Detecting Steganography in AMR Speech Based on Pulse Correlation

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

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

Abstract. This paper presents a novel methodology to detect the steganography on the fixed codebook (FCB) of adaptive multi-rate (AMR) speech stream. We have found that correlations of pulses are influenced by the steganographic operation. Based on this, two categories of features are proposed to characterize the pulse correlations, namely subframe-level pulse correlation based on self-information and track-level pulse correlation based on mutual-information, whose feature dimension is only 1/5 of the state of the art. The proposed method employs the support vector machine as the classifier and is evaluated with a large quantity of AMR speech samples. The experimental results demonstrate that the propose method is effective and has a better detection performance than the state of the arts.

Keywords: Steganography · Steganalysis · Adaptive multi-rate speech · Pulse correlation · Self-information · Mutual-information

1 Introduction

Steganography is the art of covert communication by hiding secret information in digital media, such as image [1, 2], text [3, 4], audio [5, 6] and video [7, 8]. To prevent threats and damages caused by illegal uses of steganography, steganalysis, the countermeasure of steganography, has attracted increasing attention from researchers, which aims to detect the existence of secret information in digital media [9]. In recent years, as the fast development of the Internet, Voice over Internet Protocol (VoIP) has become a popular communication service over the Internet and mobile software, which drives researches on the VoIP-based steganography [10]. Compared with traditional carriers, there are many advantages of VoIP-based carriers, such as high steganographic bandwidth, instantaneity and flexible steganographic length [11].

In VoIP service, speech signals are encoded by VoIP codecs into digital information, which are then packetized, and transmission occurs as IP packets over the Internet [12]. In general, there are three types of VoIP codecs: waveform codec, such as ITU G.711; parametric codec, such as LPC-10; hybrid codec, such as adaptive multi-rate (AMR) codec. Because the hybrid codec has a high compression ratio while

J. Li et al. (Eds.): SPNCE 2019, LNICST 284, pp. 485–497, 2019.
https://doi.org/10.1007/978-3-030-21373-2_39

keeping an acceptable speech quality, it is widely applied in VoIP scenarios. Moreover, AMR is a popular format of spoken audio, which has a widespread application in mobile devices. Therefore, AMR speech has become a hot spot of the VoIP-based steganography. To confront the challenges of the AMR-based steganography, the steganalysis of AMR speech streams is conducted in this paper.

For AMR codec, there are three suitable embedding domains including linear predictive coefficient (LPC) [13, 14], adaptive codebook (ACB) [15, 16], fixed codebook (FCB) [17, 18]. Compared with LPC and ACB, FCB accounts for a larger proportion in each frame, for example, the total bits of each frame for AMR narrow bandwidth (AMR-NB) with 12.2 kbit/s mode are 244 bits, while the total bits of FCB are 140 bits [19], which are more than a half. Furthermore, in AMR encoding process, the structure of FCB is based on the interleaved single-pulse permutation (ISPP) design and the search procedure of FCB is the depth-first tree. All the characteristics indicate that the steganography on FCB can be high-capacity and slightly perceptible. Based on this, there have been some steganographic methods on FCB, for example, Geiser and Vary [17] proposed a steganography by confining the second pulse position in the same track during the FCB search to create a covert channel, which can reach up to 2 kbit/s bandwidth with AMR-NB at 12.2 kbit/s mode. Miao et al. [18] presented an adaptive suboptimal pulse combination constrained method to embed secret information during FCB search, which further introduced a steganographic factor to control the embedding capacity.

In recent years, there have been some effective steganalysis methods to detect the above steganographic algorithms, for example, Miao et al. [20] proposed two methods to detect FCB-based steganography. One calculated the Markov transition probabilities of pulse positions in the same track as steganalysis features. The other utilized the joint entropy and conditional entropy to detect secret information. The experimental results show that both methods can detect the Geiser's [17] and Miao's [18] steganography methods, but the detection accuracy are not satisfied. Later, Ren et al. [21] studied the FCB search strategy of steganography and found that there exists a difference of the probability of the same pulse position between cover and steganographic samples. The probabilities of the same pulse position in each track are applied as steganalysis features whose dimension is only 7-dimension. Experiments show that Ren's [21] method outperforms Miao's [20] method. Recently, Tian et al. [22] proposed another effective detection method which has a better detection performance than Ren's [21] while the feature dimension is 498-dimension. In Tian's method [22], the Markov transition matrix of the pulse pairs in the same subframe and the joint probabilities of the pulse pairs in different subframe are calculated as steganalysis features. Thought Ren's method [21] has a low feature dimension, the detection accuracy is not satisfied; while Tian's method [22] is better, the feature is relatively high. To cover the shortages existing in the state of the arts [21, 22], we present a more accurate detection method based on the support vector machine by employing the correlation of pulses to fully capture the influence caused by steganography on FCB as the steganalysis features, namely subframe-level pulse correlation based on self-information and track-level pulse correlation based on mutual-information. The experimental results show that the proposed method is able to detect the steganography methods on the FCB and

outperforms the state of the arts. The main contributions of this paper can be concluded as follow:

(1) This paper proposed two effective categories of steganalysis features. The proposed features are only 100 dimensional, which is 1/5 of that of the state of the art;
(2) A support vector machine based steganalysis scheme is presented;
(3) The proposed steganalysis scheme outperforms the state of the arts.

The rest of this paper is organized as follows. Section 2 introduces the standard encoding principle of the AMR codec, the steganography conducted AMR speech streams and the state of arts for steganalysis. The proposed features are described in Sect. 3. Section 4 presents the support vector machine based steganalysis method. Experiments and analysis are shown in Sect. 5. Finally, concluding remarks are given in Sect. 6.

2 Background and Related Work

This section first introduces the standard encoding principle of AMR codec, then reviews the steganography methods [17, 18] conducted on the FCB, finally, presents the state of the arts for steganalysis [21, 22].

2.1 Standard Encoding Principle of AMR Codec

The AMR codec [19] is a multi-mode codec which supports 8 narrow band encoding modes with bit rates ranging from 4.75 kbit/s to 12.2 kbit/s and 9 wide band encoding modes with bit rates between 6.6 kbit/s and 23.85 kbit/s. The encoding algorithm of the AMR codec is algebraic code-excited linear prediction whose main functions can be divided into three parts: LPC analysis, pitch delay search and FCB search. Because this paper concentrates on the steganography for FCB, only the search strategy of FCB is illuminated, which takes AMR-NB with 12.2 kbit/s mode as the example.

There are 10 non-zeros pulses in the innovation vector, where all the pulses have two amplitudes +1 or −1. In each subframe, there are 40 positions, which are divided into 5 tracks and each track locates two pulses as shown in Table 1. The FCB search aims at minimizing the mean square error between the weighted input speech and the weighted synthesized speech. Let $b(n)$ be the presetting amplitudes which is the sum of the normalized $d(n)$ vector and normalized long-term prediction residual $res_{LTP}(n)$:

Table 1. The FCB structure of AMR-NB with 12.2 kbit/s.

Track	Pulse	Position
0	i_0, i_5	0, 5, 10, 15, 20, 25, 30, 35
1	i_1, i_6	1, 6, 11, 16, 21, 26, 31, 36
2	i_2, i_7	2, 7, 12, 17, 22, 27, 32, 37
3	i_3, i_8	3, 8, 13, 18, 23, 28, 33, 38
4	i_4, i_9	4, 9, 14, 19, 24, 29, 34, 39

$$b(n) = \frac{res_{LTP}(n)}{\sqrt{\sum_{i=0}^{39} res_{LTP}(i) res_{LTP}(i)}} + \frac{d(n)}{\sqrt{\sum_{i=0}^{39} d(i)d(i)}}. \quad (1)$$

The pulse positions with the maximum absolute values of $b(n)$ are searched firstly for five tracks, then the global maximum value for all pulse positions is selected as the position of the first pulse i_0. Next, four iterations are conducted. The position of pulse i_1 is set to the local maximum of each track in each iteration. The pulse pairs $\{i_2, i_3\}$, $\{i_4, i_5\}$, $\{i_6, i_7\}$ and $\{i_8, i_9\}$ are searched by sequentially searching in the nested loops. All the pulse starting positions except i_0 are cyclically shifted in each iteration. Therefore, the pulse pairs are altered, and the pulse i_1 is located a local maximum of a different track. The remain pulses are searched for the other positions in the tracks. There exists at least one pulse in the position corresponding to the global maximum and one pulse in position corresponding to one of the 4 local maxima.

2.2 AMR-Based Steganography

Due to the FCB search is based on the depth-first tree, only a small subset of suitable positions is searched, which leads to a suboptimal codebook. Therefore, it is possible to modify the pulse position with an imperceptible degradation on speech quality to hiding information. The AMR-based steganographic algorithms [17, 18] embed secret information by modifying the FCB search strategy based on the above characteristics.

Geiser and Vary [17] proposed a steganography which restricted the admissible position of the second pulse in each track of AMR-NB with 12.2 kbit/s mode. In Geiser's method, 2 bits secret information is embedding in each track and the pules positions of i_5, …, i_9 are selected two out of 8 possible values. Denote i_t and i_{t+5} as the first and the second pulse position in the same track respectively and $(m)_{i,j}$ as the bits at position i and j of the secret information m in binary representation. The restricted rule of can be expressed as

$$i_{t+5} = \begin{cases} g^{-1}\left(g\left(\left\lfloor \frac{i_t}{5} \right\rfloor\right) \oplus (m)_{2t,2t+1}\right) \cdot 5 + t \\ g^{-1}\left(g\left(\left\lfloor \frac{i_t}{5} \right\rfloor\right) \oplus (m)_{2t,2t+1} + 4\right) \cdot 5 + t \end{cases}, \quad (2)$$

where g and g^{-1} are Gray encoding and decoding by table lookup; "$\oplus$" is the bitwise exclusive disjunction (XOR); $\lfloor x \rfloor$ is the round down function. At the receiver end, the secret information is extracted by

$$(m)_{2t,2t+1} = \left(g\left(\left\lfloor \frac{i_t}{5} \right\rfloor\right) \oplus g\left(\left\lfloor \frac{i_{t+5}}{5} \right\rfloor\right)\right) \bmod 4. \quad (3)$$

Miao et al. [18] present another steganography to embed secret information in AMR speech by searching a suboptimal codevector to replace the cover one. Assume m_t as the secret information to be embedded and p_{ti} is the i-th pulse in the t-th track. The restricted rule of Miao's method is defined as follow

$$m_t = \left(\sum_{i=0}^{L_t} g\left(\left\lfloor \frac{p_{t_i}}{N} \right\rfloor \right) \right) \oplus \eta, \tag{4}$$

where g is Gray encoding by table lookup; N is the number of tracks; L_t is the number of non-zero pulses in track t; η is the embedding factor to control the embedding bits. Particularly, for AMR-NB with 12.2 kbit/s mode, η is usually set as 1, 2, 4. At the receiver end, the secret information is extracted by calculating Eq. (4) again.

2.3 The State of the Arts for AMR-Based Steganalysis

There have been some effective steganalysis methods to detect steganography on FCB of AMR speech streams. In this section, two of the best detection methods are introduced.

Ren et al. [21] presented a low dimension steganalysis based on the probabilities of same pulse positions. In the work, Ren et al. analyzed the impact of steganography on the pulse conditional probability in the t-th track (PCP_t), which is calculated by

$$PCP_t(i,j) = \frac{1}{N_S} \sum_{f=1}^{N_s} \delta(i_a(f,t) = i, i_b(f,t) = j), \tag{5}$$

where i, j are the possible positions and follow $0 \leq i, j \leq N_p - 1$; N_s is the number of subframes; f is the label of subframe and follows $0 \leq f \leq N_s$; i_a and i_b are the first and the second non-zero pulse position respectively. $\delta(R)$ is defined as

$$\delta(R) = \begin{cases} 1, & R \, is \, true \\ 0, & else \end{cases}. \tag{6}$$

It is observed that PCP_t (i, i), the probabilities of two pulses in the same position, are distinguishing between steganographic samples and cover samples. Therefore, in Ren's method, the average of PCP_t (i, i) in all tracks are employed as steganalysis features.

However, if the steganography avoids modifying the pulse in the same positions, Ren's method [21] will be invalid. Motivated by the shortcoming of Ren's method, Tian et al. [22] proposed more complete features including long-term features of pulse pairs (LTFS), short-term features of pulse pairs (STFS) and track to track correlation features (TTFS). The LTFS is the probability distributions of the pulse pairs, which is given by

$$P(x,y) = \frac{\sum_{i=0}^{N_s-1} \left(\delta\left(p_{i,j} = x, p_{i,j+T} = y\right) || \, \delta\left(p_{i,j} = y, p_{i,j+T} = x\right) \right)}{N_s}, \tag{7}$$

where N_s is the same as Eq. (5); ($p_{i,j}$, $p_{i,j+T}$) is the pulse pair for the j-th track in the i-th subframe. The STFS is the Markov transition matrix (MTM) of the pulse pairs of a track in the same subframe. Let p $((a, b) \mid (c, d))$ be the probability that the pulse pair (c, d) is followed by (a, b), so the MTM in the i-th track is described below

$$M_i = \begin{bmatrix} P(u_{i,0}|u_{i,0}) & P(u_{i,0}|u_{i,1}) & \cdots & P(u_{i,R-1}|u_{i,R-1}) \\ P(u_{i,1}|u_{i,0}) & \ddots & & \\ \vdots & & \ddots & \\ P(u_{i,R-1}|u_{i,0}) & \cdots & \cdots & P(u_{i,R-1}|u_{i,R-1}) \end{bmatrix}, \tag{8}$$

where R is the number of all potential pulse-position pairs for i-th track.

The TTFS is the joint probability matrix of pulse pairs, which can be written as

$$J_{i,j} = \begin{bmatrix} P(u_{i,0}, u_{i,0}) & P(u_{i,0}, u_{i,1}) & \cdots & P(u_{i,R-1}, u_{i,R-1}) \\ P(u_{i,1}, u_{i,0}) & \ddots & & \\ \vdots & & \ddots & \\ P(u_{i,R-1}, u_{i,0}) & \cdots & \cdots & P(u_{i,R-1}, u_{i,R-1}) \end{bmatrix}. \tag{9}$$

Although Tian's method [22] has a better detection performance, the feature dimension is quite high (498-dimensional). Given the shortcomings of the-state-of-the-art steganalysis methods, we present a high accurate detection method with a reasonable feature dimension.

3 The Proposed Steganalysis Features

In this section, we will introduce the proposed steganalysis features which include two categories: the subframe-level pulse correlation (SPC) based on self-information and the track-level pulse correlation (TPC) based on mutual information.

3.1 The Subframe-Pulses Features Based on Self-information

For AMR-NB with 12.2 kbit/s mode, each frame is divided into four subframes and in each subframe, there are 10 pulses. For the standard principle, the pulse pairs search order in a same subframe is: $\{i_0, i_1\}$, $\{i_2, i_3\}$, $\{i_4, i_5\}$, $\{i_6, i_7\}$ and $\{i_8, i_9\}$ as the blue line in Fig. 1 while for steganography, the search order is: $\{i_0, i_5\}$, $\{i_1, i_6\}$, $\{i_2, i_7\}$, $\{i_3, i_8\}$ and $\{i_4, i_9\}$ as the red line in Fig. 1.

The steganography modifies the search strategy of the pulse pairs in the subframe, and the encoding process is successive which means the latter pulse is related to the former pulse. Thus, the short-term stability of speech is destroyed by the steganographic operation. Therefore, the correlation among the pulses in the same subframe is affected by steganography. To describe the correlation of pulses in the same subframe, the SPC is calculated as steganalysis feature by the following equation:

$$H(x,y) = -\log_2\left(\frac{\sum_{i=0}^{L}\sum_{j=i}^{L}\left(\left(\delta(l_i = x)\,\&\,\delta(l_j = y)\right)\,||\,\left(\delta(l_i = y)\,\&\,\delta(l_j = x)\right)\right)}{C_L^2}\right), \tag{10}$$

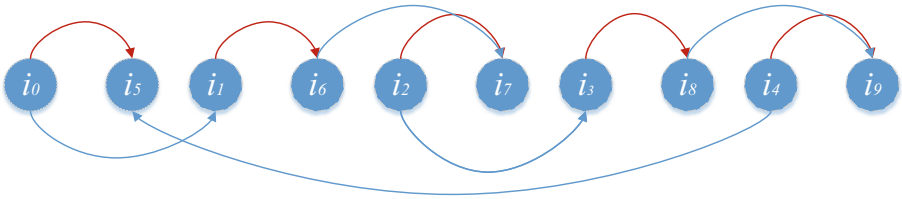


Fig. 1. The search strategy of pulse pairs between standard and steganographic principle. (Color figure online)

where L is the total of the pulse in a subframe; l_i is the i-th pulse in the subframe; x, y are the pulse positions. For AMR-NB 12.2 kbit/s mode, $L = 10$ and $x, y \in [0, 7]$. In steganalysis, the average SPC of all the subframe are applied as the steganalysis feature whose dimension is 36-D.

The comparison of SPC value between cover and steganographic sample is shown in Fig. 2. From Fig. 2 we can learn that (1) for both cover and steganographic sample, the SPC values are larger in different pulse positions than those in the same pulse positions; (2) the SPC values are influenced by steganography and the changes of SPC values in different pulse positions are more significant than those in the same pulse position.

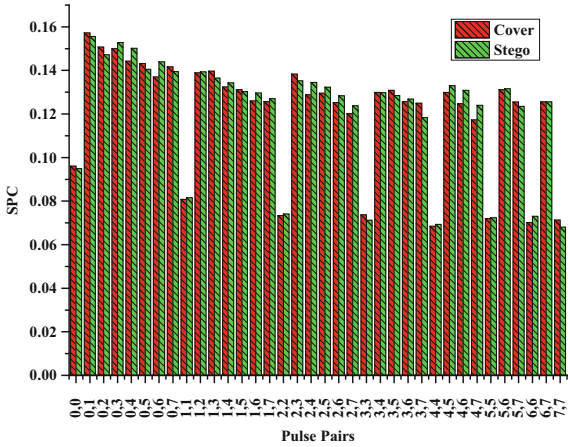


Fig. 2. The comparison of SPC between cover sample and steganographic sample at embedding rate of 100%

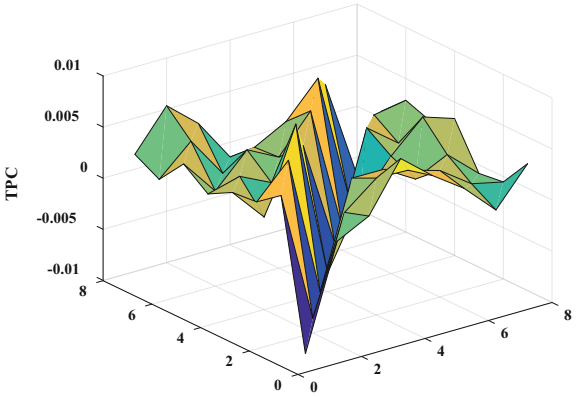
3.2 The Track-Pulses Features Based on Mutual Information

The SPC indicates the correlation of all the pulses in the subframe, while the steganography algorithms are implemented by modifying the second pulse position in the same track. To get the more precise correlation of the pulse pair in the same track, the TPC are proposed to capture the influence on the pulse pair in the same track.

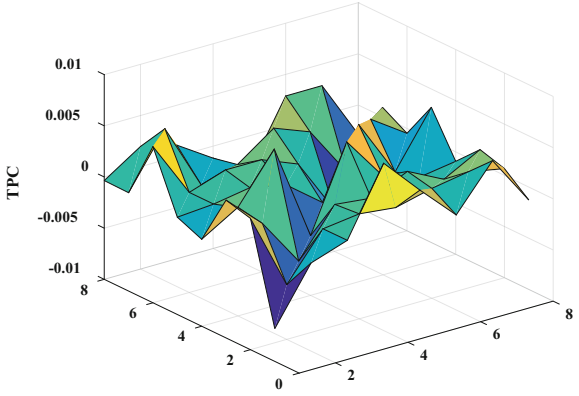
Mutual information [23] or transformation can measure the amount of information from one variable by observing another variable. Because the second pulse in the same track is modified to embed the secret information, the mutual information between the pulse pair in the same track will be influenced. The TPC is given by the following equation:

$$I(x,y) = p(x,y)\log_2\left(\frac{p(x,y)}{p(x)p(y)}\right), \tag{11}$$

where x, y are the first pulse position and the second pulse position in the same track; $p(x)$, $p(y)$ are the marginal probability distribution of x, y respectively; $p(x, y)$ is the



(a) TPC of cover sample



(b) TPC of steganographic sample

Fig. 3. The comparison of TPC between cover samples and steganographic sample at embedding rate of 100%.

joint probability distribution of x and y. Similarly, the average of TPC of all track is employed in steganalysis whose dimension is 64.

The comparison of TPC between cover sample and steganographic sample is shown in Fig. 3, from which it can be concluded that the distribution of TPC of steganographic sample is more even than that of cover sample.

Thus, the proposed features include SPC with 36-dimensional and TPC with 64-dimensional, whose feature dimension is 100-dimensional. The proposed feature dimension is only 1/5 of Tian's feature dimension.

4 Steganalysis Scheme Based on Support Vector Machine

In this section, we present a steganalysis model based on support vector machine [24] (SVM), which contains a training process and detection process. Specifically, the training process consists of the following three steps:

Step 1: Samples preparation. Collect a quantity of speech samples and encoded them with the AMR codec and conduct steganography on them at different embedding rates.
Step 2: Features extraction. Extract the proposed features in Sect. 3 from both cover samples and steganographic samples.
Step 3: Model training. Train the SVM model with the features extracted in Step 2.

Similarly, the detection process includes two steps:

Step 1: Feature extraction. Extract the proposed features from the samples to be detected.
Step 2: Detection. Input the extracted features into the trained SVM and make detection decision according to the output of the model.

5 Experiments and Analysis

5.1 Experimental Setup

To evaluate the experimental without loss generality, we collect 4000 speech samples with length of ten second as the dataset, and all the samples are encoded with AMR-NB 12.2 kbit/s mode. There are four steganography methods are detected in the steganalysis experiments, including Geiser's method [17] and Miao's methods with $\eta = 1$, 2, 4 [18]. All the detection methods are implemented on Python 3.6 with sklearn.svm [25]. Moreover, accuracy (ACC), false positive rate (FPR) and false negative rate (FNR) are employed as the metrics in the steganalysis experiment.

5.2 Detection Performance and Analysis

The steganalysis experiments for ten second samples at different rates (from 0.1 to 1.0) are conducted in this section, and we compare the proposed method with the state of

the arts [21, 22]. In each steganalysis experiment, half samples are used to trained the model and the remained samples are employed as the test set to evaluate the detection performance for each steganography. The experimental results are listed from Table 2, 3, 4 and 5.

Table 2. Detection performance for Geiser's method [17].

Embedding rate	Ren's method [21]			Tian's method [22]			The proposed method		
	ACC	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR
10%	0.5690	0.4460	0.4160	0.5703	0.4585	0.4010	**0.6212**	0.3775	0.3800
20%	0.6623	0.3805	0.2950	0.7120	0.3055	0.2705	**0.7302**	0.2650	0.2745
30%	0.7378	0.3115	0.2130	0.8260	0.1820	0.1660	**0.8455**	0.1420	0.1670
40%	0.8103	0.2260	0.1535	0.9020	0.1050	0.0910	**0.9137**	0.0750	0.0975
50%	0.8692	0.1480	0.1135	0.9523	0.0425	0.0530	**0.9550**	0.0415	0.0485
60%	0.9035	0.1105	0.0825	0.9762	0.0245	0.0230	**0.9722**	0.0250	0.0305
70%	0.9340	0.0820	0.0500	0.9850	0.0145	0.0155	**0.9858**	0.0140	0.0145
80%	0.9453	0.0600	0.0495	0.9942	0.0065	0.0050	**0.9935**	0.0045	0.0085
90%	0.9553	0.0500	0.0395	0.9958	0.0040	0.0045	**0.9968**	0.0030	0.0035
100%	0.9655	0.0405	0.0285	0.9972	0.0025	0.0030	**0.9988**	0.0015	0.0010

Table 3. Detection performance for Miao's method with $\eta = 1$ [18].

Embedding rate	Ren's method [21]			Tian's method [22]			The proposed method		
	ACC	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR
10%	0.5288	0.5070	0.4355	0.5350	0.4645	0.4655	**0.5537**	0.4590	0.4335
20%	0.5715	0.4700	0.3870	0.5998	0.4010	0.3995	**0.6415**	0.3580	0.3590
30%	0.5887	0.4130	0.4095	0.6937	0.3125	0.3000	**0.7208**	0.2860	0.2725
40%	0.6445	0.3955	0.3155	0.7520	0.2440	0.2520	**0.7947**	0.2045	0.2060
50%	0.6835	0.3515	0.2815	0.8193	0.1820	0.1795	**0.8472**	0.1515	0.1540
60%	0.7228	0.3220	0.2325	0.8748	0.1295	0.1210	**0.8965**	0.0975	0.1095
70%	0.7515	0.2730	0.2240	0.9115	0.0890	0.0880	**0.9325**	0.0640	0.0710
80%	0.7967	0.2335	0.1730	0.9440	0.0515	0.0605	**0.9540**	0.0415	0.0505
90%	0.8263	0.2000	0.1475	0.9647	0.0350	0.0355	**0.9720**	0.0215	0.0345
100%	0.8538	0.1740	0.1185	0.9792	0.0225	0.0190	**0.9782**	0.0175	0.0260

From the above tables, we can reach the following conclusions. Firstly, for each detection method, the detect accuracy is proportionate to the embedding rate while the FPR and FNR decrease according to the rise of embedding rate. Secondly, the detection performances for Miao's method $\eta = 4$ are better than the others, while detection performances for Miao's method $\eta = 1$ are the worst, which may be that the Miao's method $\eta = 4$ modifies the most pulse position and Miao's method $\eta = 1$ modifies the least pulse position. Thirdly, the proposed method outperforms the state of the arts especially at the low embedding rates. For Geiser's method, the proposed method

Table 4. Detection performance for Miao's method with $\eta = 2$ [18].

Embedding rate	Ren's method [21]			Tian's method [22]			The proposed method		
	ACC	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR
10%	0.5735	0.4545	0.3985	0.5805	0.4360	0.4030	**0.6155**	0.3895	0.3795
20%	0.6583	0.3865	0.2970	0.7175	0.2975	0.2675	**0.7458**	0.2410	0.2675
30%	0.7418	0.3060	0.2105	0.8220	0.1785	0.1775	**0.8470**	0.1445	0.1615
40%	0.8103	0.2230	0.1565	0.9055	0.0920	0.0970	**0.9150**	0.0830	0.0870
50%	0.8705	0.1460	0.1130	0.9525	0.0540	0.0410	**0.9510**	0.0420	0.0560
60%	0.9077	0.0995	0.0850	0.9755	0.0240	0.0250	**0.9725**	0.0220	0.0330
70%	0.9380	0.0720	0.0520	0.9868	0.0125	0.0140	**0.9872**	0.0110	0.0145
80%	0.9470	0.0590	0.0470	0.9930	0.0060	0.0080	**0.9918**	0.0085	0.0080
90%	0.9515	0.0535	0.0435	0.9952	0.0030	0.0065	**0.9968**	0.0030	0.0035
100%	0.9643	0.0395	0.0320	0.9978	0.0020	0.0025	**0.9988**	0.0020	0.0005

Table 5. Detection performance for Miao's method with $\eta = 4$ [18].

Embedding rate	Ren's method [21]			Tian's method [22]			The proposed method		
	ACC	FPR	FNR	ACC	FPR	FNR	ACC	FPR	FNR
10%	0.6262	0.4130	0.3345	0.6470	0.3575	0.3485	**0.6847**	0.3240	0.3065
20%	0.7610	0.2890	0.1890	0.8270	0.1890	0.1570	**0.8353**	0.1665	0.1630
30%	0.8672	0.1480	0.1175	0.9240	0.0770	0.0750	**0.9253**	0.0705	0.0790
40%	0.9300	0.0820	0.0580	0.9725	0.0250	0.0300	**0.9722**	0.0255	0.0300
50%	0.9600	0.0460	0.0340	0.9880	0.0130	0.0110	**0.9882**	0.0105	0.0130
60%	0.9673	0.0330	0.0325	0.9938	0.0070	0.0055	**0.9950**	0.0050	0.0050
70%	0.9792	0.0225	0.0190	0.9960	0.0035	0.0045	**0.9980**	0.0015	0.0025
80%	0.9852	0.0150	0.0145	0.9988	0.0020	0.0005	**0.9982**	0.0015	0.0020
90%	0.9900	0.0120	0.0080	0.9992	0.0005	0.0010	**0.9995**	0.0005	0.0005
100%	0.9930	0.0075	0.0065	1.0000	0.0000	0.0000	**1.0000**	0.0000	0.0000

reaches over 60% accuracy when the embedding rate is only 10%, nearly the accuracy achieves more than 70% for Miao's method $\eta = 4$ at the same embedding rates, which indicates that although the proposed method has a low feature dimension the steganalysis is more effective than that of the state of the art.

6 Conclusion and Future Work

In this paper, the detection of AMR-based steganography on FCB is researched. The correlations among the pulse positions in the same subframe and in the same track are explored firstly and the self-information of the pulse-pair combinations in the same subframe and the mutual information of the two pulse in the same track are proposed to describe the correlation of pulse-position difference completely. Then, the two kind of features are applied as steganalysis features, whose feature dimension is only

100-dimensional. We proposed an SVM-based steganalysis scheme with the proposed features. The experimental results show that the detection performance of the proposed method is better than the state-of-the-art methods. In the future, we will try to employed some strategies to enhance the detection performance at low embedding rates.

Acknowledgements. 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 Outstanding Youth Scientific and Technological Talents in Fujian Province University under Grant No. MJK2015-54, Promotion Program for Young and Middle-aged Teacher in Science & Technology Research of Huaqiao University under Grant No. ZQN-PY115, Program for Science & Technology Innovation Teams and Leading Talents of Huaqiao University under Grant No.2014KJTD13, Opening Project of Shanghai Key Laboratory of Integrated Administration Technologies for Information Security under Grant No. AGK201710.

References

1. Gaurav, K., Ghanekar, U.: Image steganography based on Canny edge detection, dilation operator and hybrid coding. J. Inf. Secur. Appl. **41**, 41–51 (2018)
2. El_Rahman, S.A.: A comparative analysis of image steganography based on DCT algorithm and steganography tool to hide nuclear reactors confidential information. Comput. Electr. Eng. (2016)
3. Majumder, A., Changder, S.: A novel approach for text steganography: generating text summary using reflection symmetry. Procedia Technol. **10**, 112–120 (2013)
4. Vidhya, P.M., Paul, V.: A method for text steganography using Malayalam text. Procedia Comput. Sci. **46**, 524–531 (2015)
5. Kar, D.C., Mulkey, C.J.: A multi-threshold based audio steganography method. J. Inf. Secur. Appl. **23**, 54–67 (2015)
6. Devi, R.R., Pugazhenthi, D.: Ideal sampling rate to reduce distortion in audio steganography. Procedia Comput. Sci. **85**, 418–424 (2016)
7. Dasgupta, K., Mondal, J.K., Dutta, P.: Optimized video steganography using genetic algorithm (GA). Procedia Technol. **10**, 131–137 (2013)
8. Kar, N., Mandal, K., Bhattacharya, B.: Improved chaos-based video steganography using DNA alphabets. ICT Express **4**, 6–13 (2018)
9. Cheng, J., Kot, A.C.: Steganalysis of halftone image using inverse halftoning. Signal Process. **89**, 1000–1010 (2009)
10. Mazurczyk, W.: VoIP steganography and its detection-a survey. ACM Comput. Surv. **46**, 1–21 (2013)
11. Tian, H., et al.: Optimal matrix embedding for voice-over-IP steganography. Signal Process. **117**, 33–43 (2015)
12. Kassim, M., Rahman, R.A., Aziz, M.A.A., Idris, A., Yusof, M.I.: Performance analysis of VoIP over 3G and 4G LTE network. In: IEEE, Kanazawa, Japan, pp. 37–41 (2017). http://ieeexplore.ieee.org/document/8298391/. Accessed 20 July 2018
13. Xiao, B., Huang, Y., Tang, S.: An approach to information hiding in low bit-rate speech stream. In: Proceedings of the IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference, pp. 1–5 (2008)

14. Liu, P., Li, S., Wang, H.: Steganography in vector quantization process of linear predictive coding for low-bit-rate speech codec. Multimedia Syst. **23**, 485–497 (2017)
15. Huang, Y., Liu, C., Tang, S., Bai, S.: Steganography integration into a low-bit rate speech codec. IEEE Trans. Inf. Forensics Secur. **7**, 1865–1875 (2012)
16. Nishimura, A.: Data hiding in pitch delay data of the adaptive multi-rate narrow-band speech codec. In: 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 483–486 (2009)
17. Geiser, B., Vary, P.: High rate data hiding in ACELP speech codecs. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4005–4008 (2008)
18. Miao, H., Huang, L., Chen, Z., Yang, W., Al-hawbani, A.: A new method for covert communication via 3G encoded speech. Comput. Electr. Eng. **38**, 1490–1501 (2012)
19. Speech Codec Speech Processing Functions; Adaptive Multi-Rate—Wideband (AMR-WB) Speech Codec; Transcoding Functions (Release 6), document 3GPP TS 26.190 V6.1.1 (2005)
20. 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
21. Ren, Y., Cai, T., Tang, M., Wang, L.: AMR steganalysis based on the probability of same pulse position. IEEE Trans. Inf. Forensics Secur. **10**, 1801–1811 (2015)
22. Tian, H., et al.: Steganalysis of adaptive multi-rate speech using statistical characteristics of pulse pairs. Signal Process. **134**, 9–22 (2017)
23. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 (1948)
24. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011)
25. Sklean source. https://github.com/scikit-learn/scikit-learn. Accessed 11 Nov 2018