Journal of Universal Computer Science, vol. 25, no. 9 (2019), 1131-1150
submitted: 15/10/18, accepted: 9/7/19, appeared: 28/9/19 © J.UCS

Steganalysis of Adaptive Multi-Rate Speech Using Statistical Characteristics of Pitch Delay

Hui Tian, Meilun Huang
(College of Computer Science & Technology, National Huaqiao University
Xiamen 361021, China
htian@hqu.edu.cn, mlhuang@hqu.edu.cn)

Chin-Chen Chang
(Department of Information and Computer Science, Feng Chia University
Taichung 40724, Taiwan
alan3c@gmail.com)

Yongfeng Huang
(Department of Electrical Engineering, Tsinghua University
Beijing 100084, China
yfhuang@tsinghua.edu.cn)

Jing Lu, Yongqian Du
(Network Technology Centre, National Huaqiao University
Xiamen 361021, China
jlu@hqu.edu.cn, yqdu@hqu.edu.cn)

Abstract: Steganography is a promising technique for covert communications. However, illegal usage of this technique would facilitate cybercrime activities and thereby pose a great threat to information security. Therefore, it is crucial to study its countermeasure, namely, steganalysis. In this paper, we aim to present an efficient steganalysis method for detecting adaptive-codebook based steganography in adaptive multi-rate (AMR) speech streams. To achieve this goal, we first design a new low-dimensional feature set for steganalysis, including an improved calibrated Markov transition probability matrix for the second-order difference of pitch delay values (IC-MSDPD) and the probability distribution of the odevity for pitch delay values (PDOEPD). The dimension of the proposed feature set is 14, far smaller than the feature set in the state-of-the-art steganalysis method. Employing the new feature set, we further present a steganalysis scheme for AMR speech based on support vector machines. The presented scheme is evaluated with a large number of AMR-encoded speech samples, and compared with the state-of-the-art one. The experimental results show that the proposed method is effective, and outperforms the state-of-the-art one in both detection accuracy and computational overhead.

Keywords: Steganalysis, Steganography, Adaptive multi-rate speech, Pitch delay, Support vector machines
Categories: C.2.0, K.6.5, E.4

1 Introduction

Steganography is an efficient technique used for hiding information, which can achieve covert communication by concealing secret messages into seemingly normal carriers [Zielińska et al., 14], such as, images [Zhang et al., 16], audios [Nedeljko and Tapio, 05], videos [Mazurczyk et al., 16], and network protocols [Frączek et al., 12]. In recent years, due to the increasing popularity and broad influence of Voice over IP (VoIP), VoIP-based steganography has emerged as a new branch of steganography [Mazurczyk et al., 13]. In contrast with traditional carriers, VoIP possesses many advantages, such as instantaneity, massive carrier data, high covert bandwidth and flexible carrier length [Tian et al., 14, Tian et al., 15a]. Thus, VoIP steganography is considered as a promising technique for secure communications [Tian et al., 15b, Mazurczyk et al., 10]. However, like some other security techniques, VoIP steganography is a double-edged sword. Illegal usage of this technique would facilitate cybercrime activities and thereby pose a great threat to information security. Thus, it is also very crucial to develop its countermeasure technique, VoIP steganalysis [Tu et al., 15, Tian et al., 16, Tian et al., 17, Lin et al., 18].

Generally speaking, VoIP steganography has two main implementation ways: approaches based on protocol steganography [Wendzel and Palmer, 15, Waldemar et al., 18] and approaches by embedding information into speech streams [Tian et al., 14, Tian et al., 15a, Tian et al., 15b]. By contrast, the latter has attracted the most attention in the research community [Mazurczyk et al., 13]. So far, fruitful studies have been conducted on compressed speech streams based on various codecs, such as, G. 711 [Tian et al., 15a, Tian et al., 15b], G. 729 [Wu et al., 15], G.723.1 [Tian et al., 14, Liu et al., 16a], iLBC [Wu and Sha, 17], Speex [Janicki, 16]. More recently, adaptive multi-rate (AMR) codec, due to being widely employed in not only 3G and 4G speech services [3GPP/ETSI, 16, Varga et al., 06] but also popular mobile instant messaging apps (such as WeChat, WhatsApp, Snapchat and LINE), has attracted increasing interest of the steganography research community [Ren et al., 17a, He et al., 18, Nishimura, 09, Zhang et al., 18].

AMR is a typical speech codec based on Algebraic Code Excited Line Prediction (ACELP). In general, according to the adopted parameters for embedding information, the steganographic techniques based on ACELP codecs can be divided into three categories, i.e., linear prediction coefficient (LPC) steganography [He et al., 18, Peng et al., 16, Liu et al., 15, Liu et al., 16b], adaptive-codebook (ACB) steganography [Nishimura, 09, Yu et al., 12] and fixed-codebook (FCB) steganography [Ren et al., 17a, Miao et al., 12, Geiser and Vary, 08]. Of course, all three types of parameters can be also employed to hide information together. In this work, we focus on the ACB steganography and its steganalysis in AMR encoded speech streams. In ACELP-based codecs, the ACB is often employed to model the pitch excitation to obtain a high quality of synthesized speech [3GPP/ETSI, 16]. Accordingly, the ACB parameters for each subframe involve pitch delay and pitch gain. Because it is difficult to determine the pitch periods of speech signals accurately [Hess and O'Shaughnessy, 84], the pitch delay parameters are widely considered as ideal carriers with sufficient redundancy for hiding information. In other words, secret information can be embedded in speech streams by modifying the pitch delay parameters. There have been many fruitful studies in this field. Nishimura

[Nishimura, 09] first proposed two steganographic methods for AMR codec, i.e., adaptive pitch quantization (APQ) and pitch change point (PCP). The former achieves information hiding by varying the quantization width of the pitch delay value, while the latter replaces pitch delay data in even subframes with secret information when an abrupt change in pitch is observed. Yu et al. [Yu et al., 12] proposed another steganographic method for AMR speech, which determines the pitch delay sequence according to the information to be embedded. Wu et al. [Wu et al., 15] proposed an analysis-by-synthesis based data hiding method for G.729a, whose main idea is to conceal information by modifying the least significant bits of pitch delay parameters. Janicki [Janicki, 16] presented a data hiding method for Speex, where the secret information is hidden into the approximated fine pitch values in the slowly varying regions. Liu et al. [Liu et al., 13] designed a double-layer steganographic method, in which the quantization rules for both integer and fractional pitch delay values are modified to embed secret data. Huang et al. [Huang et al., 12] presented a steganographic method, which achieves information hiding by modifying the search rule of the closed-loop adaptive codebook. The main idea is idea to use suboptimal pitch delay values to replace the optimal ones according to the secret information. Further, to reduce the embedding distortion, Yan et al. [Yan et al., 15] presented a double-layer steganography scheme, in which the first-layer embedding is realized by limiting the search ranges of the pitch periods in the even subframes, and the second one is achieved by selecting proper pitch delay values according to the information to be embedded.

To detect the ACB-based steganography, some researchers have also made useful attempts [Li et al., 14, Ren et al., 17b]. For example, Li et al. [Li et al., 14] presented a steganalysis method based on codeword network with the correlation characteristic of adjective frames, which can detect Huang's method [Huang et al., 12] successfully; Moreover, Ren et al. [Ren et al., 17b] proposed a steganalysis method for AMR-based speech employing the calibrated Markov transition probability matrix for the second-order difference of pitch delay values (C-MSDPD) as the feature, which can detect both Huang's and Yan's methods [Huang et al., 12, Yan et al., 15] effectively and outperforms Li's method [Li et al., 14]. However, the feature based on the original Markov transition probability matrix for the second-order difference of pitch delay values (MSDPD), rather than C-MSDPD feature ultimately adopted for detection, is used to determine the optimal threshold interval, which is apparently inappropriate. Our further experimental analysis show that the threshold interval of C-MSDPD (i.e., [-6, 6]) employed in Ren's method not the optimal one, and the dimensionality of the C-MSDPD feature can be further reduced by determining the optimal threshold. In addition, Ren's method only considers the influence of the steganographic operations on the pitch delay values, while ignoring the steganographic influence on the odevity of pitch delay values. That is, the C-MSDPD feature cannot sufficiently characterize the influence of the steganographic operations.

Thus, in this paper, we present two types of statistical features for steganalysis, namely, the improved C-MSDPD and the probability distribution for the odevity of pitch delay values. The dimension of the adopted feature set is 14, far smaller than 169 dimensions of the previous C-MSDPD. Employing the new feature set, we present a steganalysis of AMR speech based on support vector machines (SVM). The proposed method is evaluated with a large number of AMR-encoded speech samples, and compared with the state-of-the-art one [Ren et al., 17b]. The experimental results show that the proposed method is effective, and outperforms the state-of-the-art one in both computational overhead and detection accuracy.

The remaining of this paper is organized as follows. To make this paper self-contained, Section 2 introduces the principle of AMR encoding and adaptive codebook search, and briefly reviews typical steganographic methods based on pitch delay and the state-of-the-art steganalysis method. Section 3 describes the adopted features, and the SVM-based steganalysis scheme. The performance of the proposed scheme is comprehensively evaluated and compared with the state-of-the-art one in Section 4. Finally, the concluding remakes are given in Section 5.

2 Background and Related Work

In this section, we first introduce the principle of adaptive codebook search in brief, then give an overview of the typical steganographic methods [Huang et al., 12, Yan et al., 15] for AMR speech, finally review the-state-of-the-art steganalysis method [Ren et al., 17b].

2.1 Principle of Adaptive Codebook Search in AMR codec

AMR codec is an encoding algorithm based on ACELP. During the encoding procedure, two appropriate code vectors are respectively chosen from the adaptive codebook and a fixed codebook. Further, the codec can obtain the excitation signal by using the vectors. Moreover, the two selected vectors are sent into a linear prediction (LP) synthesis filter to construct the synthetic speech [3GPP/ETSI, 16]. The best speech quality can be obtained by an analysis-by-synthesis search procedure under the criterion of minimizing the perceptually weighted error between the original speech and synthesized one. The encoding procedure of AMR codec is shown in Figure 1.

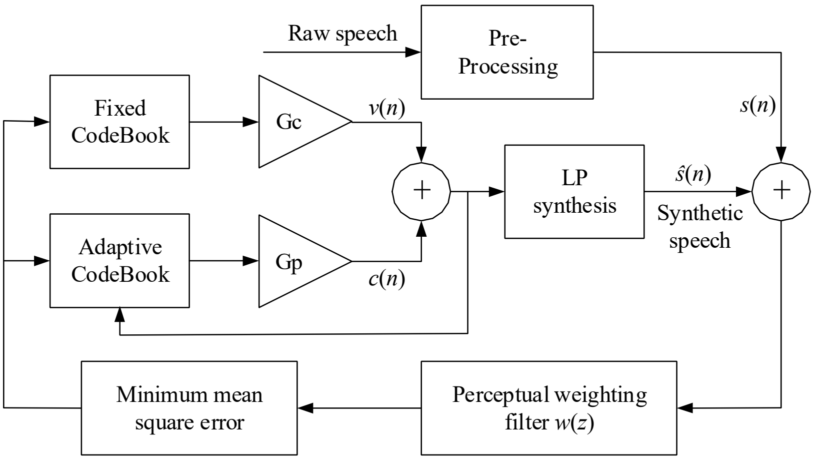


Figure 1: The encoding procedure of AMR codec

Pitch period is an important parameter for speech compression and synthesis, which means the period when the vocal cords vibrate periodicity. Adaptive codebook search, including an open-loop analysis process and a closed-loop analysis process, aims to get optimal pitch delay. The open-loop analysis process is performed on a perceptually weighted signal and aims to offer a search range of pitch delay to the closed-loop analysis process. The closed-loop analysis process is performed on a linear predictive residual signal. The pitch delay and pitch gain can be obtained after the closed-loop analysis process, where the pitch delay is the value of the pitch period. For ease of description, we take AMR at 12.2 kb/s mode as an example to introduce the related principles in the following text.

Parameter	**1st Subframe**	**2nd Subframe**	**3rd Subframe**	**4th Subframe**	**Total Per Frame**
Two LSP Sets					38
Pitch Delay	9	6	9	6	30
Pitch Gain	4	4	4	4	16
Algebraic Codebook Indices	35	35	35	35	140
Codebook Gain	5	5	5	5	20
Total					244

Table 1: Bit allocation of the AMR (12.2 kb/s) coding algorithm for a 20 ms frame [3GPP/ETSI, 16]

Table 1 shows the bit allocation of the AMR at 12.2 kb/s mode [3GPP/ETSI, 16]. In the AMR codec, there are 4 subframes in each frame of 20ms and totally 244 bits produced. According to the search criteria of minimizing the mean square error, the optimum integer pitch delay of each subframe is selected from different search ranges. During the encoding process of AMR at 12.2 kb/s mode, the open-loop pitch analysis process is performed twice in a frame, and two estimated integer lags for the first and third subframe are obtained at first. Next, the closed-loop pitch analysis process is carried out to get an optimum pitch delay around the previous two estimated lags. The search ranges of the first and third subframes can be formally described as

$$p_{0,i} = \begin{cases} [18, 24], & T_{op} < 21 \\ [T_{op} - 3, T_{op} + 3], & 21 \le T_{op} \le 140, \\ [137, 143], & T_{op} > 140 \end{cases} \tag{1}$$

where $p_{0,\,i}$ is the integer pitch delay of the first (or third) subframe in the i-th frame, and T_{op} is the corresponding open-loop estimated lag.

Further, the pitch delays of the second and fourth subframes are determined by performing the closed-loop pitch analysis process again around the integer pitch delays of previous subframes. Their search ranges can be stated as

$$p_{1,i} = \begin{cases} [18, 27], & p_{0,i} < 23 \\ [p_{0,i} - 5, p_{0,i} + 4], & 23 \le p_{0,i} \le 139, \\ [134, 143], & p_{0,i} > 139 \end{cases} \quad (2)$$

where $p_{1,i}$ is the integer pitch delay of the second (or fourth) subframe, and $p_{0,i}$ is the integer pitch delay of the first (or third) subframe.

2.2 Principle of Adaptive Codebook-based Steganography

From the principle of the adaptive codebook search, we can learn that the pitch delay values obtained during AMR encoding are suboptimal. That is to say, there are some other candidate values for constructing a high quality of synthesized speech. Benefiting from this property, the existing AMR-based steganography methods [Huang et al., 12, Yan et al., 15] incorporate the steganographic operation into the adaptive codebook search process, and implement information hiding with no perceptible distortion by replacing the original pitch delay values with "appropriate" suboptimal ones based on the content of secret messages. In other words, they embed secret data into speech streams by altering the quantization rules of pitch delay during the adaptive codebook search.

Huang et al. [Huang et al., 12] first presented a steganographic adaptive codebook search strategy, which can embed four bits of secret messages into each subframe for AMR codec at 12.2 kb/s mode. The main idea is to modify the search range of pitch delay according to the odd or even value of each secret bit. To be specific, for each subframe, if the secret bit is "1", only the odd integer pitch delay values are searched and modified for information hiding; otherwise, only the even integer pitch delay values are selected. Accordingly, the secret bit sequence can be extracted from the received pitch delay values as

$$b_{i\times 4+j} = \mathrm{mod}(p_{i,j}, 2), \quad (3)$$

where $p_{i,j}$ is the j-th subframe integer pitch delay value of the i-th frame and $b_{i\times 4+j}$ is the secret bit embedded into $p_{i,j}$.

Yan et al. [Yan et al., 15] further suggested a double-layer steganographic algorithm to reduce embedding distortion, which, however, shares the similar idea with Huang's method, i.e., modulating the search rule of pitch delay in each frame to conceal secret messages in the adaptive codebook search process. However, differing from Huang's steganography, to reduce the distortion caused by the change of pitch delay values, only the pitch delays in the second and fourth subframes in each frame are employed to hide messages. Moreover, the exclusive OR relationship between the pitch delay values is used to embed more secret messages. The adopted exclusive OR operation can be described as

$$m_3 = \lfloor (p_{1,i} \bmod 4)/2 \rfloor \oplus \lfloor (p_{3,i} \bmod 4)/2 \rfloor, \quad (4)$$

where m_3 is the third secret bit, while $p_{1,i}$ and $p_{3,i}$ are respectively the modified pitch delay values of the second and fourth subframes in the i-th frame. Accordingly, the secret message can be easily extracted from each frame at the receiver side according to Eqs. (3) and (4).

2.3 Review of the-State-of-the-Art Detection Method for Adaptive Codebook-based Steganography

To detect the existing AMR-based steganography methods described above [Huang et al., 12, Yan et al., 15], Ren et al. [Ren et al., 17b] proposed a steganalysis method based on the fact that the pitch delay values in cover speech samples are more stable than those in steganographic ones. Accordingly, the Markov transition probability matrix of the second-order difference of pitch delay (MSDPD) is employed to describe the differences between the cover and the steganographic samples. The MSDPD feature (denoted as *M*) can be calculated as follows:

$$M(x,y)=\frac{\sum_{i=0}^{N-4} P(\Delta(i+1)=y, \Delta(i)=x)}{\sum_{i=0}^{N-4} P(\Delta(i)=x)}, \tag{5}$$

where *M*(*x*, *y*) is the transition probability that Δ(*i*+1) = *y* while Δ(*i*) = *x*, Δ(*i*) is the second-order difference of pitch delay in the *i*-th subframe, *P*(Δ(*i*) = *x*) is the probability that Δ(*i*) = *x*, and *P*(Δ(*i*+1) = *y*, Δ(*i*) = *x*) is the probability thatΔ(*i*+1) = *y* and Δ(*i*) = *x* in the meantime.

The calibration technique is a common means to strengthen the effect of steganalysis features [Kodovsky, 09], which is also used to improve the MSDPD feature. In the work, the ranges from [-1, 1] to [-10, 10], totaling ten intervals, were tested to get the optimal threshold of MSDPD feature. Experimental results showed that the detection accuracy was the best when the threshold was [-6, 6]. At this moment, the dimension of the feature set is 13 × 13 = 169. Further, the calibrated feature, defined as C-MSDPD, can be obtained as

$$C-MSDPD = MSDPD_C - MSDPD_O, \tag{6}$$

where the *MSDPD_O* and *MSDPD_C* are respectively the features before and after calibration. Figure 2 illustrates the extraction procedure of the C-MSDPD feature.

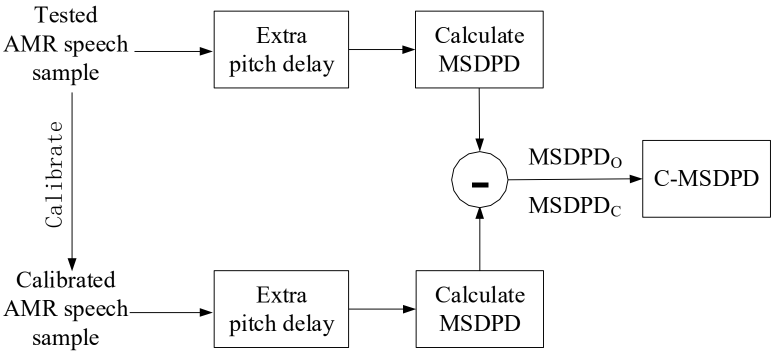


Figure 2: The extraction procedure of the C-MSDPD feature

It has been proved that C-MSDPD is feasible for detecting the existing two steganographic algorithms [Huang et al., 12, Yan et al., 15]. However, there still are two obvious defects in this work:

- Ren et al. adopted an inappropriate way to determine the optimal threshold interval. Specifically, they used the MSDPD feature, rather than C-MSDPD feature ultimately adopted for detection, to determine the optimal threshold interval. Moreover, the dimension of the C-MSDPD feature is 169, which inevitably induces considerable computational costs.
- The feature set is not complete enough to characterize the change rules of pitch delay values in the existing steganographic methods. As mentioned above, both the two steganographic algorithms exploit the odevity of pitch delay values to hide messages. That is to say, the steganographic operations would affect not only the distribution of pitch delay values but also their parity distribution. The former was taken into account, while the latter was ignored in Ren's work.

Given the shortcomings, we are motivated to find out more accurate and more complete features, and accordingly, present a more effective steganalysis of AMR speech based on pitch delay in this paper.

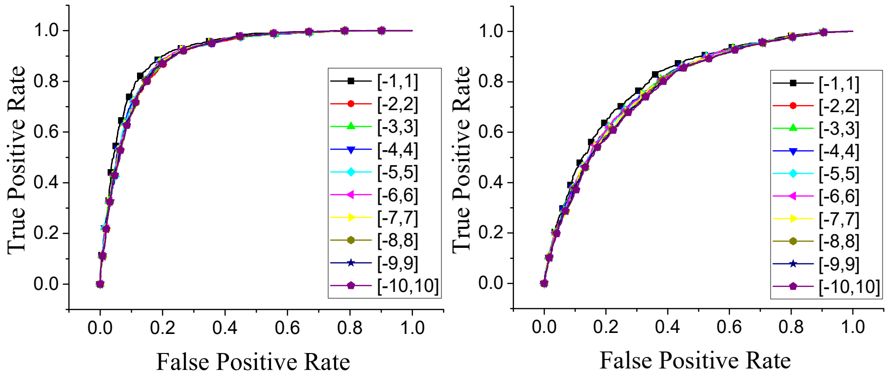
3 Proposed Steganalysis Scheme

In this section, we first give an improved set of the C-MSDPD feature by re-examining the threshold for the second-order difference of pitch delay (denoted as IC-MSDPD), and present a new feature based on the probability distributions of the odevity for pitch delay (PDOEPD). Finally, incorporating the presented feature set, we propose an SVM-based steganalysis scheme.

3.1 The optimal feature set based on calibrated Markov transition probability matrix for the Second-order Difference of Pitch Delay (IC-MSDPD)

In Ren's method [Ren et al., 17b], for the MSDPD feature, the totally ten candidate thresholds from [-1, 1] to [-10, 10] were tested, and the results demonstrated that the optimal one was [-6, 6]. This threshold was also directly considered as the optimal one for the C-MSDPD feature. However, our further observation and analysis show that the threshold of [-6, 6] is the best choice for MSDPD, but not for C-MSDPD. Thus, we seek to find out the optimal threshold for the C-MSDPD feature by reexamining the thresholds from [-1, 1] to [-10, 10].

In the experiment, all the speech samples are encoded with the AMR codec at 12.2 kb/s mode, and the steganographic samples are produced by performing the existing steganographic methods at the embedding rate of 30%. Figure 3 shows the ROC (Receiver Operating Characteristic) curves for different thresholds, from which we can learn that the differences among the ten cases are slight. To further compare them, we calculate the area under the curve (AUC) of each case, as shown in Table 2. The larger the AUC is, the better performance the classifier has. Apparently, the AUC for the threshold of [-1, 1] is the largest, indicating that the threshold of [-1, 1] is the best choice for the C-MSDPD feature. In this case, the dimension of the feature set is $3 \times 3 = 9$, far smaller than the 169 dimensions at the case of [-6, 6]. That is to say, the C-MSDPD feature with the threshold of [-1, 1] (denoted as IC-MSDPD) can not only achieve better detection performance but also reduce the computational costs.



(a) The ROC curve for detecting Huang's method

(b) The ROC curve for detecting Yan's method

Figure 3: The ROC curves for detecting the existing steganographic methods at the embedding rate of 30% with the C-MSDPD feature under different thresholds. (a) The ROC curve for detecting Huang's method. (b) The ROC curve for detecting Yan's method.

	Area Under Curve (AUC)	
Thresholds	**For Huang's method**	**For Yan's method**
[-1, 1]	0.9174	0.8012
[-2, 2]	0.9010	0.7838
[-3, 3]	0.9048	0.7853
[-4, 4]	0.9056	0.7857
[-5, 5]	0.9051	0.7842
[-6, 6]	0.9047	0.7838
[-7, 7]	0.9038	0.7802
[-8, 8]	0.9026	0.7773
[-9, 9]	0.9026	0.7753
[-10, 10]	0.9013	0.7736

Table 2: The area under the curve of each case for detecting two steganographic methods

3.2 Probability Distributions of the Odevity for Pitch Delay (PDOEPD)

As mentioned above, the existing steganographic methods share the same idea, namely, exploiting the odevity of pitch delay values to hide secret messages. As is well known, the encrypted message can be regarded as a random binary sequence, where the bits of "0" and "1" are evenly distributed. Since the pitch delay values are determined according to the parity of each secret bit, the probability distribution of odevity for pitch delay conforms to the distribution of embedded secret bits. To verify this inference, we perform a statistical experiment to compare the probability distribution of odevity for pitch delay in cover speech samples and steganographic ones.

In the experiment, we construct three sample sets, namely, the cover set including 1680 speech samples encoded with the AMR codec at 12.2 kb/s mode, the steganographic set produced by Huang's method at the embedding rate of 100%, and the steganographic set produced by Yan's method at the embedding rate of 100%. Moreover, we treat four pitch delay values in a frame f_i as a statistical unit, and record the number of odd pitch delay values (called parity factor) as v_i. Apparently, there are 5 possible values for v_i, i.e., $v_i \in \{0, 1, 2, 3, 4\}$. For each speech sample, we extract the parity factor sequence, $V = \{v_1, v_2, ..., v_N\}$, where N is the number of the frames. Theoretically, the steganographic parity factors are uniformly distributed between 0 and 4. Namely, the appearance probability for v_i is

$$p(v_i) = \frac{C_4^i}{2^4}. \tag{7}$$

Figure 4 shows the probability distributions of the parity factors in the three speech sample sets, from which we can learn that the probability distributions of the parity factors in the two steganographic sample sets meet the theoretical probability distribution of steganographic parity factors on the whole, particularly at the cases of $v_i = 0$, 1 or 4. That is to say, just as we inferred above, the steganographic operations would inevitably cause apparent changes in the probability distribution of odevity for pitch delay. Therefore, we employ the probability distribution of the odevity for pitch delay (PDOEPD) as another steganalysis feature, whose dimension is 5.

3.3 SVM-based Steganalysis Scheme

In this section, employing the popular machine learning tool, SVM[Gu et al., 15, Gu and Sheng, 16, Yuan et al., 16], we present a steganalysis scheme for detecting adaptive codebook-based steganography in AMR speech streams, as illustrated in Figure 5. In this scheme, we employ a 14-dimensional feature set, which consists of the 9-dimensional IC-MSDPD feature and 5-dimensional PDOEPD feature. Moreover, our scheme includes two phases, i.e., the training phase and the detection phase.

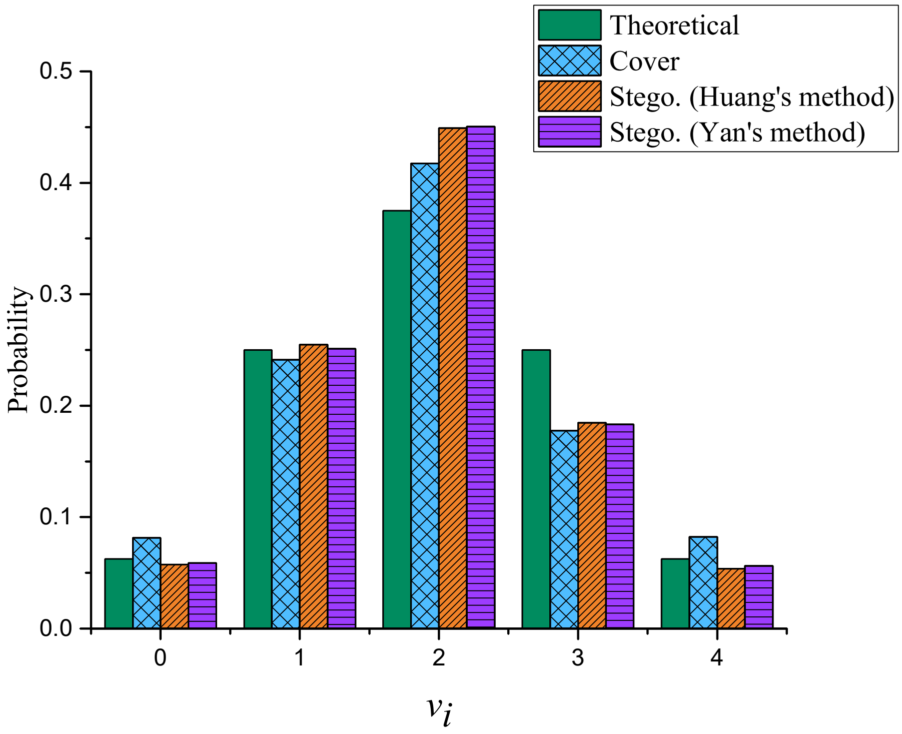


Figure 4: Probability distributions of the parity factors

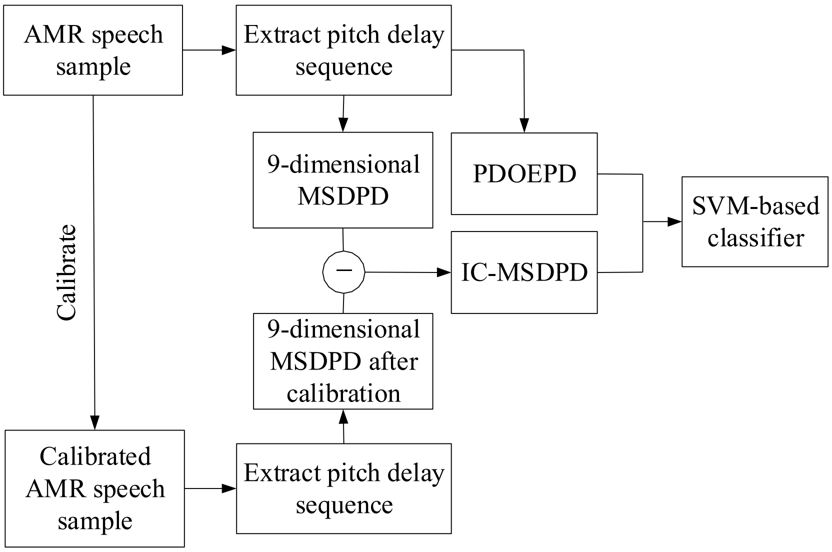

AMR speech sample
Extract pitch delay sequence
9-dimensional MSDPD
PDOEPD
SVM-based classifier
−
IC-MSDPD
Calibrate
9-dimensional MSDPD after calibration
Calibrated AMR speech sample
Extract pitch delay sequence

Figure 5: The process of the presented steganalysis scheme

Specifically, the training phase involves the following steps.

Step 1 Preparation of Sample Set. Collect a large number of AMR speech samples, and perform Huang's method and Yan's method on each cover sample at the established embedding rates to produce two groups of steganographic sample sets.

Step 2 Feature Extraction. For each speech sample, first extract the 5-dimensional PDOEPD feature and 9-dimensional MSDPD feature; further, extract 9-dimensional MSDPD feature from the corresponding calibrated sample, and get the IC-MSDPD feature according to Eq. (6); finally, obtain the 14-dimensional feature set by combining the PDOEPD feature and IC-MSDPD feature.

Step 3 Classifier Training. Train the SVM-based classifier with the selected 14-dimensional feature set.

Accordingly, the detection phrase contains two steps listed as below:

Step 1 Feature Extraction. For each speech sample to be detected, first extract the PDOEPD feature and original 9-dimensional MSDPD feature; further, extract 9-dimensional MSDPD feature from the corresponding calibrated sample, and get the IC-MSDPD feature according to Eq. (6); finally, obtain the 14-dimensional feature set by combining the PDOEPD feature and IC-MSDPD feature.

Step 2 Detection (Classification). Feed the feature set into the trained SVM-based classifier to detect whether the given speech sample is an original object or a covert one.

4 Performance Evaluation

4.1 Experimental setup and evaluation criteria

In the experiment, the well-known open-source tool LibSVM [Chang, 11] is used as the classifier. The default parameters, namely c = 1 and g = 1/1064, and the linear SVM (C-style) with RBF kernel are adopted. To evaluate the performance of the proposed scheme, 3367 ten-second speech samples are collected from language-learning materials, which consist of two categories: Chinese speech and English speech, each includes male speech and female speech. Each sample is mono, sampled at 8 KHz and quantized at 16 bits/s. In the experiments, all the samples are further encoded with 3GPP public floating-point AMR codec at 12.2 kb/s mode to obtain the cover samples. In addition, two steganographic methods, i.e., Huang's method and Yan's method, are involved. Accordingly, there are two groups of steganographic samples in each steganographic experiment. To evaluate the detection performance in various scenarios, for each speech sample, we generate its steganographic versions at various embedding rates (from 10% to 100%) and the versions with different lengths (from 1s to 10s) at the embedding rate of 100%, respectively using Huang's method and Yan's method. Table 3 shows all the speech sample sets used in the evaluation experiments.

Number of Cover samples	Steganographic samples for each steganographic method			
	Embedding rate	Number	Length	Number
3367	10%	3367	1s	3367
	20%	3367	2s	3367
	30%	3367	3s	3367
	40%	3367	4s	3367
	50%	3367	5s	3367
	60%	3367	6s	3367
	70%	3367	7s	3367
	80%	3367	8s	3367
	90%	3367	9s	3367
	100%	3367	10s	3367

Table 3: The speech sample sets used in the evaluation experiments.

In all steganalysis experiments, three statistical results on the accuracy (ACC) rate, false-positive rate (FPR) and false-negative rate (FNR) are employed for measuring the performance of the proposed scheme. ACC is the proportion of true detention results among the total number of test samples, namely,

$$ACC = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \tag{8}$$

where N_{TP} is the number of steganographic samples correctly identified, N_{TN} is the number of cover samples correctly identified, N_{FP} is the number of cover samples inaccurately identified, and N_{FN} is the number of steganographic cover samples inaccurately identified.

FPR, usually known as false alarm rate, is the proportion of cover samples inaccurately identified among the total number of all cover samples, namely,

$$FPR = \frac{N_{FP}}{N_{FP} + N_{TN}}. \tag{9}$$

FNR, usually known as missed-event rate, is the proportion of steganographic samples inaccurately identified among the total number of all steganographic samples, namely,

$$FNR = \frac{N_{FN}}{N_{FN} + N_{TP}}. \tag{10}$$

4.2 Performance comparison of the proposed method and existing one

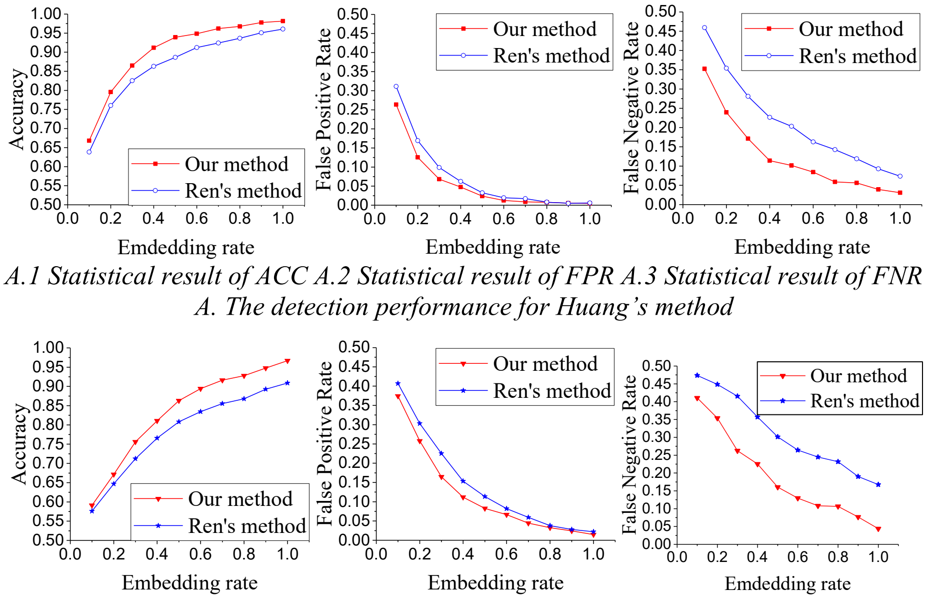
To compare the performance of the proposed method with the state-of-the-art one (i.e., Ren's method [Ren et al., 17b]), we conduct steganalysis experiments respectively for ten-second samples at various embedding rates (from 10% to 100%) and the samples with different lengths (from 1s to 10s) at the embedding rate of

100%. In each steganalysis experiment, a total of 3368 samples, including 1684 samples (with odd indices) selected from the cover samples and the corresponding samples from each group of steganographic samples, are used to train the SVM-based classifier; and for a given steganographic method, the remainder 1683 samples (with even indices) from cover samples and their corresponding steganographic samples are employed to evaluate the detection performance.

Figures 6 and 7 respectively show the experimental results for the ten-second samples at various embedding rates (from 10% to 100%) and the samples with different lengths (from 1s to 10s) at the embedding rate of 100%. From these results, we can learn the following two facts:

First, for the two steganalysis methods, the detection performance has positive correlations with the embedding rates of a given steganographic method. That is, the higher the embedding rate, the better the detection performance. However, for all the steganographic methods, the proposed steganalysis method can achieve significantly higher detection accuracies as well as markedly lower FPRs and FNRs at any given embedding rate than the state-of-the-art one. Particularly, for detecting Huang's method [Huang et al., 12], the accuracy of the proposed method is higher than 90% when embedding rate is only 40%, while the state-of-the-art one achieves the same accuracy rate when embedding rate is 60% or above. For detecting Yan's method [Yan et al., 15], the advantage of our method is more obvious. The accuracy of the proposed method is about 6% higher than that of the state-of-the-art one for any samples with the embedding rates larger than 50%.

Second, for the two steganalysis methods, the detection performance has positive correlations with the lengths of the speech samples. That is, as the length of the speech sample increases, the detection efficiencies of both steganalysis methods are improved, but our method can still achieve better performance than the state-of-the-art one. Particularly, in terms of detecting Yan's method, the accuracy of the proposed method is about 6% higher than that of the state-of-the-art one for any samples with the length between 2s and 10s at the embedding rate of 100%.



A.1 Statistical result of ACC A.2 Statistical result of FPR A.3 Statistical result of FNR

A. The detection performance for Huang's method

B.1 Statistical result of ACC B.2 Statistical result of FPR B.3 Statistical result of FNR

B. The detection performance for Yan's method

Figure 6: Experimental results for ten-second speech samples at various embedding rates. A. The detection performance for Huang's method. B. The detection performance for Yan's method.

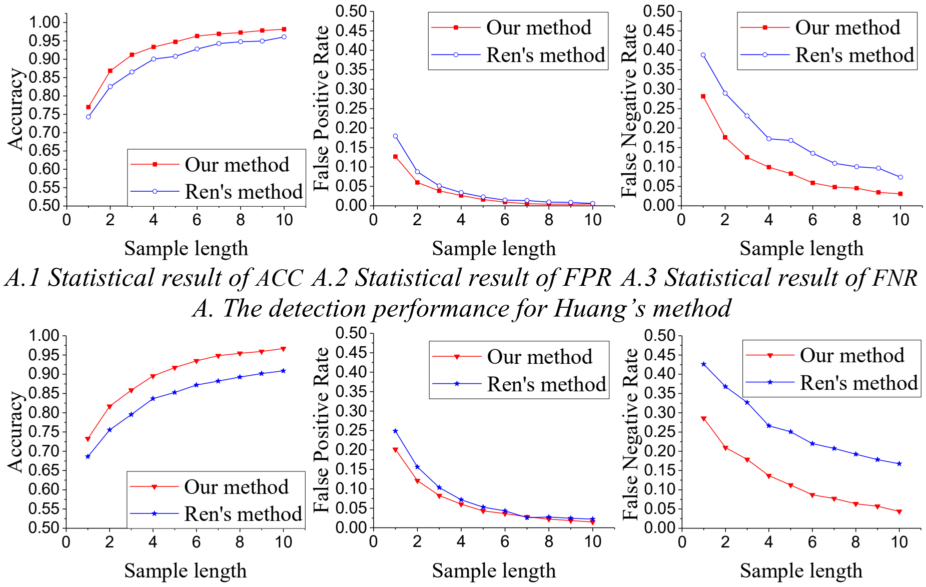
To show the performance of the two steganalysis methods more clearly, we give receiver-operating-characteristic (ROC) curves for detecting the existing steganographic methods at typical embedding rates of 30%, 60% and 100% with the sample lengths of 1s, 5s, and 10s, respectively, as shown in Figure 8. The results demonstrate once again that the two steganalysis methods are really feasible and effective for detecting steganographic methods, while the proposed method can offer better detection performance than the state-of-the-art one.

In addition, to show the advantages of our low-dimensional feature set, we test the SVM-classifier average training and detecting time with the sample length of 10s, as shown in Table 4. This experiment is conducted on an HP workstation with an Intel Core i3-3220 Duo CPU at 3.30 GHz, 8 GB RAM and 7200 RPM 500 GB Serial ATA drive with a 16 MB buffer.

As Table 4 shows, the training and detecting time in our work is far less than that in Ren's method, meaning that our proposed method is much more efficient in terms of computational overhead. In other words, the proposed method would reduce much less delay and achieve much better real-time capability for detection.

Average time (ms)	**For Huang's method**		**For Yan's method**	
	Ren's method	Proposed method	Ren's method	Proposed method
Training	76.22	6.60	98.06	10.15
Detecting	16.84	1.44	18.74	1.72

Table 4: The average training and detecting time for the two steganalysis methods



A.1 Statistical result of ACC A.2 Statistical result of FPR A.3 Statistical result of FNR

A. The detection performance for Huang's method

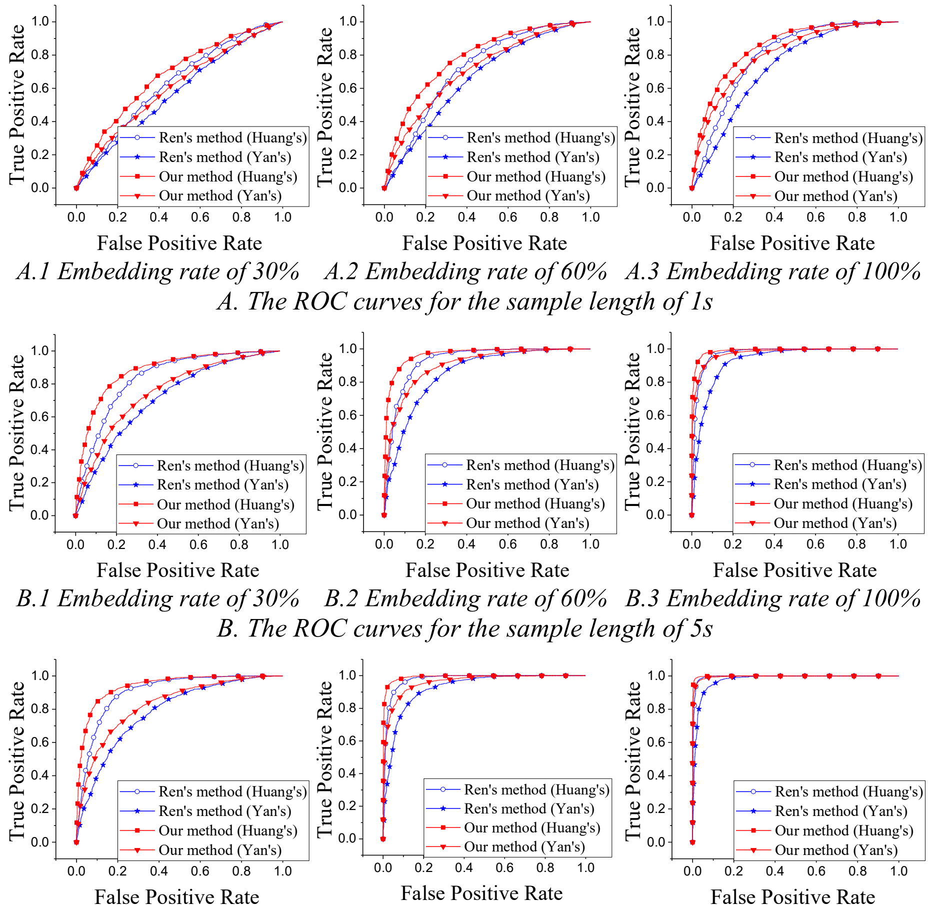
B.1 Statistical result of ACC B.2 Statistical result of FPR B.3 Statistical result of FNR

B. The detection performance for Yan's method

Figure 7: Experimental results for speech samples with different lengths at the embedding rate of 100%. A. The detection performance for Huang's method. B. The detection performance for Yan's method.

6 Conclusion

In this paper, we proposed an effective steganalysis method for detecting steganography of AMR speech based on the modification of pitch delay. We design two types of steganalysis features, namely, IC-MSDPD and PDOEPD. The first one is the improved version of C-MSDPD employed in the state-of-the-art method, while the latter one describes the odevity of pitch delay values. The total number of dimensions of the adopted feature set is 14, far smaller than that of previous C-MSDPD. Further, we design an SVM-based steganalysis scheme. With a large number of AMR-encoded speech samples, the proposed method is comprehensively evaluated and compared with the previous one. The experimental results show that the proposed method is feasible and effective, and achieves better detection accuracy than the state-of-the-art one while requiring much less computational overhead. It is worth noting that, although the method is described and evaluated with the AMR codec, it can be extended to other ACELP-based codecs, such as G.723.1, G.729, and QECLP.



A.1 Embedding rate of 30% *A.2 Embedding rate of 60%* *A.3 Embedding rate of 100%*

A. The ROC curves for the sample length of 1s

B.1 Embedding rate of 30% *B.2 Embedding rate of 60%* *B.3 Embedding rate of 100%*

B. The ROC curves for the sample length of 5s

C.1 Embedding rate of 30% *C.2 Embedding rate of 60%* *C.3 Embedding rate of 100%*

C. The ROC curves for the sample length of 10s

Figure 8: The ROC curves for the two steganographic methods with the sample lengths of 1s, 5s, and 10s at embedding rates of 30%, 60% and 100%. A. The ROC curves for the sample length of 1s. B. The ROC curves for the sample length of 5s. C. The ROC curves for the sample length of 10s.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China under Grant Nos. U1536115 and U1405254, Natural Science Foundation of Fujian Province of China under Grant No. 2018J01093, Program for New Century Excellent Talents in Fujian Province University under Grant No. MJK2016-23, Program for Out-standing Youth Scientific and Technological Talents in Fujian Province University under Grant No. MJK2015-54 and Opening Project of Shanghai Key Laboratory of Integrated Administration Technologies for Information Security under Grant No. AGK201710.

References

[3GPP/ETSI, 16] 3GPP/ETSI. Digital Cellular Telecommunications System (Phase 2+); Universal Mobile Tele-communications System (UMTS); LTE; Mandatory Speech Codec Speech Processing Functions; Adaptive Multi-Rate (AMR) Speech Codec; Trans-coding Functions (3GPP TS 26.090 version 13.0.0 Release 13), Technical Report TR 126 090, Sophia Antipolis Cedex, France, January 2016.

[Chang, 11] Chang, CC. LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol., 2011, 27: 1–27: 27.

[Frączek et al., 12] Frączek W., Mazurczyk W., Szczypiorski K.: Multilevel steganography: improving hidden communication in networks, 2012, 1967-1986.

[Geiser and Vary, 08] Geiser B., Vary P.: High rate data hiding in ACELP speech codecs. International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV: IEEE, 2008, 4005-4008.

[Gu et al., 15] Gu B., Sheng V.S, Tay K., et al.: Incremental support vector learning for ordinal regression, IEEE Trans. Neural Netw. Learn. Syst., 2015, 1403–1416.

[Gu and Sheng, 16] Gu B., Sheng V.S.: A robust regularization path algorithm for ν-support vector classification, IEEE Trans. Neural Netw. Learn. Syst. 2016.

[Hess and O'Shaughnessy, 84] Hess W., O'Shaughnessy D.: Pitch determination of speech signals: algorithms and devices. Journal of the Acoustical Society of America, 1984, 1277-1278.

[He et al., 18] He J., Chen J., Xiao S., et al.: A novel AMR-WB speech steganography based on diameter-neighbor codebook partition, Security and Communication Networks, 2018, 1-11.

[Huang et al., 12] Huang Y., Liu C., Tang S.: Steganography integration into a low bit rate speech codec, IEEE Transactions on Information Forensics and Security, 2012, 1865-1875.

[Janicki, 16] Janicki A.: Pitch-based steganography for Speex voice codec, Security and Communication Networks, 2016, 2923-2933.

[Kodovsky, 09] Kodovsky JFJ.: Calibration revisited. Proceedings of ACM Multimedia and Security Workshop, 2009, 63-74.

[Li et al., 14] Li S., Jia Y., Fu J., et al.: Detection of pitch modulation information hiding based on codebook correlation network, Chinese Journal of Computers, 2014, 2107–2117.

[Lin et al., 18] Lin Z., Huang Y., Wang J.: RNN-SM: fast steganalysis of VoIP streams using recurrent neural network, IEEE Transactions on Information Forensics and Security, 2018, 1-15.

[Liu et al., 13] Liu C., Bai S., Huang Y., et al.: An information hiding algorithm based on pitch prediction, Computer Engineering, 2013, 137-140.

[Liu et al., 15] Liu P., Li S., Wang H.: Steganography in vector quantization process of linear predictive coding for low bit rate speech codec, Multimedia Systems, 2015, 1-13.

[Liu et al., 16a] Liu J., Tian H., Lu J., et al.: Neighbor-index-division steganography based on QIM method for G.723.1 speech streams, Journal of Ambient Intelligence and Humanized Computing, 2016, 139-147.

[Liu et al., 16b] Liu P., Li S., Wang H.: Steganography integrated into linear predictive coding for low bit-rate speech codec, Multimedia Tools and Applications, 2016, 1-23.

[Mazurczyk et al., 10] Mazurczyk W., Lubacz J.: LACK-A VoIP steganographic method, Telecommunication Systems, 2010, 153-163.

[Mazurczyk et al., 13] Mazurczyk W.: VoIP steganography and its detection—A survey, ACM Computing Surveys, 2013, 1-21.

[Mazurczyk et al., 16] Mazurczyk W, Szczypiorski K, Janicki A.: YouSkyde: information hiding for Skype video traffic, Multimedia Tools and Applications, 2016, 13521-13540

[Miao et al., 12] Miao H., Huang L., Chen Z., et al.: A new scheme for covert communication via 3G encoded speech, Computers and Electrical Engineering, 2012, 1490-1501

[Nedeljko and Tapio, 05] Nedeljko C., Tapio S.: Increasing Robustness of LSB Audio Steganography by Reduced Distortion LSB Coding, Journal of Universal Computer Science, 2005, 56-65.

[Nishimura, 09] Nishimura A.: Data hiding in pitch delay data of the adaptive multi-rate narrow-band speech codec, Proceedings of International Conference on Intelligent Information Hiding and Multi-media Signal Processing, 2009, 483-486.

[Peng et al., 16] Peng S., Huang F., Li F.: A steganography scheme in a low bit rate speech codec based on 3D-Sudoku matrix, IEEE International Conference on Communication Software and Networks, 2016, 13-18.

[Ren et al., 17a] Ren Y, Wu H., Wang L.: An AMR adaptive steganography algorithm based on minimizing distortion, Multimedia Tools and Applications, 2017, 1-16.

[Ren et al., 17b] Ren Y., Yang J., Wang J., et al.: AMR steganalysis based on second-order difference of pitch delay, IEEE Transactions on Information Forensics and Security, 2017, 345-1357.

[Tian et al., 14] Tian H., Liu J., Li S.: Improving security of quantization-index-modulation steganography in low bit-rate speech streams. Multimedia Systems, 2014, 143-154.

[Tian et al., 15a] Tian H., Qin J., Guo S., et al.: Improved adaptive partial-matching steganography for voice over IP, Computer Communications, 2015, 95-108.

[Tian et al., 15b] Tian H., Qin J., Huang Y., et al.: Optimal matrix embedding for voice-over-IP steganography, Signal Process, 2015, 33-43.

[Tian et al., 16] Tian H., Wu Y., Chang C., et al.: Steganalysis of analysis-by-synthesis speech exploiting pulse-position distribution characteristics, Security and Communication Networks, 2016, 2934-2944.

[Tian et al., 17] Tian H., Wu Y., Chang C., et al.: Steganalysis of adaptive multi-rate speech using statistical characteristics of pulse pairs, Signal Processing, 2017, 9-22.

[Tu et al., 15] Tu S., Tao H., Huang Y.: Detection of instant voice communication steganography using semi-supervised learning, Tsinghua Science and Technology, 2015, 1246-1252.

[Waldemar et al., 18] Waldemar G., Jacek K., Krzysztof S.: SOMSteg - Framework for covert channel and its detection within HTTP, Journal of Universal Computer Science, 2018, 864-891.

[Varga et al., 06] Varga I., Lacovo R., Usai P.: Standardization of the AMR wide-band speech codec in 3GPP and ITU-T, IEEE Communications Magazine, 2006, 66-73.

[Wendzel and Palmer, 15] Wendzel S., Palmer C.: Creativity in Mind: evaluating and maintaining advances in network steganographic research, Journal of Universal Computer Science, 2015, 1684-1705.

[Wu et al., 15] Wu J, Cao H., Li D.: An approach of steganography in G.729 bitstream based on matrix coding and interleaving, Chinese Journal of Electronics, 2015, 157-165.

[Wu and Sha, 17] Wu Z., Sha Y.: An implementation of speech steganography for iLBC by using fixed codebook, IEEE International Conference on Computer and Communications, 2017: 1970-1974.

[Yan et al., 15] Yan S., Tang G., Sun Y.: Steganography for low bit rate speech based on pitch period prediction, Application Research of Computers, 2015, 1774-1777.

[Yu et al., 12] Yu C., Huang L., Yang W., et al.: A 3G speech data hiding method based on pitch period, Journal of Chinese Computer Systems, 2012, 1445-1449.

[Yuan et al., 16] Yuan C., Sun X., Lv R.: Fingerprint liveness detection based on multi-scale LPQ and PCA, China Communications, 2016, 60–65.

[Zhang et al., 16] Zhang Y., Luo X., Yang C., et al.: A framework of adaptive steganography resisting JPEG compression and detection, Security and Communication Networks, 2016, 2957-2971.

[Zhang et al., 18] Zhang Y., Qin C., Zhang W., et al.: On the Fault-tolerant Performance for a Class of Robust Image Steganography, Signal Processing, 2018, 99-111.

[Zielińska et al., 14] Zielińska E., Mazurczyk W., Szczypiorski K.: Trends in steganography, Communications of the ACM, 2014, 86-95.