Steganalysis of Adaptive Multiple-Rate Speech Using Parity of Pitch-Delay Value

Xiaokang Liu[1], Hui Tian[1(✉)], Jie Liu[1], Xiang Li[1], and Jing Lu[2]

[1] College of Computer Science and Technology,
National Huaqiao University, Xiamen 361021, China
htian@hqu.edu.cn
[2] Network Technology Center,
National Huaqiao University, Xiamen 361021, China

Abstract. Exploiting the fact that the pitch period parameter in speech parameter encoding is difficult to predict, a large number of steganographic strategies choose to embed secret information in the pitch period. Several detection methods for these steganography strategies based on the pitch period have also been proposed so far, but it is still a challenge to detect the steganography accurately. In this work, a new steganalysis scheme is proposed to detect pitch period based steganography, which has lower complexity and higher accuracy compared with the existing steganalysis schemes. Firstly, we regard a frame as a calculation unit within which the parity of four sub-frames can be obtained. Secondly, after filtering and merging into 14-dimensional PBP (parity Bayesian probability) features, these features are classified by the support vector machine (SVM). We evaluate the performance of the proposed strategy with numerous speech samples encoded by the adaptive multi-rate audio codec (AMR) and compare it with the state-of-the-art strategies. The experimental results illustrate that proposed method can effectively detect the pitch-delay based steganography. It is not only superior to the existing steganalysis methods in detection accuracy, but also has outstanding real-time detection performance and robustness because of its lower feature dimension and complexity.

Keywords: Steganalysis · Adaptive multi-rate codec · Pitch delay · Bayes's theorem · Support vector machine (SVM)

1 Introduction

Steganography is a common means of hiding secret information in the carrier without perceptible distortion, and it has been applied to very broad areas from war to politics since the ancient Greek era. The carrier of information (namely, the steganographic carrier) has been changing over the ages. Current steganography chiefly relies on networks such as the Internet protocols [1] and digital multimedia (text [2, 3], image [4–6] voice [7, 8], video [9]). In recent years, mobile device-based voice communication protocol has been greatly technologically advanced. Therefore, a large number of steganographic researches are attracted to the field of voice transmission [8, 10–12] because of not only the wide range of applications but also reliability real-time and considerable redundancy. AMR [13–15] has become a hotspot in steganography

J. Li et al. (Eds.): SPNCE 2019, LNICST 284, pp. 282–297, 2019.
https://doi.org/10.1007/978-3-030-21373-2_21

research because it is widely used in 3G and 4G networks. Moreover, AMR has considerable coding redundancy, which eventually makes it practical and efficient to apply AMR-based steganography into secure communication. However, on the one hand, the development of steganography provides a better choice for the safe transmission of data. On the other hand, it also provides a better choice for the illegal transmission of data. If the technology is exploited illegally, it will pose a threat to network security. Therefore, steganalysis, aimed at detecting steganography in the communication process, has increasingly received widespread attention.

AMR adopts an optimized compressed speech coding mode [16], and it is extensively employed by almost all cell phone. Specifically, AMR is a multi-rate ACELP (Algebraic Code Excited Linear Prediction) encoder with 8 modes, from 4.75 kbit/s to 12.2 kbit/s and sampling frequency of 8000 Hz. The process of mode integration and bit rate conversion is mainly operated by changing the quantization parameters, which provides seamless switching of 20 ms frame boundaries [17]. Additionally, the encoding method of AMR leads to a large amount of redundancy in the encoding, which makes AMR an ideal voice steganographic carrier. Correspondingly, steganalysis for AMR also evolves with steganography [18–21].

As described above, AMR is a typical algebraic code excited line prediction (ACELP) coding. The structure of ACELP mainly includes Liner Prediction Coefficient (LPC) analysis, Fixed Code Book (FCB) searching and Adaptive Code Book (ACB) searching. The adaptive codebook is to match the pitch period, and the fixed codebook is founded on the algebraic codebook search. Although ACELP has been developed for quite a long time and many prediction algorithms have been proposed for it, these parts of predictive coding are still hard to be accurate. The unavoidable redundancy provides favorable conditions for steganography. In the LPC analysis, Liu et al. [22] proposed a new quantized index modulated LPC steganography algorithm. Not only does it improve the efficiency, but it also reduces the distortion of speech. After that, Liu et al. [22] also proposed a new method based on matrix embedding information, and the method greatly improves security and the efficiency of embedding. Similarly, there are also many steganography schemes for the fixed codebook [11, 13, 23, 24]. Geiser et al. [13] introduced a method of hiding data at a higher rate in the FCB, in which the information bandwidth can reach 2 kbit/s. Based on the similar principle, Miao et al. [15] proposed a 3G speech encoding steganography scheme, which bases on the Adaptive Suboptimal Pulsed Combination Constraints (ASOPCC) method. Compared with the linear prediction analysis and fixed codebook searching, the adaptive codebook searching is more flexible and the range of the pitch period of speech is wider when considering the complexity of the speech itself. The redundancy caused by inaccurate predictions drives the development of steganography in recent years. Based on the AMR pitch delay, Huang et al. [25] presented a new steganography scheme. It divides the adaptive codebook into two parts, and then introduces a random location selection to adjust the embedding rate dynamically and improve the security of steganography. The experimental results show that not only are the considerable capacity and real-time performance ensured, but also the quality and the anti-detection ability of the steganographic speech remain high. Based on Huang, Yan et al. [26] proposed a twice-layer steganography algorithm using low-rate speech as a carrier. The first layer of steganography is implemented by limiting the search set of the pitch period value of the speech sub-frame. The second layer of steganography is

implemented by exploiting the randomness of the pitch period in the search set. In the process of twice-layer embedding, the value of the pitch period is determined by the principle of minimizing the amplitude of modification. The advancement of steganography directly promoted the generation and the development of another opposite technology-steganalysis.

Compared with steganography the development process of steganalysis is always lagging. However, there are still massive valid works have been proposed [18, 21, 22, 27–30]. For the steganography on the LPC, Li et al. [28] observed that the correlation properties of the split vector quantization (VQ) codebook of linear predictive coding filter coefficients changed after the QIM steganography. Based on this observation, they construct the QCCN (Quantitative Codebook Correlation Network) model and obtain the eigenvector after quantifying the fixed-point related features of the pruned network. These steganalysis methods have got great feedback. Tian et al. [21] employed probability distribution of pulse pairs as the long-term distribution features and employed Markov transition probability matrix of pulse pair as the short-term invariant features. Moreover, adaptive boosting was introduced to optimize these features, and finally, the feature classification results obtained are superior to the existing detection methods. Some efficient detection methods for coping with the new pitch delay steganography was also proposed. Li et al. [31] proposed a method for detecting quantization index modulation (QIM) steganography in a G.723.1 bit stream. They extract these eigenvectors based on the correlation and the imbalance of each quantized index (codebook) distribution in the quantized index sequence. Based on the correlation and the imbalance of each quantized index (codebook) distribution in the quantized index sequence, a kind of novel eigenvector is extracted, and then the extracted features are combined with support vector machines to construct a classifier for detecting QIM steganography in the G.723.1 coding stream. Experiments show that this method has achieved good results in detecting steganography. However, Ren et al. [20] proposed a new steganalysis scheme for AMR speech and achieved better results. Based on the differences in the continuity of the adjacent pitch delays between the original and steganographic AMR speech, they calculated the second-order Markov transition probability (MSDPD) feature matrix and then obtained C-MSDPD by subtracting the MSDPD after calibration. Experiments show that the effect of C-MSDPD is better than Li et al. and is the most efficient detection method for pitch delay steganography presently.

The rest of this paper is organized as follows. To make this paper self-contained, Sect. 2, firstly, introduces the structure of AMR codec and the principle of the state-of-the-art steganalysis. Section 3 explores the steganalysis features based on the parity of the pitch delay. The steganalysis scheme is revealed in Sect. 4, which is followed by the evaluation and experimental results that are presented in Sect. 5. Finally, the conclusions are drawn and directions for further work are suggested in Sect. 6.

2 Background and Related Work

In this section, the principle of the pitch delay searching in AMR codec is introduced. Firstly, the defects of the coding principle exploited by steganography will be explored. Secondly, the state-of-the-art steganography and steganalysis schemes based on the pitch delay are introduced in detail.

2.1 The Principle and Analysis of AMR Codec

The structure of AMR mainly consists of linear prediction, adaptive codebook searching, fixed codebook searching, gain quantization, post-processing and error concealment. The fundamental function of the linear prediction is to obtain the 10 coefficients of a 10-order LPC filter, and convert them into line spectra to quantify the parameters LSF. Adaptive codebook searching, including open-loop pitch analysis and closed-loop pitch analysis, obtains the pitch delay and the pitch gain. While fixed codebook searching, including the quantization of the quantization, is to obtain algebraic codebook gain [32]. AMR coding structure is shown in Fig. 1.

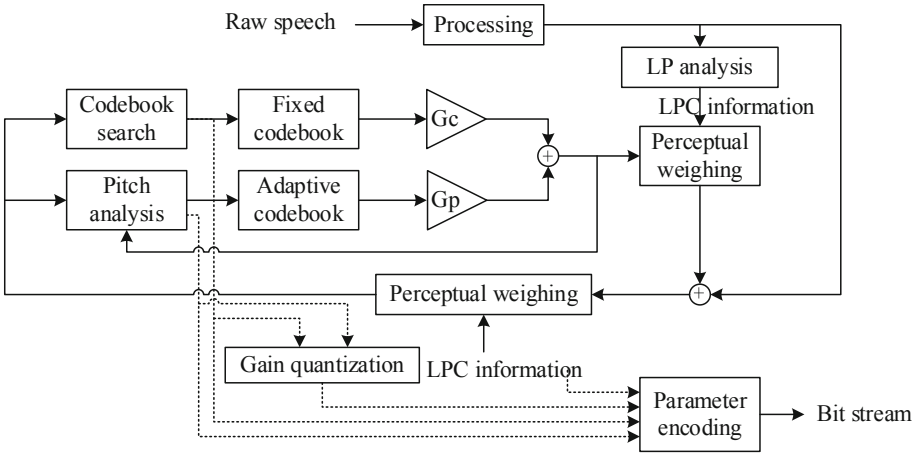


Fig. 1. AMR structure

AMR sets the speech rate to 8 kHz as the sampling rate. One frame has 240 sampling points and is divided into four sub-frames (T_0, T_1, T_2, T_3). In order to illustrate the AMR search principle more clearly, we take the case of 12.2 kb/s mode. The open-loop pitch delay T_{OP} is calculated based on a weighted speech signal which is the output of the original signal input to the perceptual weighting filter. The purpose of predicting the open-loop pitch period is to reduce the computational complexity of the closed-loop pitch period. Then the closed-loop pitch delay can be calculated by the signal f and the cross-correlation formula $C_{OL}(j)$ searching a maximum within a certain range around the open-loop pitch delay. The formula is as follows.

$$C_{OL}(j) = \frac{\left(\sum_{n=0}^{119} f[n]f[n-\mathrm{j}]\right)^2}{\sum_{n=0}^{119} f[n-j]f[n-\mathrm{j}]} \; 18 \le j \le 142. \tag{1}$$

The coefficient j of the maximum $C_{OL}(j)$ is selected as the open-loop pitch within the search range according to formula (1). The closed-loop pitch T_0 and T_2 are searched on the basis of the open-loop pitch, then the closed-loop pitch of T_1 and T_3 are calculated based on the closed-loop pitch delay of T_0 and T_2 which have been obtained. The range of the pitch delay for T_0, T_2 is determined by formula (2).

$$\mathrm{T}_0 = \begin{cases} [18, 24], & \mathrm{T}_{OP} \le 21 \\ [\mathrm{T}_{OP} - 3, \mathrm{T}_{OP} + 3], & 21 \le \mathrm{T}_{OP} \le 140 \\ [137, 143], & \mathrm{T}_{OP} > 140 \end{cases}, \tag{2}$$

where T_{OP} is the open-loop pitch, and T_0 is the first sub-frame of closed-loop pitch.

$$\mathrm{T}_1 = \begin{cases} [18, 27], & \mathrm{T}_0 \le 23 \\ [\mathrm{T}_0 - 5, \mathrm{T}_0 + 4], & 23 \le \mathrm{T}_0 \le 139 \\ [134, 143], & \mathrm{T}_0 > 139 \end{cases}, \tag{3}$$

where T_1 is the second sub-frame of the closed-loop pitch. From (3), T_1 and T_3 are searched based on T_0 and T_2. According to the nature of speech, the correlation between the pitch delay of adjacent sub-frames in a frame is quite stable, especially the pair of (T_0, T_1) and (T_2, T_3). Therefore, the probability of a change in parity between T_0 and T_1 or T_2 and T_3 is less than the probability of invariance, which is also confirmed in later experiments.

2.2 The Principle of Pitch-Based Steganography

The accurate prediction of the pitch delay in speech coding is still a challenge, even though many algorithms have been contributed to it. Nevertheless, this uncertainty can be exploited to design steganography algorithms. According to the characteristics of pitch delay searching, Huang et al. [25] divided the adaptive codebook into two parts. One of them contains only odd numbers and the other just even numbers. The closed-loop pitch is calculated by (1) and (2), yet a restriction is added when searching for the closed-loop pitch so that mod (T_t, 2) equals the secret information to be embedded. In the process of extracting secret information, the procedure is exactly the opposite, namely, adding the judgment mod (T_t, 2) to extract the secret information. T_t is the pitch delay of the t-th sub-frame. Yan et al. [24] proved that Huang's steganography could undermine the quality of speech at high embedding rates through experiments. Moreover, they discovered that if the changes were T_1, T_3 (T_0, T_2 unchanged), then there would be less impact on the quality of speech. In order to improve the embedding rate, they proposed a twice-layer steganography method, which calculates λ under the condition of Huang's T_0 and T_2 embedding.

$$\lambda = [(T_1 \bmod 4)/2 \oplus (T_3 \bmod 4)/2], \tag{4}$$

where T_1 and T_3 are determined by controlling the value of λ (0, 1) (T_1 is the second sub-frame closed-loop pitch value and T_3 is the fourth sub-frame closed-loop pitch value).

2.3 Review of the AMR-Based Steganalysis

After analyzing the variance of first-order difference and second-order difference of the pitch delay, Ren et al. [20] found that Huang's steganography method [25] makes obvious changes of some features before and after steganography, especially the second-order difference. Therefore, Ren et al. choose the second-order difference of the pitch delay as the classification feature.

$$D_T^2(t) = T_{t+2} - 2T_{t+1} + T_t, \tag{5}$$

where $D_T^2(t)$ is the second-order difference of the t-th sub-frame and T_t is the pitch delay of the t-th sub-frame. According to the inherent principle of speech, the changes of the adjacent second-order difference should be more concentrated and Markov transfer probability is expert in illustrating this connection. Therefore, the second-order difference construction is employed to construct Markov transition matrices as the classification characteristics.

$$M_{D_T^2} = \frac{\sum_{t=0}^{N-4} \delta(D_T^2(t), D_T^2(t+1) = i)}{\sum_{t=0}^{N-4} \delta(D_T^2(t) = i)}, \tag{6}$$

where $M_{D_T^2}$ is the transition probability of the current second-order difference to the next second-order difference. According to the experimental data, (–6, 6) is selected as the threshold. Therefore, there are 169 kinds of Markov transition probability.

What is special is that the appointment of a method called calibration [33] has greatly improved the accuracy of the MSDPD feature detection. The specific process is to divide the steganographic file into A and B parts. After decoding the part B, the original AMR encoder is re-encoded to obtain B_1, then the MSDPD obtained by decoding the extracted pitch delay of A and B_1. Finally, the subtraction of MSDPD of A and MSDPD of B_1 is

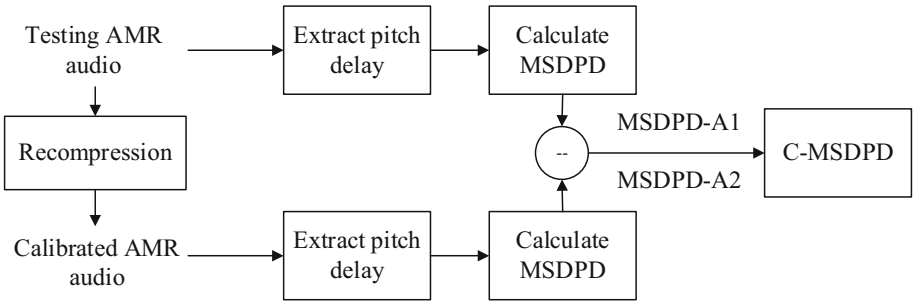


Fig. 2. C-MSDPD extraction process

C-MSDPD. The process is shown in Fig. 2. The experimental results show that using C-MSDPD as the classification feature by SVM can acquire the best results.

3 Background and Related Work

In this section, our Bayes's theorem features are illustrated below, which followed by the exact representation of our features. For the convenience of description, the AMR-NB at 12.2 kbps is taken as the example. Afterward, the characteristics, including advantages and disadvantages between the proposed features and C-MSDPD, are elaborated and compared.

3.1 Bayes's Theorem Features of Pitch Delay Parity

From the analysis of the pitch delay searching principle, the closed-loop pitches of T_1 and T_3 are searched based on T_0 and T_2. According to the analysis of the AMR coding principle above, the steganography methods, embedding secret information through changing the closed-loop pitch, distorts the connection between sub-frames. After determining the current sub-frame, the range of the next sub-frame pitch delay has narrowed and the possibility of the change of the parity has narrowed accordingly. Nevertheless, the distribution of the pitch delay value tends to be random under normal condition but concentrated under the condition of existing steganography, as shown in Fig. 3a. It is inferred that the existing steganography destroys the parity-correlation of the closed-loop and has a negative influence on the stability of the pitch delay distribution. Subsequently, we apply the probability distribution of the parity distribution within a frame to illustrate the effect of the steganography in the pitch delay. It is assumed that an AMR speech sample contains N frames with T sub-frames per frame (T = 4 AMR-NB codec at mode 12.2 kbps).

The parity of each sub-frame has only two states, and each sub-frame is an independent event. Therefore, there are 2^4 states in all four sub-frames. Assuming the probability of four odd sub-frames are P_{T0}, P_{T1}, P_{T2}, and P_{T3} respectively, the distribution probability P_k (k = 1, 2, 3, …, 16) of each of the 16 states is as follows.

$$P_k = P_{T0}^{\alpha_0} \cdot (1 - P_{T0})^{1-\alpha_0} \cdot P_{T1}^{\alpha_1} \cdot (1 - P_{T1})^{1-\alpha_1} \cdot P_{T2}^{\alpha_2} \cdot (1 - P_{T2})^{1-\alpha_2} \cdot P_{T3}^{\alpha_3} \cdot (1 - P_{T3})^{1-\alpha_3},$$
$$k = 1, 2, \ldots, 16. \alpha_0 = 0, 1. \alpha_1 = 0, 1. \alpha_2 = 0, 1. \alpha_3 = 0, 1. \tag{7}$$

If P_{T0}, P_{T1}, P_{T2}, and P_{T3} are unequal, the probability of P_k are various. If P_{T0}, P_{T1}, P_{T2}, and P_{T3} are equal to P_T, some of P_k is equivalent and the formula become a binomial distribution formula described as follow.

$$P_k\{X = \alpha\} = \binom{n}{a} P_T^{\alpha} (1 - P_T)^{n-\alpha}, \alpha = 0, 1, 2, 3, 4, \tag{8}$$

where n is equal to 4. If P_T is equal to 1/2, all P_k are equal to 1/16. Actually, according to ANR-NB coding rules, P_{T0}, P_{T1}, P_{T2}, and P_{T3} are unequal. However, typically the

probability that the secret information appears 0 (or 1) is similar to approximately 1/2. Therefore, P_k will definitely change before and after steganography.

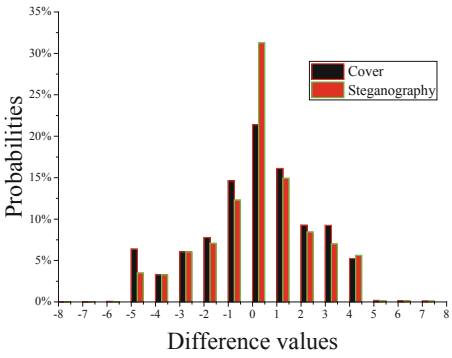


Fig. 3a. The distribution of pitch difference.

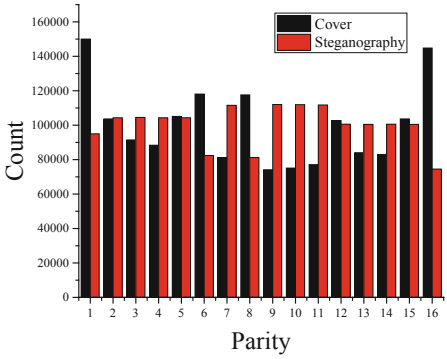


Fig. 3b. The distribution of P_k

In order to validate our conjecture, the following experiment was designed. The differences between T_0 and T_1 and the difference between T_2 and T_3 in 3600 samples were statistically analyzed to obtain Fig. 3a. The statistics of the number of consecutive odd-numbered sub-frames in the sample are shown in Fig. 3b.

Figure 3a is a representation of the relationship between T_0, T_1 and T_2, T_3 within a frame, from which the difference between before and after steganography can be observed clearly. Figure 3b illustrates that the probability distribution of P_k are more average for steganographic samples. From Fig. 3a, it can be learned that the difference in the original sample is more concentrated and the distribution after steganography is evener, which verifies our previous theoretical analysis. This conclusion can be drawn in Fig. 3b that the parity distribution within one frame after steganography is evener. In Fig. 3b, the parity distribution is not uniform within one frame without steganography. Therefore, we choose the parity of the difference to characterize this change. However, the statistical result is not enough to demonstrate the correlation of each sub-frame within a frame. Therefore, we describe their correlation by the Bayes's theorem of the parity of the current sub-frame and the parity of the next sub-frame. Then the conditional probability is regarded as the feature of classification, and finally, SVM is applied to classify to judge whether the sample is steganographic.

3.2 Description of the Features Based on Bayes Theorem

Three features based on Bayes' theorem are depicted in this section. From the above description, it is known that there is an obvious difference between the pitch delay parity Bayes probability in the original sample and the steganographic sample. In order to describe this difference, the Bayesian formula of the pitch delay is considered as the feature. Since the state of the pitch delay is either odd or even and these are two

mutually exclusive events, only one of them needs to be recorded as a feature. For the convenience of describing the following features, only the Bayesian probability of odd pitch delay is calculated. Assume that the four sub-frames are odd-numbered events A_0, A_1, A_2, A_3, and the even-numbered events are $A_0, \bar{A}_1, \bar{A}_2, \bar{A}_3$.

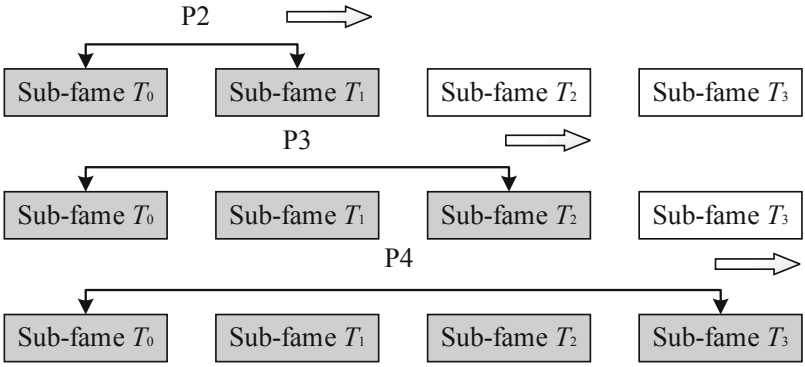


Fig. 4. Features based on Bayes's theorem

The first kind of feature is the relationship between the second sub-frame and the first sub-frame, which include two cases: P_1 and P_2. P_1 ($P_1 = P(A_1 \mid A_0)$) is the conditional probability of the first sub-frame occurred odd under the condition the second sub-frame occurred odd. P_2 ($P_2 = P(A_1 \mid A_0)$) is the conditional probability of the first sub-frame occurred odd under the condition the second sub-frame occurred even.

$$P_1 = P(A_1|A_0) = \frac{P(A_1A_0)}{P(A_0)}. \tag{9}$$

In the same way, the second kind of feature is the relationship between the first sub-frame, the second sub-frame and the third sub-frame. These features include four cases, which are $P_3 = P(A_2 \mid A_0\ A_1)$, $P_4 = P(A_2 \mid \bar{A}_0A_1)$, $P_5 = P(A_2 \mid A_0\bar{A}_1)$, $P_6 = P(A_2 \mid \bar{A}_0\bar{A}_1)$ respectively.

$$P_3 = P(A_2 \mid A_0A_1) = \frac{P(A_0A_1A_2)}{P(A_0A_1)}. \tag{10}$$

The third feature is that the fourth sub-frame is an odd Bayesian probability in the case where the first sub-frame, the second sub-frame, and the third sub-frame are determined. These features include eight cases, which are $P_7 = P(A_3 \mid A_2A_0A_1)$, $P_8 = P(A_3 \mid A_2A_0\bar{A}_1)$, $P_9 = P(A_3 \mid A_2\bar{A}_0A_1)$, $P_{10} = P(A_3 \mid A_2\bar{A}_0\bar{A}_1)$, $P_{11} = P(A_3 \mid \bar{A}_2A_0A_1)$, $P_{12} = P(A_3 \mid \bar{A}_2A_0\bar{A}_1)$, $P_{13} = P(A_3 \mid \bar{A}_2\bar{A}_0A_1)$, $P_{14} = P(A_3 \mid \bar{A}_2\bar{A}_0\bar{A}_1)$ respectively.

$$P_7 = P(A_3 \mid A_0A_1A_2) = \frac{P(A_0A_1A_2A_3)}{P(A_0A_1A_2)}. \tag{11}$$

Figure 4 demonstrates the extraction processing of features based on the Bayes formula for the speech sample encoded with the AMR-NB codec at 12.2 kbps mode. Finally, all of these 14-dimensional are combined to the PBP features.

4 Steganalysis Scheme

In this section, we will introduce steganalysis steps using SVM (support vector machines) as a classifier, which has become an increasingly popular tool for classification. Moreover, its accuracy and detection efficiency have a huge advantage, especially in the small sample set (Figs. 5 and 6).

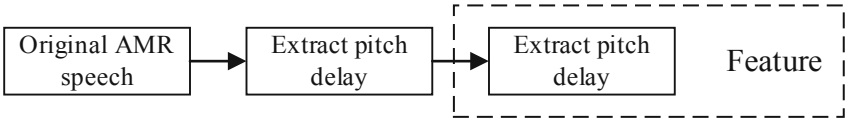


Fig. 5. Extract feature steps

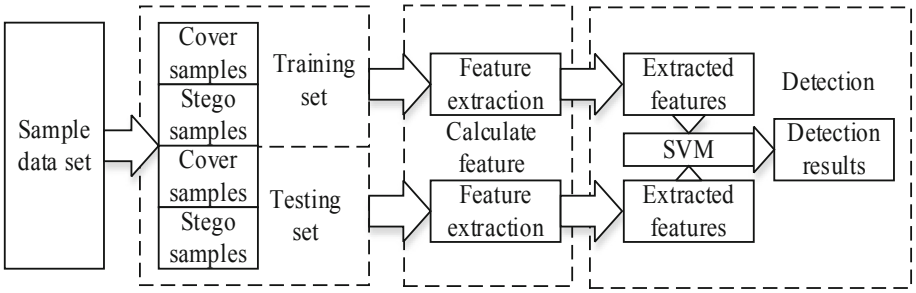


Fig. 6. Classification steps

The SVM training set main steps are shown as follows:

STEP 1. Collect a large number of speech samples randomly and divide it into two parts, half of which are used to steganography and the other half encoded by the original encoder.
STEP 2. Calculate the proposed features through formulas in Sect. 3.2.
STEP 3. Train the steganographic and original speech samples with original and steganographic tags using the above features.

The SVM testing set main steps are shown below:

STEP 1. Recode the voice samples of the test set with the same standard.
STEP 2. Calculate the proposed features in both the original and re-encoded samples.
STEP 3. Enter the feature vectors into the trained classifier to determine whether the sample is steganographic.

5 Performance Evaluation and Analysis

5.1 Experimental Settings

In this paper, SVM open source library is used as a classifier to evaluate experimental results, in which parameters, for example, Gaussian radial basis function kernel for SVM classification, are default setting. Our database consists of 3367 PCM voices, which has been adopted by many papers [21, 34–36]. Each PCM voice is a mono, 8 kHz and 16-bit quantized code, with 10-s dimensions per length. According to different languages, these voices can be divided into four categories: Chinese male voice, Chinese female voice, English male voice, and English female voice. And the proportion of such voices is equivalent. All speeches were encoded at 12.2 kb/s in the AMR using 10% to 100% embedding rate from 1 s to 10 s applying the steganography method of Huang et al. [25] and Yan et al. [26]. Half of the databases are randomly selected as the training set and the other half as the test set, while the embedded information is composed of (0, 1) random number produced by random seed 3367. In the following experiments, we will analyze and evaluate the data from the Accuracy (ACC) False-positive rate (FPR) False-negative rate (FNR) data. ACC has introduced to determine the correct proportion, that is, whether the steganographic sample or the cover sample can be judged correctly. The ACC expression is:

$$ACC = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \tag{12}$$

where N_{TP} is the number of positive instances which are judged to be correct, namely, the steganographic samples are identified. N_{TN} is the number of passive instances which are judged to be negative, namely, the samples that are not steganographic determined to be not steganographic. N_{FP} is the number of negative instances which are judged to be positive, that is, the samples with no steganography are mistakenly considered as a steganographic sample. N_{FN} is the number of positive instances which are judged to be negative, that is, steganographic samples are considered non-steganographic samples. FPR is the proportion of negative instances which are mistaken in all negative instances.

$$FPR = \frac{N_{FP}}{N_{FP} + N_{TN}}, \tag{13}$$

where FNR is the proportion of positive instances which are mistaken in all positive instances.

$$FNR = \frac{N_{FN}}{N_{TP} + N_{FN}}. \tag{14}$$

The result obtained in our experiments is that the method of Ren et al. [19] have an advantage when using Huang's steganography method in a relatively short time compared with the features mentioned in the previous section, but in other cases, the effect is close. It should be noted that the feature dimension mentioned is much lower than the C-MSDPD feature and therefore has better robustness. In the following data, for convenience, only partial results are shown, which are the features of the contrast chart of one-second to ten-second samples.

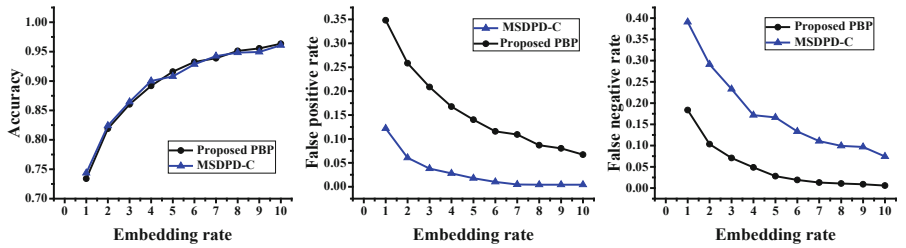
5.2 The Method Proposed Under High Embedding Ratio Is Compared with the Existing Method

Table 1 shows the comparison of the features when the embedding rate is 100% at different times. When Huang's steganography is exploited as the detection object, the classification effects of the PBP features are similar to the C-MSDPD features. However, when Yan's steganography is exploited as the detection object, the classification effects of the PBP features are better than C-MSDPD's. Moreover, when Yan's steganography is exploited as the detection object, the classification effects of the PBP features are better than C-MSDPD's. Yan's steganography method is an improvement of Huang's, and aim to promote the anti-detection ability. As can be seen from the table, compared with Huang's method, the detection accuracy of C-MSDPD in Yan's method is dropped markedly. However, different steganography methods have a minor effect on PBP's performance, which has better adaptability compared with C-MSDPD.

Table 1. The detection accuracies for three features when the embedding rate is 100%

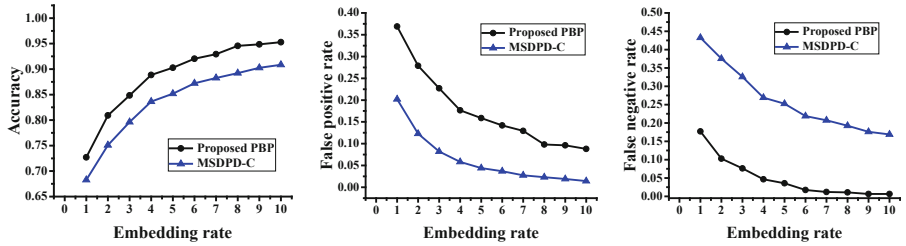
Time (100%)	Huang		Yan	
	C-MSDPD	PBP	C-MSDPD	PBP
1 s	74.36%	73.41%	68.27%	72.70%
2 s	82.41%	81.91%	75.07%	80.93%
3 s	86.45%	86.04%	79.62%	84.85%
4 s	90.02%	89.19%	83.63%	88.86%
5 s	90.79%	91.59%	85.18%	90.29%
6 s	92.84%	93.26%	87.23%	92.04%
7 s	94.24%	93.88%	88.27%	92.93%
8 s	94.83%	95.13%	89.22%	94.56%
9 s	94.95%	95.54%	90.26%	94.86%
10 s	96.08%	96.35%	90.85%	95.28%

As can be seen in Figs. 7 and 8, compared with the MSDPD-C features the ACC, FPR and FNR of the proposed features have obvious advantages. Figure 7 is a comparison of ACC, FPR and FNR for the proposed features and the MSDPD-C feature when the sample length is 1 s to 10 s for an embedding rate of 100% in Huang's steganography while Fig. 8 in Yan's steganography. From Figs. 7 and 8, the accuracy increases with the increase of sample length, while the FPR and FNR decrease. Compared with the MSDPD-C, the FPR of proposed PBP is underperforming. However, it has better performance in Accuracy and the FNR. Figure 9 is a comparison of ROC for the proposed features and the MSDPD-C feature when the sample length is 10 s, and the embedding rate is from 30% to 90% in Huang's method while Fig. 10 in Yan's method.



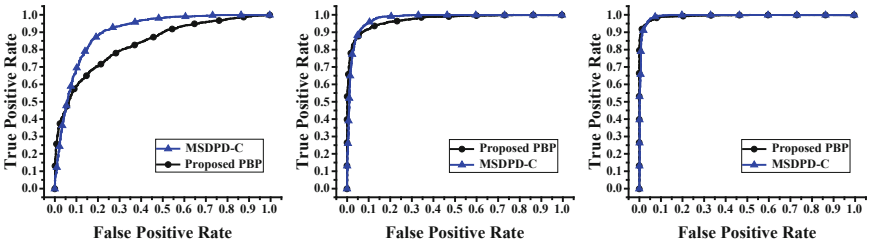
7.1 Statistical results of ACC 7.2 Statistical results of FPR 7.3 Statistical results of FNR

Fig. 7. The detection accuracies for Huang's method in 100% embedding rate



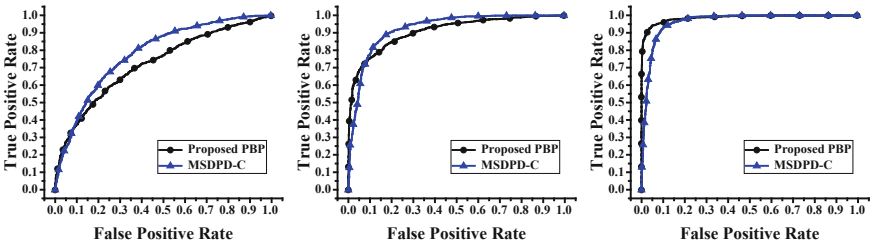
8.1 Statistical results of ACC 8.2 Statistical results of FPR 8.3 Statistical results of FNR

Fig. 8. The detection accuracies for Yan's method in 100% embedding rate



9.1 Embedding rate of 30% 9.2 Embedding rate of 60% 9.3 Embedding rate of 90%

Fig. 9. The ROC curves for detecting Huang's method



10.1 Embedding rate of 30% 10.2 Embedding rate of 60% 10.3 Embedding rate of 90%

Fig. 10. The ROC curves for detecting Yan's method

6 Conclusion

Because of the unpredictability for the pitch of speech parameter encoding, many steganographic methods are presented for secret communication. Motivated by present difficulties, a practical steganalysis scheme is developed in this paper. Distinct from existing works, we treat a frame as a calculation unit and pay more attention to the change in numerical parity rather than just the change in these values. Finally, SVM is employed to classify the PBP features. We evaluate the performance of the proposed method with plenty of speech samples coded by adaptive multi-rate audio coder (AMR), and compare it with the state-of-the-art methods. The experimental results illustrate that our method can effectively detect the pitch delay-based steganography and achieve superior performance than other state-of-the-art methods on ACC, FPR, and FNR. Particularly, the proposed method can provide excellent real-time performance and robustness because of its lower feature dimension and complexity. Therefore, the proposed method can support credible practicability in the steganalysis scenario for real-time speech streams.

Acknowledgements. This research is funded by the National Natural Science Foundation of China under Grant Nos. U1536115 and U1405254, the Natural Science Foundation of Fujian Province of China under Grant No. 2018J01093, the Program for New Century Excellent Talents in Fujian Province University under Grant No. MJK2016-23, the Program for Outstanding Youth Scientific and Technological Talents in Fujian Province University under Grant No. MJK2015-54, the Promotion Program for Young and Middle-aged Teachers in Science and Technology Research of Huaqiao University under Grant No. ZQN-PY115, Program for Science and Technology Innovation Teams and Leading Talents of Huaqiao University under Grant No. 2014KJTD13, the Opening Project of Shanghai Key Laboratory of Integrated Administration Technologies for Information Security under Grant No. AGK201710 and the Subsidized Project for Postgraduates' Innovative Fund in Scientific Research of Huaqiao University No. 17014083010.

References

1. Mazurczyk, W., Lubacz, J.: LACK - a VoIP steganographic method. arXiv:08114138 Cs, November 2008
2. Kar, D.C., Mulkey, C.J.: A multi-threshold based audio steganography scheme. J. Inf. Secur. Appl. **23**(Supplement C), 54–67 (2015)
3. Djebbar, F., Ayad, B., Meraim, K.A., Hamam, H.: Comparative study of digital audio steganography techniques. EURASIP J. Audio Speech Music Process. **2012**(1), 25 (2012)
4. Cheddad, A., Condell, J., Curran, K., Mc Kevitt, P.: Digital image steganography: survey and analysis of current methods. Signal Process. **90**(3), 727–752 (2010)
5. Zhang, Y., Qin, C., Zhang, W., Liu, F., Luo, X.: On the fault-tolerant performance for a class of robust image steganography. Signal Process. **146**, 99–111 (2018)
6. Luo, X., et al.: Steganalysis of HUGO steganography based on parameter recognition of syndrome-Trellis-Codes. Multimed. Tools Appl. **75**(21), 13557–13583 (2016)
7. Mazurczyk, W.: VoIP steganography and its detection—a survey. ACM Comput. Surv. **46** (2), 20:1–20:21 (2013)
8. Tian, H., et al.: Optimal matrix embedding for voice-over-IP steganography. Signal Process. **117**, 33–43 (2015)
9. Sadek, M.M., Khalifa, A.S., Mostafa, M.G.M.: Video steganography: a comprehensive review. Multimed. Tools Appl. **74**(17), 7063–7094 (2015)
10. Neal, H., ElAarag, H.: A reliable covert communication scheme based on VoIP steganography. In: Shi, Y.Q. (ed.) Transactions on Data Hiding and Multimedia Security X. LNCS, vol. 8948, pp. 55–68. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46739-8_4
11. Tian, H., Liu, J., Li, S.: Improving security of quantization-index-modulation steganography in low bit-rate speech streams. Multimed. Syst. **20**(2), 143–154 (2014)
12. Tian, H., et al.: Improved adaptive partial-matching steganography for voice over IP. Comput. Commun. **70**, 95–108 (2015)
13. Geiser, B., Vary, P.: High rate data hiding in ACELP speech codecs, pp. 4005–4008 (2008)
14. Ren, Y., Wu, H., Wang, L.: An AMR adaptive steganography algorithm based on minimizing distortion. Multimed. Tools Appl. **77**(10), 12095–12110 (2018)
15. Miao, H., Huang, L., Chen, Z., Yang, W., Al-Hawbani, A.: A new scheme for covert communication via 3G encoded speech. Comput. Electr. Eng. **38**(6), 1490–1501 (2012)
16. Luo, D., Yang, R., Huang, J.: Identification of AMR decompressed audio. Digit. Signal Process. **37**, 85–91 (2015)

17. Ekudden, E., Hagen, R., Johansson, I., Svedberg, J.: The adaptive multi-rate speech coder. In: IEEE Workshop on Speech Coding Proceedings. Model, Coders, and Error Criteria (Cat. No. 99EX351), pp. 117–119 (1999)
18. Miao, H., Huang, L., Shen, Y., Lu, X., Chen, Z.: Steganalysis of compressed speech based on Markov and entropy. In: Shi, Y.Q., Kim, H.-J., Pérez-González, F. (eds.) IWDW 2013. LNCS, vol. 8389, pp. 63–76. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43886-2_5
19. Ren, Y., Cai, T., Tang, M., Wang, L.: AMR steganalysis based on the probability of same pulse position. IEEE Trans. Inf. Forensics Secur. **10**(9), 1801–1811 (2015)
20. Ren, Y., Yang, J., Wang, J., Wang, L.: AMR steganalysis based on second-order difference of pitch delay. IEEE Trans. Inf. Forensics Secur. **12**(6), 1345–1357 (2017)
21. Tian, H., et al.: Steganalysis of adaptive multi-rate speech using statistical characteristics of pulse pairs. Signal Process. **134**(Supplement C), 9–22 (2017)
22. Liu, P., Li, S., Wang, H.: Steganography in vector quantization process of linear predictive coding for low-bit-rate speech codec. Multimed. Syst. **23**(4), 485–497 (2017)
23. Liu, J., Tian, H., Lu, J., Chen, Y.: Neighbor-index-division steganography based on QIM method for G.723.1 speech streams. J. Ambient Intell. Humaniz. Comput. **7**(1), 139–147 (2016)
24. Yan, S., Tang, G., Chen, Y.: Incorporating data hiding into G.729 speech codec. Multimed. Tools Appl. **75**(18), 11493–11512 (2016)
25. Huang, Y., Liu, C., Tang, S., Bai, S.: Steganography integration into a low-bit rate speech codec. IEEE Trans. Inf. Forensics Secur. **7**(6), 1865–1875 (2012)
26. Yan, S., Tang, G., Sun, Y.: Steganography for low bit-rate speech based on pitch period prediction. Appl. Res. Comput. **32**(6), 1774–1777 (2015)
27. Xiao, B., Huang, Y., Tang, S.: An approach to information hiding in low bit-rate speech stream. In: IEEE Global Telecommunications Conference, IEEE GLOBECOM 2008, pp. 1–5 (2008)
28. Li, S., Jia, Y., Kuo, C.C.J.: Steganalysis of QIM steganography in low-bit-rate speech signals. IEEE/ACM Trans. Audio Speech Lang. Process. **25**(5), 1011–1022 (2017)
29. Ghasemzadeh, H., Tajik Khass, M., Khalil Arjmandi, M.: Audio steganalysis based on reversed psychoacoustic model of human hearing. Digit. Signal Process. **51**, 133–141 (2016)
30. Ma, Y., Luo, X., Li, X., Bao, Z., Zhang, Y.: Selection of rich model steganalysis features based on decision rough set α-positive region reduction. IEEE Trans. Circuits Syst. Video Technol. **29**, 336–350 (2018)
31. Li, S., Tao, H., Huang, Y.: Detection of quantization index modulation steganography in G.723.1 bit stream based on quantization index sequence analysis. J. Zhejiang Univ. Sci. C **13**(8), 624–634 (2012)
32. Manjunath, S., Gardner, W.: Variable rate speech coding, US7496505B2, 24 February 2009
33. Kodovský, J., Fridrich, J.: Calibration revisited. In: Proceedings of the 11th ACM Workshop on Multimedia and Security, New York, NY, USA, pp. 63–74 (2009)
34. Tian, H., et al.: Distributed steganalysis of compressed speech. Soft Comput. **21**(3), 795–804 (2017)
35. Tian, H., et al.: Steganalysis of low bit-rate speech based on statistic characteristics of pulse positions. In: 10th International Conference on Availability, Reliability and Security, pp. 455–460 (2015)
36. Tian, H., et al.: Steganalysis of analysis-by-synthesis speech exploiting pulse-position distribution characteristics. Secur. Commun. Netw. **15**(9), 2934–2944 (2016)