that a cover is classfied as a cover.
FP: False Positive, the occurrence that a cover is classified as a stego.
FN: False Negative, the occurrence that a stego is classified as a cover.

Previous stegnalysis methods illustrate the accuracies according to the hidden ratios. However, in practice, before we analyze the suspicious files, we do not know the exact hidden ratio, so this kind of evaluation only gives an experimental evalution of a steganalysis method. How can we evaluate a steganalysis method in a practical case when we do not know the hidden ratio? To close this gap, in this paper, we proposed a general evaluation method. We used a dataset in which the HCR is roughly uniformly distributed, the maximum hidden message is close to the maximum hidden ratio and the minimum hidden message is just several bits input from the keyboard, which include almost all possible HCRs in reality, Therefore we can give a general evaluation of a steganalysis method on a dataset with wide-ranged HCR. We introduced a general detection score to estimate the performance of a steganalysis method, the score is the accuracy obtained from a general test dataset.

4.2 Experiment result

In our experiment, we attacked four steganography algorithms, Hide4PGP, S-Tools, StegoMagic and Xiao Steganography. For every hiding algorithm, we used 2000 covers to

produce 2000 stegos. We compared our method with D-MC [11] and 2D-Mel [12]. In Table 2, we used HCR to replace the hidden ratio, which is convenient to compare the detection performance among different hiding algorithms. The HCRs for four hiding algorithms are all close to the maximum hidden ratio, as a general belief, a higher hidden ratio leads to a better performance. The classification accuracies according to HCRs for four steganography algorithms are shown in Table 2. As is shown in Table 2, our method has an obvious advantage over the previous D-MC and 2D-Mel methods.

Table 2 suggests that when the HCR is high, the analysis will get a good detection accuracy. Take Hide4PGP as an example, the HCR is 0.2485, which means that the hidden message is 0.2485 of the cover, in our experiment, the cover size is 861KB, for Hide4PGP, the hidden message is 214KB. So as the other three hiding algorithms, the hidden message for S-Tools is 205KB, for StegoMagic is 94KB and for Xiao Steganography is 97KB. When the HCR is close, S-tools and Xiao Steganalysis show a weaker resistance compared with Hide4PGP and StegoMagic respectively. To this extent, the HCR can be used to compare the resistance among hiding algorithms.

Figure 4 provides the Receiver Operating Characteristic (ROC) curves on four steganography algorithms using three different steganalysis methods.

Receiver Operating Characteristic (ROC) curve is a direct and natural tool to select possibly optimal models. The curve is created by plotting the true positive (TP) rate against the false positive (FP) rate at various threshold settings. The diagonal divides the ROC space. Points above the diagonal represent good classification results (better than random), points below the line represent poor results (worse than random), the curve is more close to the left corner, it will indicate a better performance. As is shown in Fig. 4, the ROC curve of our linear prediction based method is more approximate to the upper left corner than the other two methods, it has a better performance than D-MC and 2D-Mel methods.

Additionally, We set our general test dataset, in which, the HCR is uniformly distributed. In every dataset for different steganography algorithm, the hidden message size is almost uniformly distributed. Take Hide4PGP for example, the HCR ranges from 0.0001 to 0.2485, we used 2000 covers and stegos to create the distribution, so as the other three steganography algorithms. For S-Tool, the HCR ranges from 0.0001 to 0.2308, 2000 covers and stegos, for StegoMagic the HCR ranges from 0.0001 to 0.1090 and for Xiao Steganography the HCR ranges from 0.0001 to 0.1120. As is shown in Table 4, we compared our method with D-MC and 2D-Mel, experiment results show that our method gains higher detection scores.

In Table 3, we set the detection accuracy as the score under the general evaluation according to a wide-ranged HCR. This score provides a general evaluation on the performance of steganalysis methods. Using this evaluation, we can objectively evaluate the performance of a steganalysis method. Consider this scenario, when a steganalysis method delivers a 99% accuracy with HCR at 0.2500, but in practice, the suspicious audio files probably have a

Table 2 The detection accuracy using three methods on four different hiding algorithms

Hiding Algorithm	HCR	D-MC	2D-Mel	LP Based
Hide4PGP	0.2485	94.6%	97.8%	**98.3%**
S-Tools	0.2380	96.7%	98.3%	**99.5%**
StegoMagic	0.1090	90.4%	92.9%	**98.2%**
Xiao Steganography	0.1120	91.3%	94.2%	**96.2%**

The best detection accuracy result are shown in bold

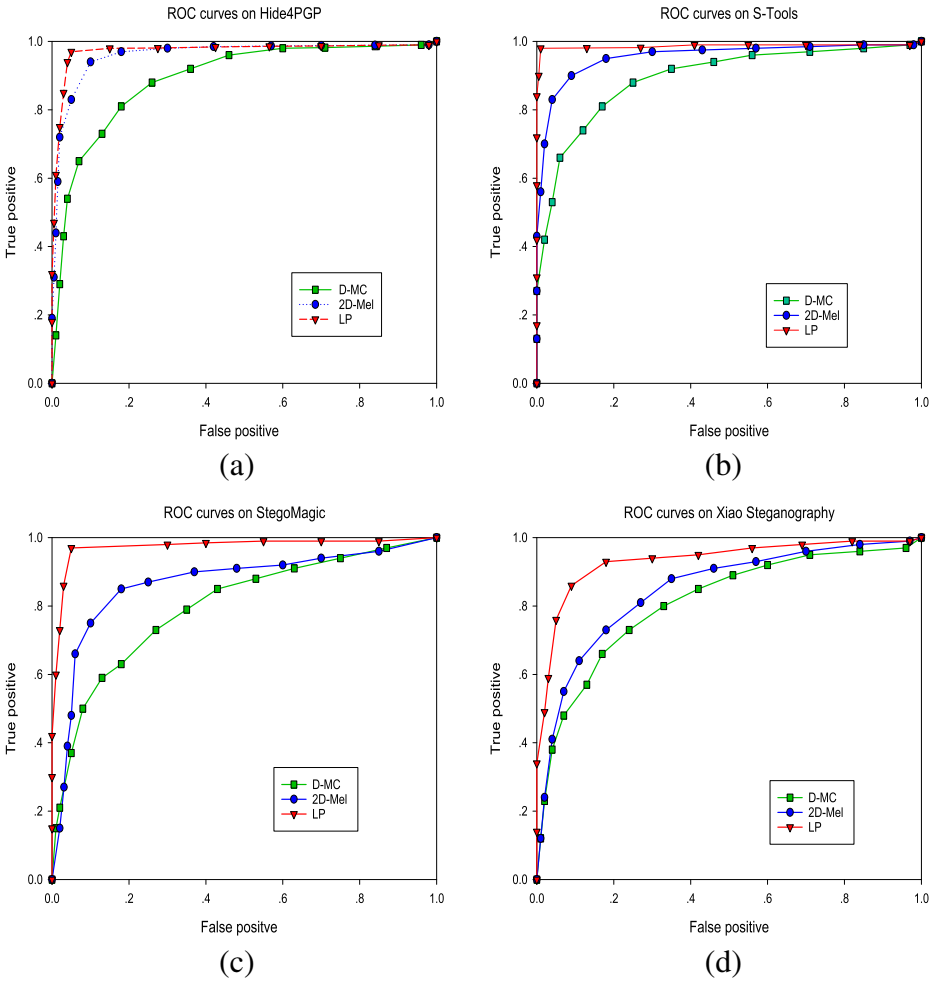


Fig. 4 ROC curves on four steganography algorithms using three different steganalysis methods

very small HCR. As a result, when we do not know the exact HCR, how can we evaluate the performance of some particular steganalysis method? According to this evaluation, the HCR ranging from 0.0001 (the hidden message is only several bits) to the maximum hidden ratio (according to different hiding algorithms), we obtain a general detection accuracy, we set this accuracy as the score of the steganalysis method. To our belief, a steganalysis method with a higher score will has a better general detection performance.

4.3 Efficiency

We used 1000 pair of covers and stegos processed by Hide4PGP algorithm with consistent hidden message to cover ratio (HCR) as 0.2485 to investigate the processing time of proposed linear prediction method and the other two prestigious methods: D-MC and 2D-Mel. Table 4 shows the configuration of our experiment environment. Table 4 also shows the processing time using a support vector machine loaded a 4-fold cross validation method as the

Table 3 The general score of three methods on four different hiding algorithms

Hiding Algorithm HCR	Hide4PGP 0.0001-0.2485	S-tools 0.0001-0.2380	StegoMagic 0.0001-0.1090	Xiao Steganography 0.0001-0.1120
D-MC	83.1	82.0	79.6	78.5
2D-Mel	85.2	84.7	85.4	83.6
LP Based	**88.1**	**90.6**	**88.4**	**87.1**

The highest general score are shown in bold

classifier. The time includes extracting features, training the machine and classifying the test dataset.

As shown in Table 4, our linear prediction based method has a slight overhead than D-Mel but uses less time than D-MC method. Notice that, this slight extra overhead averaged to a single audio file is negligible. Considering the high accuracy brought by linear prediction based method, we believe that this extra overhead will not impose limitations on the practical use of our linear prediction based method.

5 Discussion

In this section, we considered a wavelet transform on audio signal before linear prediction as a trial and tested the influence to detection accuracy by using different linear prediction solutions. In addition we tested our method on MP4 files, experiment results show that our method can deliver a good performance attacking Huffuman codebook based staganography algorithms. At the end of the discussion, we used Artificial Neural Network (ANN) as an optional classifier to conduct our experiment and presented its performance.

5.1 Wavelet transform before linear prediction

We employed linear prediction on wavelet coefficients as a trial. We tried different wavelet functions, such as ‘db2’, ‘db8’ and ‘haar’. We found that when using ‘db8’ wavelet function, we obtained more obvious differences between covers and stegos and a better detection performance. A ‘db8’ wavelet transform was done on the audio signal, then followed by feature

Table 4 Configuration of experiment environment and the processing time for linear prediction based method and two other methods (D-MC and 2D-Mel)

Parameters	Configuration		
System	Windows 7 pro		
RAM	4.00GB		
Processor	Intel(R) Core (TM) i7-4790 CPU 3.60GHz		
Software	MATLAB2016A		
Audio File Size	861KB		
Number of sample	2000		
Classifier	SVM with 4-fold cross validation		
Method	D-MC	2D-Mel	LP
Average Time	60.80 s	45.45 s	48.10 s

extraction procedure on the detail sub-band coefficients to obtain the LPC, LP residual, LP spectrum and LPCC as features.

We performed another experiment, extracted the features from wavelet coefficients, and the other parts of the experiment were the same as what had been used in Section 4.

The wavelet based linear prediction did not bing consistent increase or decline on detection accuracy. As for S-Tools, the mean accuracy value increases by 0.2%, Xiao Steganagrphy is 0.7%, StegoMagic is 1.3%. On the contrary, the test accuracy for Hide4PGP decreases by 0.5%. So we suppose that the wavelet transform will not bring a great difference to our steganalysis method.

5.2 Different solution to linear prediction

We utilized a lattice based model to solve the linear prediction equation for an exploration. We replaced the autocorrelation method which was used in our scheme with a lattice based method. There are several kinds of lattice, such as forward lattice, geometric mean lattice, Burg, etc. We chose the geometric mean lattice for a replacement. The LPCs obtained by autocorrelation method and geometric mean lattice method (GML) are a little different, which are represented in Table 5.

Though there are some difference in LPC between autocorrelation method and geometric mean lattice method, they are very close in power spectrum. Figure 5 shows the power spectrum comparison between autocorrelation method and geometric mean lattice method. In our experiments, we used autocorrelation method and geometric mean lattice method in linear prediction. In the same way, the classification features, samples and hiding algorithms were the same as what had been used in Section 4. The experiment results show that the different solutions for linear prediction do not bring an obvious difference in detection accuracy.

5.3 Analysis of Other File Formats

For the general versatility of our linear prediction based method, theoretically, our method can be used to analyze several kinds of audio files. Table 6 lists eight audio formats that can be analyzed by our method. However, it is remained to be investigated whether our linear prediction based method can be effective enough for these format audio files. For example, MP3Stego is a steganography tool for MP3 files. This steganography algorithm has a very low hidden ratio, in this case, our method will bring inferior performance than some excellent steganalysis methods [32, 34] designed for attacking MP3Stego. On the other hand,

Table 5 LPC of the cover and the stego by using different linear prediction solutions

Order	Autocorrelation	GML	Order	Autocorrelation	GML
a_1	-0.0500	−0.0310	a_7	−0.1742	−0.1847
a_2	−0.1042	−0.0856	a_8	−0.0252	−0.0329
a_3	0.0162	0.0508	a_9	−0.0717	−0.1010
a_4	−0.1207	−0.1095	a_{10}	−0.0560	−0.0784
a_5	−0.2061	−0.1957	a_{11}	0.0071	−0.0005
a_6	−0.0807	−0.0930	a_{12}	0.0089	−0.0050

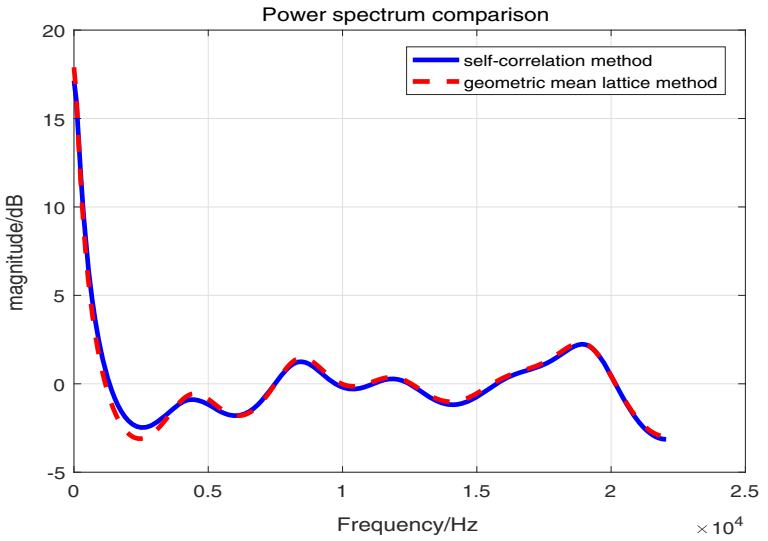


Fig. 5 The power spectrum comparison between autocorrelation method and geometric mean lattice method

our method can deliver good performance as some steganalysis methods designed for AAC (MP4 files).

We tested MP4 files (AAC, Advanced Audio Coding), which are widely used on YouTube, iPhone, Android and some audio systems equipped in cars. There have been many kinds of steganography tools for AAC files. While some are not resistant to analysis methods or bring distortion to the audio quality. Huffman codebook steganography schemes [25, 37] proposed recently are better steganography tools, which will not introduce any distortion to audio quality and have good imperceptivity. At a result, this kind of steganography tools are more challenging to steganalysis methods. We tended to use our linear prediction based method to attack this kind of steganography tools.

Because this kind of steganography tools for AAC have not been developed into software and the source code is disclosed. We need to develop a program to accomplish Huffman codebook steganography tool firstly, then using this tool to produce stegos for our experiment. We would like to reserve this work as our future work. For the same reason, we can not provide the comparison with other steganalysis methods for AAC audio files. Since we need to develop programs to accomplish their methods and run these methods on the same data sets. But we have ask some pair of samples (cover and stego) from the authors of work [22] and we tested our method on a small set of samples.

Table 6 File formats that our linear prediction based method can work on

Format	Format
WAV (.wav)	AIFF (.aiff, .aif)
OGG (.ogg)	AIFC (.aifc)
FLAC (.flac)	MP3 (.mp3)
AU (.au)	MPEG-4 AAC (.m4a,.mp4)

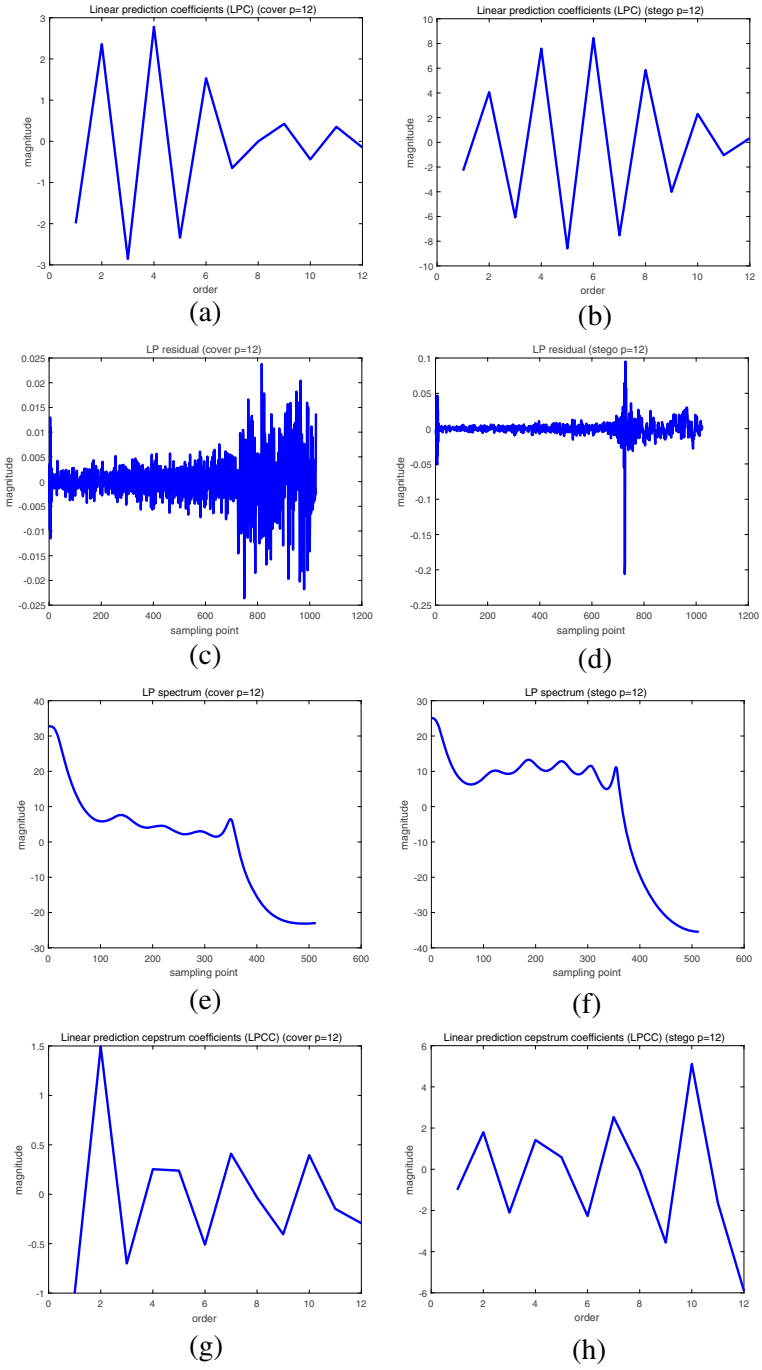


Fig. 6 The difference of linear prediction based features between a cover audio signal (left) and its associated stego audio signal (right), including the comparison of LPC shown in (**a**) and (**b**), the difference of LP residual given in (**c**) and (**d**), the difference of LP spectrum shown in (**e**) and (**f**) and distinction in LPCC shown in (**g**) and (**h**). The cover and stego MP4 audio files are 128kbps and the relative embedding rate (RER) is 1

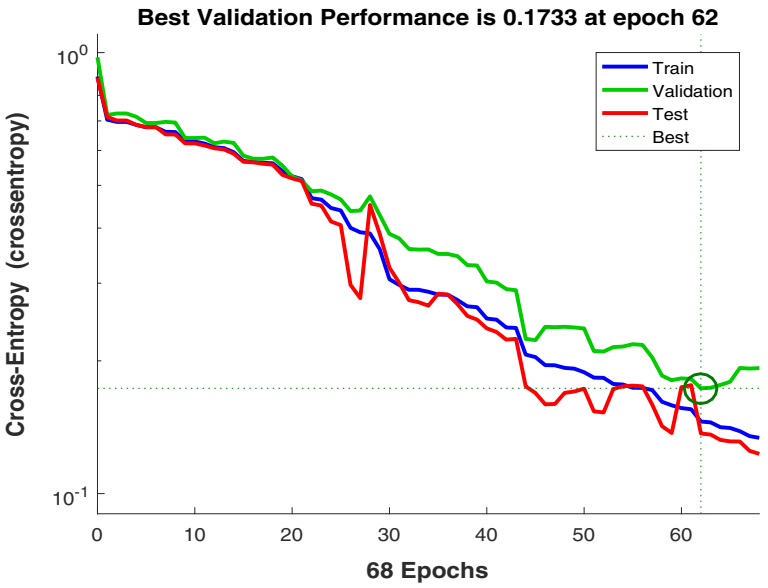


Fig. 7 The best validation performance

Figure 6 shows the feature differences between a cover and its associated stego using our linear prediction based method. The cover and the stego audio files is 128kbps and the relative embedding rate (RER) is the maximum hidden ratio.

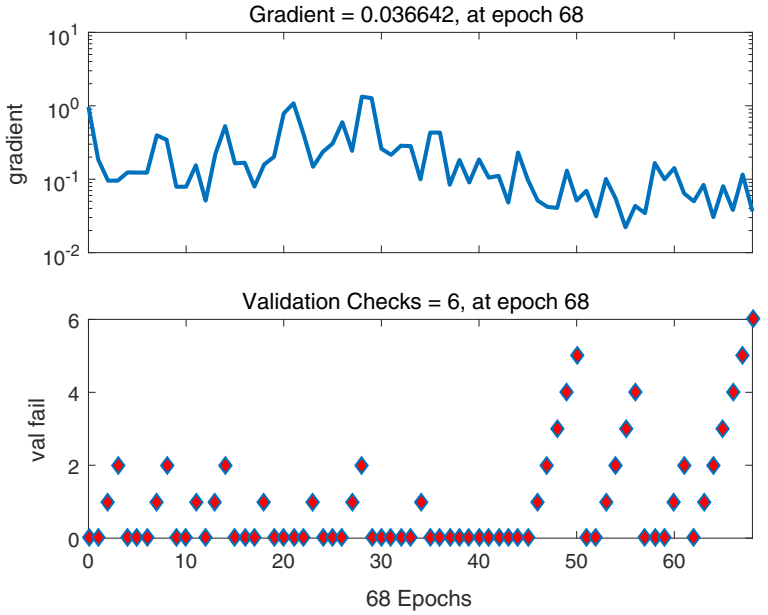


Fig. 8 The process of training

Linear prediction based feature distinctions between the cover and the stego are obvious. From this point of view, we believe our linear prediction based method is also effective to analyze AAC (MP4) audio files. We obtained 50 pair of covers and stegos from authors in [22] and the samples are 128 kb per second, the embedding rate in 1 (means the hidden messages reach the maximum hiding capacity), our method achieves a 100% accuracy, the same accuracy as method in [22].

5.4 Artificial neural network as the classifier

We tried to use Artificial Neural Network (ANN) as an exploration. We used a set of samples to employ ANN as our classifier. Firstly, we extracted the linear prediction based features and labeled every cover as 1 and stego as 0. Then we used the Pattern Recognition Tool as our model. We test 1000 samples, 500 covers and 500 stegos produced by Hide4PGP, with hidden message size is 214KB, cover size is 861KB and the HCR (hidden message to cover ratio) is 0.2485. We used 700 samples as training samples, 150 samples as validation samples and 150 samples as testing samples. The number of layers is 2 and there are 10 neurons in every layer. Figure 7 shows the performance of the ANN classifier in our experiment. The best validation performance occurs at epoch 15 and the train processing is followed in Fig. 8. Figure 9 shows the error histogram of the classification in this ANN. As shown in Fig. 9, the error is small and close to zero error line. Figure 10 shows the confusion matrix of the ANN, which is a criterion for perception classifier, we can see that there are some situations that a 0 labeled sample is classified as 1 labeled sample and a 1 labeled sample is classified as a 0 labeled sample. Figure 11 shows the ROC (Receiver Operating Characteristic) curve of this ANN classifier. The curve is more close to the left corner, it will indicate

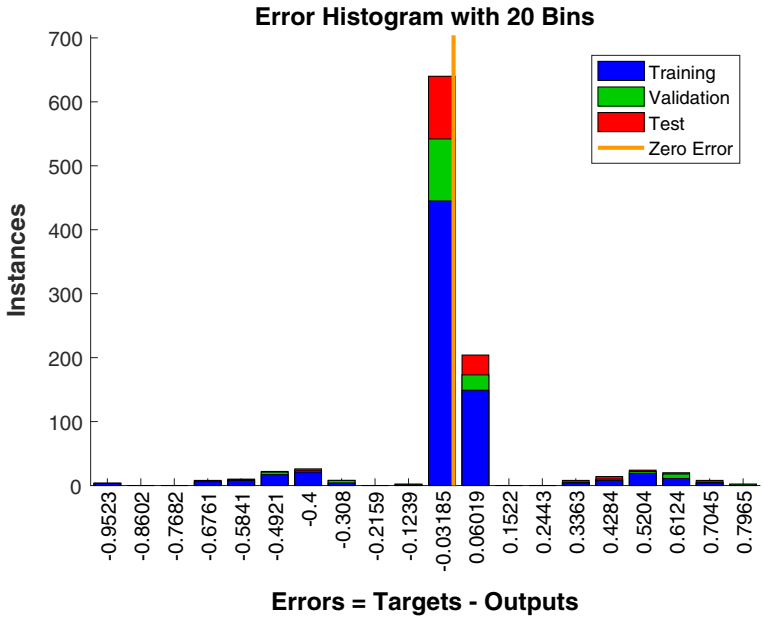


Fig. 9 The Error histogram of the classification

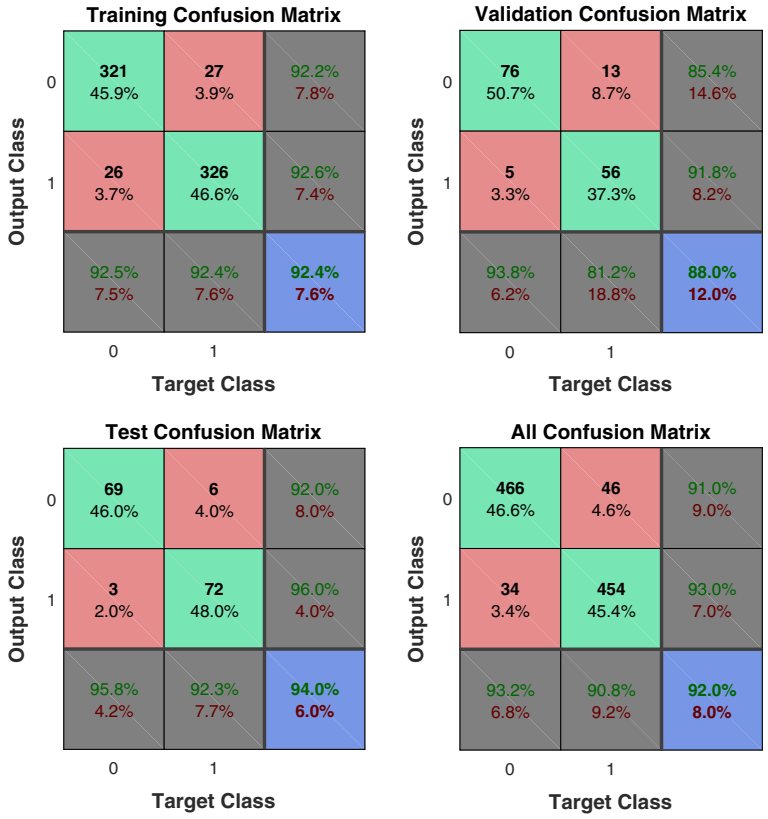


Fig. 10 The Confusion Matrix for training, validation and test samples, and a confusion matrix for all samples

a better performance. The curve gained in this ANN classifier is approximate to the upper left corner and indicates a good performance.

6 Future work

As shown in discussion part, our linear prediction based method can also work on ACC (MP4) files. In our future work, we will conduct experiments to show the performance of our method and give comparisons with other steganalysis methods for AAC (MP4) files. Additionally, we will try to detect other format audio files, such as AU and AMR files and present our experiment results.

In our support vector machine, we used a k-fold cross validation method to find the best parameters and to improve the detection accuracy. According to the advance in SVM, an adaptive method could be a better choice. We investigated some research papers about adaptive methods to find the optimal parameters of SVM in machine learning field [1, 2, 5, 13, 28, 29, 36]. We tend to use Genetic Algorithm based SVM (GA-SVM), using genetic algorithm to find the optimal parameters. We would like to reserve this improvement as our future work.

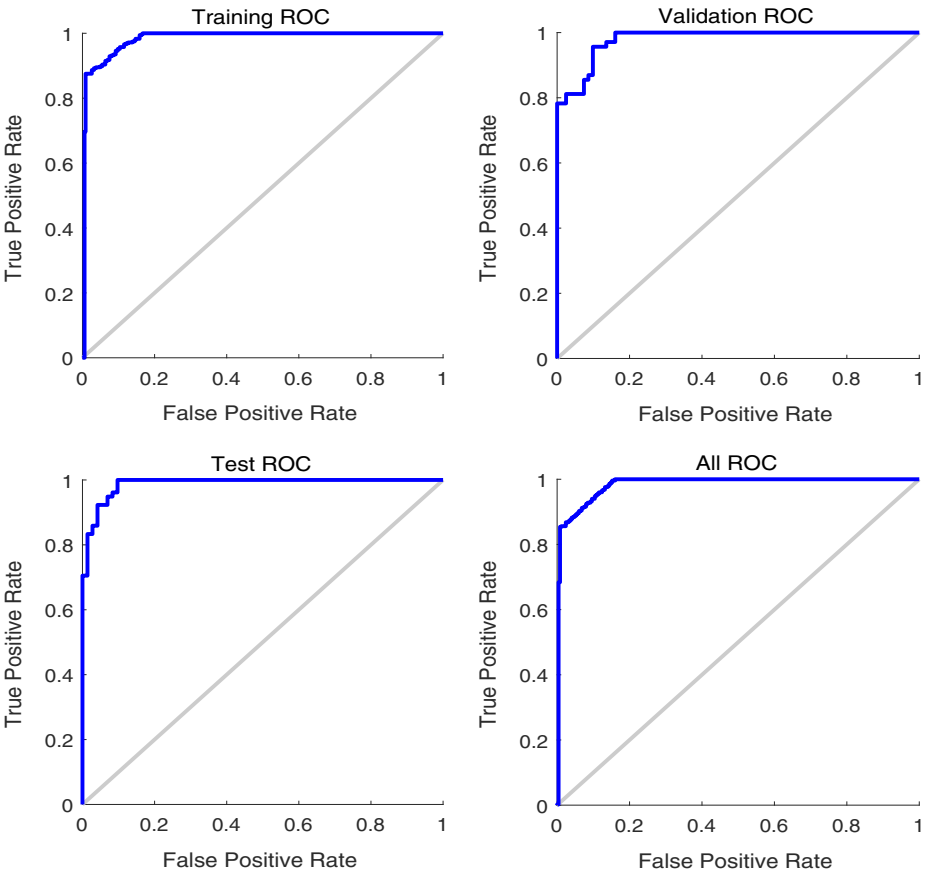


Fig. 11 The Receiver Operating Characteristic curve (ROC) of the ANN classifier

7 Conclusion

In this paper, we proposed a linear prediction based audio steganalysis method, introducing linear prediction technique into audio steganalysis, which suffices to bring a significant difference between the cover and the stego. By extracting linear prediction based coefficients, applying SVM as the classifier, our experiments ended up with excellent performance attacking four steganography algorithms. Additionally, a hidden message to cover ratio is proposed in this paper to provide a uniform criterion to compare the detection performance among hiding algorithms. A general evaluation method was used to practically estimate steganalysis methods. In discussion part, we tested the influence brought by wavelet transform and different solution to linear prediction equation. Experiments results show that the effect is negligible. In addition, we tested our method on another audio file format and tried to use another machine model as an optional classifier, our test experiments delivered good performance. This linear prediction based steganalysis method maybe an essential and foundational insight and can bring improvement to audio steganalysis field.

Acknowledgements The authors are supported by National Natural Science Foundation of China (No.61402471, 61472414). We wish to thank Professor Tang and Professor Zuo for their substantial support, insightful comments and suggestions, Dr. Chen Gong, Dr. Hailong Zhang and Professor Li for their discussion and guidance. Special thanks goes to Mr and Mrs Han for their understanding and support.

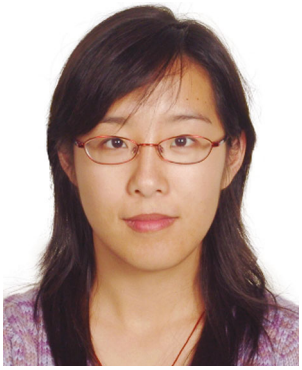
References

1. Avci E (2009) Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithmCsupport vector machines: HGASVM. Expert Syst Appl 36(2):1391–1402
2. Couellan N, Jan S, Jorquera T (2015) Self-adaptive support vector machine: A multi-agent optimization perspective. Expert Syst Appl 42(9):4284–4298
3. Desai MB, Patel SV, Prajapati B (2016) ANOVA and fisher criterion based feature selection for lower dimensional universal image steganalysis. Int J Image Process (IJIP) 10(3):145
4. Djebbar F, Ayad B (2012) Audio steganalysis based on lossless data-compression techniques. In: International Conference on Information and Communications Security. Springer, Berlin Heidelberg, pp 1–9
5. Du SC, Huang DL, Wang H (2015) An adaptive support vector machine-based workpiece surface classification system using high-definition metrology. IEEE Trans Instrum Measur 64(10):2590–2604
6. Geiser B, Vary P (2008) High rate data hiding in ACELP speech codecs. In: IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE Xplore, pp 4005–4008
7. Ghasemzadeh H, Arjmandi MK (2014) Reversed-Mel cepstrum based audio steganalysis. In: 4th International eConference on Computer and Knowledge Engineering (ICCKE). IEEE, pp 679–684
8. Ghasemzadeh H, Khass MT, Arjmandi MK (2016) Audio steganalysis based on reversed psychoacoustic model of human hearing. Digit signal process 51:133–141
9. Kraetzer C, Dittmann J (6505) Mel-cepstrum based steganalysis for VoIP steganography. Proc SPIE - Int Soc Opt Eng 6505:650505-650505-12
10. Liao X, Shu C (2015) Reversible data hiding in encrypted images based on absolute mean difference of multiple neighboring pixels. J Vis Commun Image Represent 28:21–27
11. Liu Q, Sung AH, Qiao M (2009) Temporal derivative-based spectrum and mel-cepstrum audio steganalysis. IEEE Trans Inf Forensic Secur 4(3):359–368
12. Liu Q, Sung AH, Qiao M (2011) Derivative-based audio steganalysis. ACM Trans Multimed Comput Commun Applicat (TOMM) 7(3):18
13. Liu D, Niu D, Wang H (2014) Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 62:592–597
14. Makhoul J (1975) Linear prediction: A tutorial review. Proc IEEE 63(4):561–580
15. Miao H, Huang L, Chen Z et al (2012) A new scheme for covert communication via 3G encoded speech. Comput Electr Eng 38(6):1490–1501
16. Molau S, Pitz M, Schluter R et al (2001) Computing mel-frequency cepstral coefficients on the power spectrum. In: 2001. Proceedings. (ICASSP'01). 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol 1. IEEE, pp 73–76
17. Nouri A, Nazari A (2016) Improving Image steganalysis performance using a graph-based feature selection method. Adv Comput Sci: Int J 5(3):33–39
18. Ozer H, Avcibas I, Sankur B et al (2003) Steganalysis of audio based on audio quality metrics. Int Soc Opt Photonics Electron Imaging 2003 2003:55–66
19. Qian Y, Dong J, Wang W et al (2016) Learning and transferring representations for image steganalysis using convolutional neural network. In: IEEE International Conference on Image Processing (ICIP). IEEE, pp 2752–2756
20. Qiao M, Sung AH, Liu Q (2013) MP3 audio steganalysis. Inf Sci 231(9):123–134
21. Ren Y, Cai T, Tang M et al (2015) AMR steganalysis based on the probability of same pulse position. IEEE Trans Inf Forensic Secur 10(9):1–1
22. Ren Y, Xiong Q, Wang L (2016) Steganalysis of AAC using calibrated Markov model of adjacent codebook. In: IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp 2139–2143

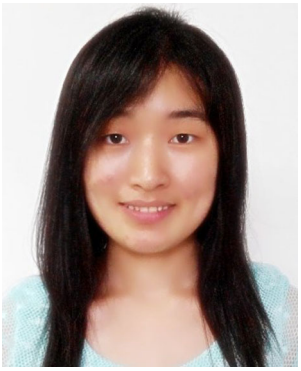
23. Ru XM, Zhang HJ, Huang X (2005) Steganalysis of audio: Attacking the steghide. In: Proceedings of 2005 International Conference on Machine Learning and Cybernetics, vol 7. IEEE, pp 3937–3942
24. Srivastava S, Nandi P, Sahoo G et al (2014) Formant based linear prediction coefficients for speaker identification. In: 2014 International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, pp 685–688
25. Tang BT, Guo L, Liu ZH (2008) An information hiding method in advanced audio coding (AAC). Tech Acoust 27(4):533–538
26. Tian H, Wu Y, Chang CC et al (2016) Steganalysis of adaptive multi-rate speech using statistical characteristics of pulse pairs. Signal Process 134:9–22
27. Tint Y, Mya KT (2012) Audio steganalysis using features extraction and classification. International Journal of Research and Reviews in Computer Science (IJRRCS) 3(2):1593–1595
28. Wu Q (2010) Power load forecasts based on hybrid PSO with Gaussian and adaptive mutation and Wv-SVM. Expert Syst Appl 37(1):194–201
29. Wu CH, Tzeng GH, Goo YJ (2007) A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. Expert Syst Appl 32(2):397–408
30. Wu A, Feng G, Zhang X et al. (2016) Unbalanced JPEG image steganalysis via multiview data match. J Vis Commun Image Represent 34:103–107
31. Xia Z, Wang X, Sun X, Liu Q, Xiong N (2016) Steganalysis of LSB matching using differences between nonadjacent pixels. Multimed Tools Appl 75(4):1947–1962
32. Yan D, Wang R, Yu X et al (2013) Steganalysis for MP3Stego using differential statistics of quantization step. Digit Signal Process 23(4):1181–1185
33. Yavanoglu U, Ozcakmak B, Milletsever O (2012) A new intelligent steganalysis method for waveform audio files. In: 11th International Conference on Machine Learning and Applications (ICMLA), vol 2. IEEE, pp 233–239
34. Yu X, Wang R, Yan D et al (2012) MP3 audio steganalysis using calibrated side information feature. J Comput Inf Syst 8(10):4241–4248
35. Yuan C, Xia Z, Sun X (2017) Coverless image steganography based on SIFT and BOF. J Internet Technol 18(2):435–442
36. Zhai S, Jiang T (2015) A new sense-through-foliage target recognition method based on hybrid differential evolution and self-adaptive particle swarm optimization-based support vector machine. Neurocomputing 149:573–584
37. Zhu J, Wang RD, Li J et al (2011) A huffman coding section-based steganography for AAC audio. Inf Technol J 10(10):1983–1988

Chunling Han received the B.S Degree from China University of Geosciences in Wuhan China 2015. She is currently a Ph.D candidate in State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security at University of Chinese Academy of Sciences. Her research interests include cryptography, theory of computation and information hiding.

Rui Xue received the M.S and Ph.D degrees from Department of Mathematics at Beijing Normal University, in 1988 and 1999. He is currently a researcher and doctoral supervisor in State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security at University of Chinese Academy of Sciences. His research interests include cryptography, security protocols, cloud data security and multimedia security.



Rui Zhang received a Ph.D degree in Information Security from Beijing Jiaotong University, Beijing in 2011. She is now an assistant researcher at the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security at University of Chinese Academy of Sciences. Her research interests include cloud data security, privacy preservation and security protocols.



Xueqing Wang received the B.S. degree in information and computing science from Yunnan University in 2013. She is currently a Ph.D candidate in the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, School of Cyber Security at University of Chinese Academy of Sciences. Her research interests are public-key cryptography, multimedia security and security protocols.