IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 7, 2019, accepted December 15, 2019, date of publication December 24, 2019,
date of current version January 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962009

Detecting Steganography in Inactive Voice-Over-IP Frames Based on Statistic Characteristics of Fundamental Frequency

HUI TIAN[1,2,3], (Senior Member, IEEE), JIE LIU[1,2,3], CHIN-CHEN CHANG[4,5], (Fellow, IEEE), YONGFENG HUANG[6], (Senior Member, IEEE), AND YIQIAO CAI[1,3], (Member, IEEE)

[1]College of Computer Science and Technology, National Huaqiao University, Xiamen 361021, China
[2]State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[3]Fujian Key Laboratory of Big Data Intelligence and Security, National Huaqiao University, Xiamen 361021, China
[4]Department of Information and Computer Science, Feng Chia University, Taichung 40724, Taiwan
[5]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
[6]Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

Corresponding author: Hui Tian (htian@hqu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61972168, Grant U1536115, and Grant U1405254, in part by the Natural Science Foundation of Fujian Province of China under Grant 2018J01093, in part by the Opening Program of State Key Laboratory of Information Security of China under Grant 2019-ZD-09, in part by the Program for New Century Excellent Talents in Fujian Province University under Grant MJK2016-23, and in part by the Program for Outstanding Youth Scientific and Technological Talents in Fujian Province University under Grant MJK2015-54.

ABSTRACT Steganography in inactive Voice-over-IP frames is a new technique of information hiding, which can achieve large steganographic capacity while maintaining excellent imperceptibility. To prevent the illegitimate use of this technique, the entropy-based and poker test-based steganalysis methods have been presented. However, the detection performance of these two methods is not so good for the cases of having small quantity of inactive frames or low embedding rates. Thus, we present a new steganalysis method based on statistic characteristics of fundamental frequency. Specifically, we employ the statistics for zero-crossing count (ZCC), including the average ZCC of inactive frames, the ratio between the average ZCC of inactive frames and that of all frames, and the difference between the average ZCC of inactive frames and their calibrated versions, to characterize the frame-level dynamic characteristic of speech signals; we utilize the average values of Mel-frequency cepstral coefficients (MFCCs) to represent the invariant characteristic of inactive frames; further, using the feature set consisting of the zero-crossing statistics and average MFCCs, we propose a support-vector-machine based steganalysis for inactive speech frames. The proposed steganalysis method is evaluated with a large number of ITU-T G.723.1 encoded speech samples, and compared with the existing methods. The experimental results demonstrate that the proposed method significantly outperforms the previous ones on detection accuracy, false positive rate and false negative rate for any given embedding rates or using the same number of inactive frames. Particularly, the proposed method can provide accurate detecting results for the existing steganographic methods only using very small quantity of inactive frames, and thereby be employed to detecting potential inactive-frame steganography behaviors in real-time speech streams.

INDEX TERMS Steganography, Steganalysis, Voice over IP, Inactive frames, Fundamental frequency.

I. INTRODUCTION

Steganography is a technique of covert communication by hiding information into digital media (such as image [1], video [2], audio [3] and text [4]) without causing any perceptible distortion. In recent years, Voice over Internet Protocol (VoIP), which enables phone calls based on IP networks, has been widely applied in people's daily life, for its convenience, low costs and high speech quality [5]. With the increasing popularity of VoIP, VoIP-based steganography has attracted extensive attention from research communities [6]–[11]. Compared with traditional carriers, there are

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaochun Cheng.

many advantages of VoIP for information hiding, such as instantaneity, huge amounts of carrier data, high steganographic bandwidth, and alterable conversation length [6]–[9]. Therefore, VoIP-based steganography is popularly regarded as one of ideal solutions for secure communications. However, it might be also abused by lawbreakers, terrorists and hackers for cybercriminal activities, because unauthorized information flow with its help can covertly pass through firewalls and monitors without being noticed [12]. Therefore, in order to improve cybersecurity, it is indispensable to develop the corresponding countermeasure technique, i.e., steganalysis of VoIP, whose primary aim is to detect covert communications based on VoIP accurately [13]–[15]. In this paper, we focus on detecting steganography in inactive frames in VoIP streams, which is still an open problem.

In general, VoIP-based steganography can be divided into two categories [6]–[9]. One employs the network protocols as carriers [10], [16]–[18], while the other hides information by modifying payloads in speech streams [7]–[9], [19]–[28]. Due to its high steganographic capacity, the second category has been the mainstream of VoIP-based steganography. In VoIP, to obtain required low data rates, speech signals are often encoded into digital frame streams using code excited linear prediction (CELP) codecs, such as ITU-T G. 723.1, ITU-T G.729a, Speex, Internet Low Bitrate Codec (iLBC) and adaptive multi-rate (AMR) codec. Accordingly, most of the payload-based steganographic algorithms achieve information hiding by modifying some specific parameters in speech frames, including linear predictive coefficients (LPC) [19]–[21], fixed codebook (FCB) parameters [22]–[24] and adaptive codebook (ACB) parameters [25]–[27]. In addition, differing from the steganographic methods based on the modification of specific parameters, Huang *et al.* [11] presented a novel high-capacity steganographic algorithm by hiding information into inactive frames of VoIP streams. The work suggested that the steganography in inactive frames can achieve much larger steganographic capacity than that in active frames, while maintaining the same imperceptibility. For the speech streams encoded with ITU-T G.723.1 codec at 6.3 kbps mode, the proposed method can obtain the steganographic bandwidth of up to 101 bits per frame. Further, Lin [28] extended the idea into speech streams encoded with ITU-T G.723.1 codec at 5.3 kbps mode, whose experimental results show that the presented method can achieve the steganographic bandwidth of up to 81 bits per frame without causing perceptible degradation of speech quality.

As for the steganalysis of VoIP, there have been also many fruitful studies [13]–[15], [29]–[35]. For example, Lin *et al.* [15] introduced a recurrent neural network to detect quantization index modulation-based steganography for LPC parameters in G.729a speech streams, which can achieve excellent detection performance, even for very short speech samples, and significantly outperforms the steganalysis based on quantization codeword correlation network [29]. To detect steganography for FCB parameters in AMR speech streams, Tian *et al.* [14] presented a support-vector-machine (SVM) based steganalysis method using three kinds of statistical features for pulse pairs, namely, long-term distribution features based on the probability distributions of pulse pairs, short-term invariant features based on Markov transition probabilities of pulse pairs, and track-to-track correlation features based on the joint probability matrices of pulse pairs. Moreover, they proposed a feature selection mechanism based on adaptive boosting to optimize the feature set as well as reduce its dimension. The experimental results demonstrate their method can effectively detect the state-of-the-art steganography based on FCB parameters, and achieve much better detection performance than the steganalysis based on Miao *et al.* [31] and on the probability of same pulse position [32]. At the aspect of detecting steganography for ACB parameters, Ren *et al.* [34] presented an SVM-based steganalysis method, which uses the matrix of the second-order difference of pitch delay (MSDPD) as the detection features. Moreover, they employed the calibration method to obtain the calibrated MSDPD features to further enhance the detection accuracy. The experimental results demonstrated that it is by far the best method for detecting steganography based on ACB parameters in AMR speech streams. By contrast, the steganalysis for inactive speech frames is largely unexplored. Recently, two classical statistics, i.e., entropy and poker test statistic, were employed as the steganalysis features to detect steganography in inactive speech frames [35]. The experimental results show that these two methods are feasible, while the latter is better than the former. However, the detection performance of these methods is not so good for the cases of short sample lengths or low embedding rates. Moreover, our observations through research and experiments suggest that the embedding operations would significantly impact on the statistical characteristics of fundamental frequency for inactive speech frames. Specifically, the statistics for zero-crossing count (ZCC) are employed as the steganalysis features, including the average ZCC of inactive frames, the ratio between the average ZCC of inactive frames and that of all frames, and the difference between the average ZCC of inactive frames and their calibrated versions. These statistics are used to characterize the frame-level dynamic characteristic of speech signals. Moreover, the average values of Mel-frequency cepstral coefficients (MFCCs) are employed to represent the invariant characteristic of inactive frames. Note that, differing from the previous Mel-frequency cepstrum-based steganalysis schemes [36], [37], we directly employ the original 12-dimensional MFCCs without calculating the first-order or second-order differences of MFCCs, because the MFCCs in inactive frames are independent, meaning that there is no correlation between any two MFCCs. Further, using the feature set consisting of the zero-crossing statistics and average values of MFCCs, an SVM-based steganalysis for inactive speech frames is presented. The proposed method is evaluated with a large number of ITU-T G.723.1 encoded speech samples, and compared with entropy-based [38] and poker test-based [35]

methods. The experimental results demonstrate that the proposed method significantly outperforms the previous ones in detection accuracy, false positive rate and false negative rate for any given embedding rates or using the same quantity of inactive frames.

The rest of this paper is organized as follow. Section 2 analyses how the steganography in inactive frame impacts on the statistical characteristics of fundamental frequency, and presents two types of detection features, i.e., the zero-crossing statistics and average values of MFCCs. An SVM-based steganalysis scheme is proposed in Section 3. The performance evaluation through comprehensive experiments is described in Section 4. Finally, Section 5 offers the concluding remarks.

II. CHARACTERISTICS OF FUNDMENTAL FREQUENCY FOR INACTIVE FRAMES

Assume that a speech signal contains N_S inactive frames, each of which is sampled n times. Let the set of inactive frames be $S = \{s_i | i = 1, 2, \ldots, N_S\}$, and each inactive frame be $s_i = \{r_{i,j} | j = 1, 2, \ldots, n\}$, where $r_{i,j}$ is the j-th sample of s_i. For each inactive frame, the encoding process can be described as

$$s_i^* = \varphi(s_i), \tag{1}$$

where s_i^* is the i-th encoded inactive frame. Accordingly, the set of the encoded inactive frames can be denoted as $S^* = \{s_i^* | i = 1, 2, \ldots, N_S\}$. Further, the process for embedding secret information into an encoded inactive frame can be stated as

$$\tilde{s}_i = \psi\left(s_i^*\right), \tag{2}$$

where $\psi(.)$ is the steganographic operation and $\tilde{s}_i$ is the steganographic version of the i-th encoded inactive frame and the decoding process of s_i^* is

$$\varphi^{-1}(s_i^*) = s_i, \tag{3}$$

According to the additive noise model for steganography [39], [40], the decoding process of $\tilde{s}_i$ is

$$\varphi^{-1}(\tilde{s}_i) = \varphi^{-1}(s_i^*) + \varepsilon_i = s_i + \varepsilon_i, \tag{4}$$

where ε_i is the additive noise generated by the steganographic operation on the i-th inactive frame. This equation suggests that the steganographic operation would inevitably impact on the signal decoding of inactive frames.

In addition, fundamental frequency estimation is popularly applied in the field of speech signal processing [43]–[46], particularly in voice activity detection [45], [46]. Inspired by these successful applications, we study the impact of steganography in inactive frames on fundamental frequency characteristics, and find out that the statistics for zero-crossing count and Mel-frequency cepstral coefficients are eminently suitable for discriminating the cover and steganographic speech samples. In the following text, we will introduce how we exploit these fundamental frequency statistics as the steganalysis features in detail.

A. STATISTICS FOR ZERO-CROSSING COUNTS

In the field of signal processing, zero-crossing counts (ZCC) are widely employed to characterize the frequency of a given signal [47]. Particularly, the zero-crossing counts can help differentiate between active and silent speech. In general, for the i-th speech frame in a sample, denoted as $f_i = \{r_{i,j} | j = 1, 2, \ldots, n\}$, the ZCC λ can be calculated as

$$\lambda = \sum_{j=2}^{n} \delta\left(\xi(r_{i,j}) \cdot \xi(r_{i,j-1}) < 0\right), \tag{5}$$

where $\xi(x)$ is the sign function, namely,

$$\xi(x) = \begin{cases} 1, & if\ x > 0 \\ 0, & if\ x = 0 \\ -1, & if\ x < 0, \end{cases} \tag{6}$$

and $\delta(x)$ is a discriminant function, which is given by

$$\delta\left(x\right) = \begin{cases} 0, & if\ x\ is\ true \\ 1, & else. \end{cases} \tag{7}$$

For a normal inactive frame, the ZCC should be 0 in theory, since the value of each sample in the inactive frame is equal to 0. However, as mentioned above, if secret information is embedded into the inactive frame, the values of some samples are no longer equal to 0, due to the noises induced by steganography. Accordingly, the ZCC for a steganographic inactive frame would be not equal to 0. In this sense, the ZCC can be used to distinguish between the normal and steganographic inactive frames. Moreover, with the increase of embedding rate, more sample values in the inactive frame would be modified by the steganographic operation, which suggests that the change of the ZCC for a steganographic inactive frame at a high embedding rate is larger than that at a low embedding rate.

In addition, because there are different numbers of inactive frames in different speech samples, it is hard to use all the ZCCs of the inactive frames as the detection feature. Instead, the average ZCC of the inactive frames in each speech sample is utilized in practice. To verify the above deduction, we compare the distribution of the average ZCCs for randomly chosen 1000 cover speech samples with that for the corresponding steganographic samples at the embedding rate of 100%, as shown in Figure 1. The experimental results show there are obvious distinctions between the average ZCCs for cover speech samples and those for steganographic samples, meaning that it is feasible to employ the average ZCC as a steganalysis feature to detect the steganographic behavior in inactive frames.

In the steganography for inactive frames, the secret information is embedded into inactive frames. Thus, only the ZCCs of inactive frames would increase, while those of active frames are unaffected. For a given cover speech sample, the ratio between the average ZCC of inactive frames and that of all frames, denoted as ω, can be calculated as

$$\omega = \frac{\overline{\lambda_S}}{\overline{\lambda}}, \tag{8}$$

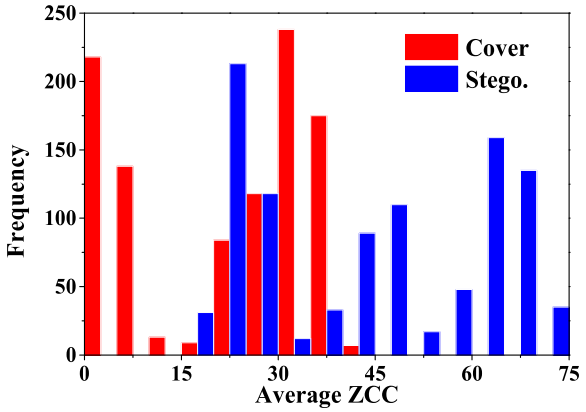


FIGURE 1. The distribution contrast of average ZCCs between cover samples and steganographic samples at the embedding rate of 100%.

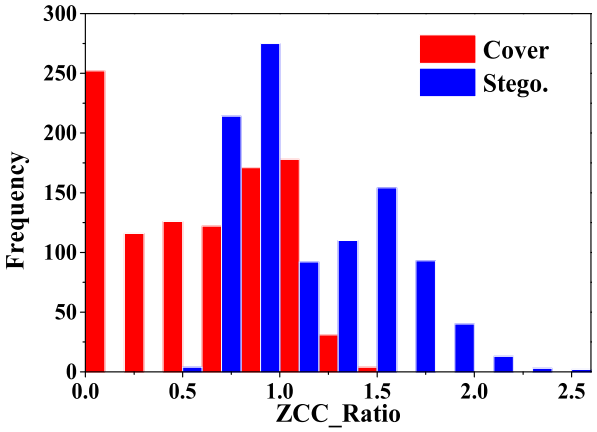


FIGURE 2. The distribution contrast of ZCC_Ratios between cover samples and steganographic samples at the embedding rate of 100%.

where $\bar{\lambda}_S$ is the average ZCC of inactive frames, and $\bar{\lambda}$ is the average ZCC of all frames. Assume that, the number of inactive frames is N_S, the sum of ZCCs of all inactive frames is Z_S, the number of all active frames is N_A, the sum of ZCCs of all active frames is Z_A, then ω can be further written as

$$\omega = \frac{\frac{Z_S}{N_S}}{\frac{Z_S+Z_A}{N_S+N_A}} = \frac{N_S+N_A}{N_S} \cdot \frac{1}{1+\frac{Z_A}{Z_S}}, \tag{9}$$

Apparently, for the steganographic version of the given speech sample, the sum of ZCCs of all inactive frames (denoted by Z'_S) is larger than Z_S. Accordingly, we have $\omega' > \omega$, where ω' is the ratio between the average ZCC of inactive frames and that of all frames for the steganographic sample. That is to say, the ratio between the average ZCC of inactive frames and that of all frames (simply called ZCC_Ratio) would be changed by the steganographic operation, and could be thereby employed to detect the steganography in inactive frames. Similarly, we compare the distribution of ZCC_Ratios for randomly chosen 1000 cover speech samples with that of ZCC_Ratios for the corresponding steganographic samples at the embedding rate of 100%, as shown in Figure 2. The experiment results show that there are obvious distinctions between the ZCC_Ratios for cover speech samples and those for steganographic samples, and thereby demonstrate the feasibility of using ZCC_Ratio as a steganalysis feature.

In addition, like Ren et al. 's work [34], we employ the calibration technique to estimate the cover signal of a given speech signal. To obtain the calibrated speech sample, the given speech sample, is first recompressed, namely, encoded and decoded again, whether it is a cover or steganographic one. As shown in Figure 3, we can respectively extract the average ZCC from the given sample and the average calibrated ZCC from the calibrated version. Finally, we can obtain the third type of ZCC statistic, i.e., the difference between the average ZCC of inactive frames and their calibrated versions, called DIF-ZCC and denoted as ν,

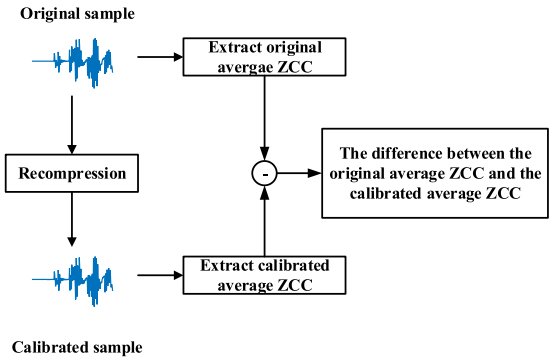


FIGURE 3. The extraction process for the third type of ZCC statistic.

namely,

$$\nu = \overline{\lambda_O} - \overline{\lambda_C}, \tag{10}$$

where $\bar{\lambda}_O$ is the average ZCC of inactive frames in the original speech sample, and $\bar{\lambda}_C$ is the average ZCC of inactive frames in the calibrated speech sample. Similarly, we compare the distribution of DIF-ZCC for randomly chosen 1000 cover speech samples with that of the corresponding steganographic samples at embedding rate of 100% as shown in Figure 4. The experimental results show that there are obvious distinctions between them. Thus, it is valid to employ DIF-ZCC as a steganalysis feature.

B. MEL-FREQUENCY CEPSTRAL COEFFICIENTS FOR INACTIVE FRAMES

Mel-frequency cepstral coefficients (MFCCs) are often used to describe the frequency characteristics similar to the human auditory system's response, and commonly applied in speech processing. In general, MFCCs are calculated by applying a Mel-scaled filter-bank to the short-term fast Fourier transform (FFT) magnitude spectrum to obtain a perceptually meaningful smoothed gross spectrum [48], [49]. Figure 5 shows the procedure of extracting the FFT-based MFCCs from a speech signal.

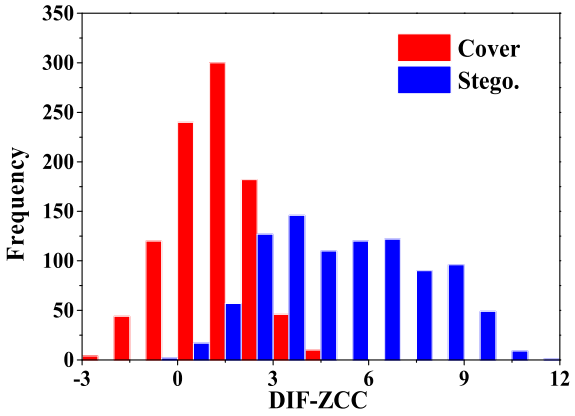


FIGURE 4. The distribution contrast of DIF-ZCCs between cover samples and steganographic samples at the embedding rate of 100%.

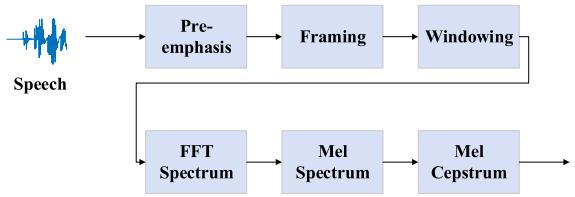


FIGURE 5. The procedure of extracting the FFT-based MFCCs [51].

As mentioned above, in accordance with the additive noise model for steganography, the FFT spectrum of a steganographic inactive frame $\mathcal{S}'$ can be stated as

$$\mathcal{S}' = \mathcal{S} + \mathcal{S}_\varepsilon, \tag{11}$$

where $\mathcal{S}$ is the FFT spectrum of the corresponding cover inactive frame and $\mathcal{S}_\varepsilon$ is the FFT spectrum of the additive noise. Further, let $\mathscr{F}_{L,i}$, $\mathscr{F}_{C,i}$ and $\mathscr{F}_{H,i}$ respectively denote the low limit frequency, center frequency and high limit frequency of the i-th ($i = 1, 2, \ldots, T$) triangular overlapping window of Mel-scaled filter-bank, where T is the number of the involved filters. T is the number of triangular overlapping windows, and usually set as 24. The relationship among the adjacent triangular overlapping windows can be described below

$$\mathscr{F}_{C,i} = \mathscr{F}_{H,i-1} = \mathscr{F}_{L,i+1}. \tag{12}$$

All the spectrums for the frames would be passed through their corresponding Mel-filters. For the given inactive frame, an output value of the i-th Mel-filter, denoted as θ_i, can be obtained by calculating

$$\theta_i = \sum_{x=\mathscr{F}_{L,i}}^{\mathscr{F}_{H,i}} W_i(x) \left|\mathcal{S}\right|, \tag{13}$$

where $W_i(x)$ is the frequency response function of the i-th Mel-filter, and can be determined as

$$W_i(x) = \begin{cases} \dfrac{x - \mathscr{F}_{L,i}}{\mathscr{F}_{C,i} - \mathscr{F}_{L,i}}, & \mathscr{F}_{L,i} \le x \le \mathscr{F}_{C,i} \\ \dfrac{\mathscr{F}_{H,i} - x}{\mathscr{F}_{H,i} - \mathscr{F}_{C,i}}, & \mathscr{F}_{C,i} \le x \le \mathscr{F}_{H,i} \end{cases} \tag{14}$$

Correspondingly, the output of the i-th Mel-filter for the steganographic inactive frame, denoted as θ_i', is given by

$$\theta_i' = \sum_{x=\mathscr{F}_{L,i}}^{\mathscr{F}_{H,i}} W_i(x) \left|\mathcal{S} + \mathcal{S}_\varepsilon\right|. \tag{15}$$

MFCCs are the result of a discrete cosine transform (DCT) operation on the logarithm of the Mel-filter outputs. There are 24 filters in the Mel bank, which leads to 24 DCT coefficients. However, due to the decorrelation property of DCT, only the first few coefficients are chosen in practice. In this work, L is equal to 12, following the convention of speech processing [50]–[52]. For the given cover inactive speech frame, each MFCC, denoted as $\eta_j (j = 1, 2, \ldots, L, L = 12)$, which can be written as

$$\eta_j = \sum_{i=1}^{T} \left(\lg\left(\theta_i\right)^* \cos\left(\frac{\pi(i+0.5)j}{T} \right) \right), \tag{16}$$

Correspondingly, each MFCC for the steganographic inactive speech frame, denoted as $\eta_j' (j = 1, 2, \ldots, L)$, is

$$\eta_j' = \sum_{i=1}^{T} \left(\lg\left(\theta_i'\right)^* \cos\left(\frac{\pi(i+0.5)j}{T} \right) \right), \tag{17}$$

Apparently, $\exists j \in [1, L]$, $\eta_j \neq \eta_j'$, since it is largely possible that $\theta_i \neq \theta_i'$ for $i \in [1, T]$, which suggests that the set of MFCCs for the inactive frames can be employed as the steganalysis feature.

To verify this deduction, we compare the distributions of MFCCs for the inactive frames in randomly chosen 1000 cover speech samples with those in the corresponding steganographic samples at embedding rate of 100%, as shown in Figure 6. All the MFCCs are calculated with a window of 256 samples and overlapping length of 80 sampling points. The experimental results show that the steganographic operation indeed induces effects on the distributions of MFCCs for the inactive frames, although the impacts caused on the different MFCCs vary. Therefore, we can safely conclude that the set of MFCCs for the inactive frames is very suitable for distinguishing between the cover and the steganographic samples.

Resembling the average ZCC, the average MFCCs (i.e., $\bar{\eta}_1, \bar{\eta}_2, \ldots, \bar{\eta}_{12}$) of the inactive frames in each speech sample are employed as the steganalysis feature, since there are different numbers of inactive frames in different speech samples.

III. PROPOSED STEGANALYSIS SCHEME

Combining the above two types of features, namely, the statistics for zero-crossing counts and the average MFCCs, we can obtain a 15-dimensional steganalysis feature $\phi = \{\bar{\lambda}_s, \omega, \nu, \bar{\eta}_1, \bar{\eta}_2, \ldots, \bar{\eta}_{12}\}$. Further, incorporating the SVM [53], we present a steganalysis scheme, as shown in Figure 7, which includes two processes. Specifically, the training process includes the following steps:

STEP 1: Sample preparation. Collect a large number of speech samples, encode them with ITU-T G.723.1 codec

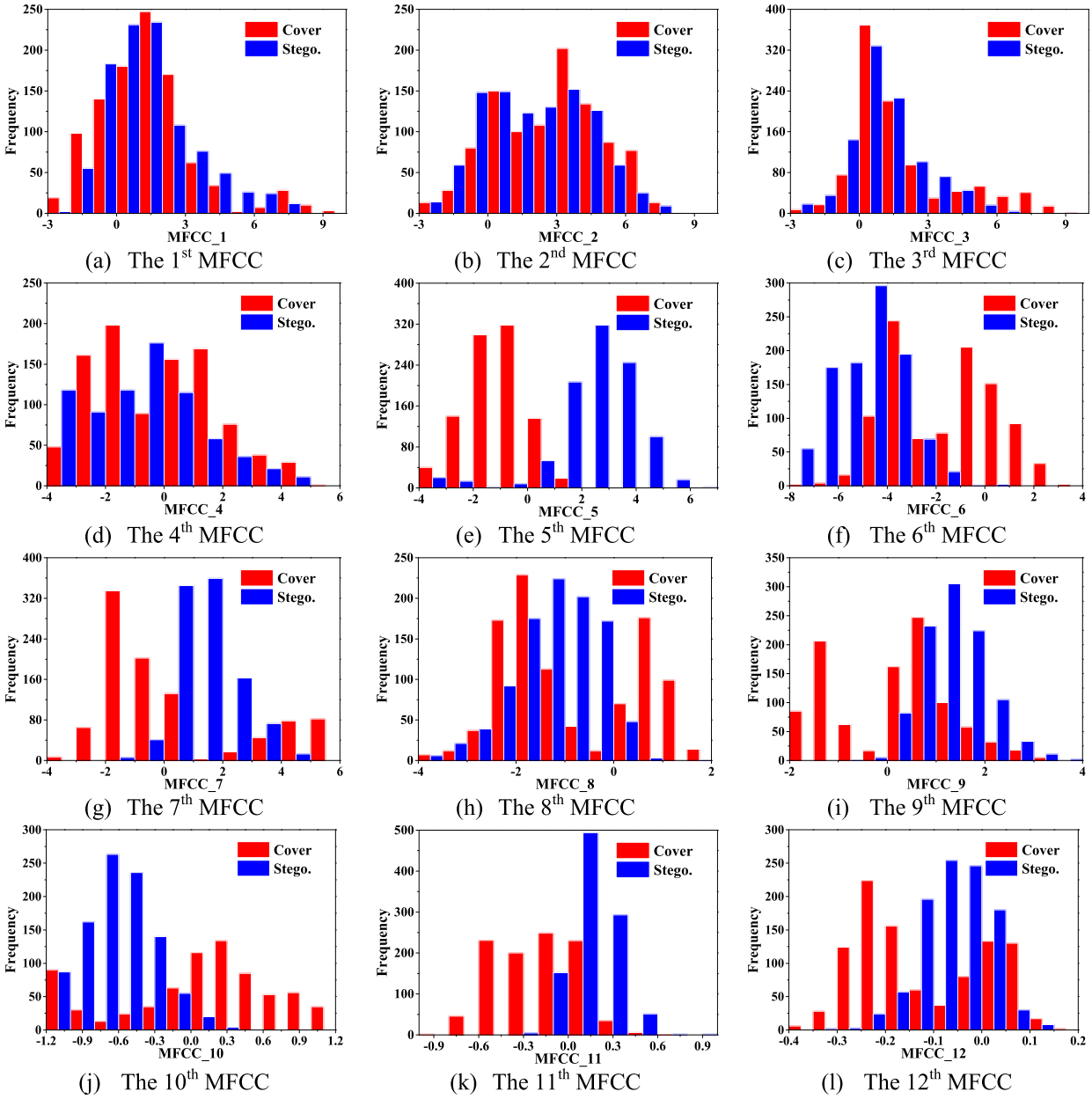


FIGURE 6. The distribution contrast of MFCCs for the inactive frames between cover samples and their steganographic samples at the embedding rate of 100%.

at 6.3 kbps (or 5.3 kbps) mode, and finally embed random information into these samples with the steganography for inactive frames at the designated embedding rate.

STEP 2: Feature extraction. For each sample, extract the 15-dimensional steganalysis feature ϕ.

STEP 3: Classifier training. Train the classifier based on SVM with the feature set ϕ.

Correspondingly, the detection process consists of two steps as follows.

STEP 1: Feature extraction. Extract the feature set ϕ from each sample to be detect.

STEP 2: Detection. Input the feature set into the well-established SVM-based classifier, and decide whether the given test sample contains secret information in accordance with the output of the classifier.

IV. PERFORMANCE EVALUATION

A. EXPERIMENTAL SETUP

To evaluate the performance of our proposed scheme, we collect a total of 2200 ten-second speech samples, which are PCM coded files with 8 kHz sampling rate, 16 bits quantization and mono. The sample set consists of two categories, i.e., English and Chinese. In each category, there are male and female speech samples. All speech samples are encoded with the ITU-T G.723.1 codec at 6.3 kbps mode and that

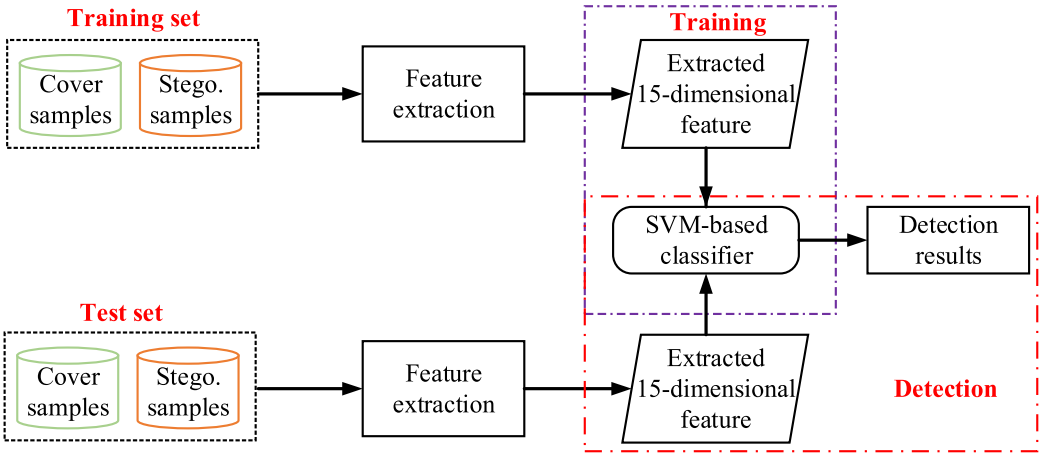


FIGURE 7. The proposed SVM-based scheme for detecting the steganography in inactive frames.

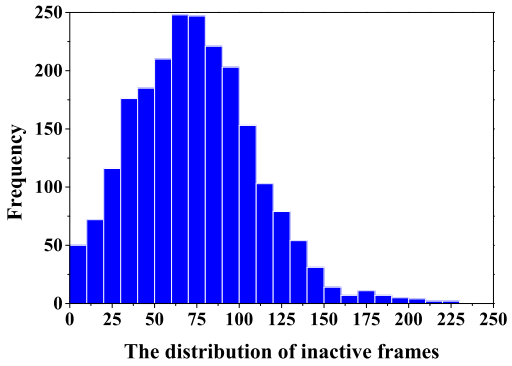


FIGURE 8. The distribution for the number of inactive frames in all the samples.

at 5.3 kbps mode, respectively. Each speech sample after encoding includes 333 frames. Figure 8 shows the distribution for the numbers of inactive frames in all the samples, which indicates that each speech sample contains a certain number of inactive frames, and the numbers of inactive frames in different speech samples are various. Moreover, in the steganographic experiments, Huang et al.'s method [11] and Lin's method [28] are respectively carried out on the speech samples encoded at 6.3 kbps mode and those encoded at 5.3 kbps mode. In all the experiments, the embedded messages are randomly generated.

In this section, we evaluate the performance of the proposed method, and compare it with the entropy-based and poker test-based methods [35]. In the steganalysis experiments, all the SVM-based classifiers are implemented based on LibSVM [53] with RBF kernel, where the default parameter setting is adopted. In each steganalysis, three statistics, namely, accuracy (ACC), false positive rate (FPR), and false negative rate (FNR) are employed to evaluate the detection performance of the steganalysis schemes. ACC is the proportion of true results and is calculated by

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

where N_{TP} is the total of true positives; N_{TN} is the total of true negatives; N_{FP} is the total of false negatives and N_{FN} is the total of false negatives. FPR is calculated as the ratio between the number of negatives wrong categories as positives and the total number of actual negatives, which is given by

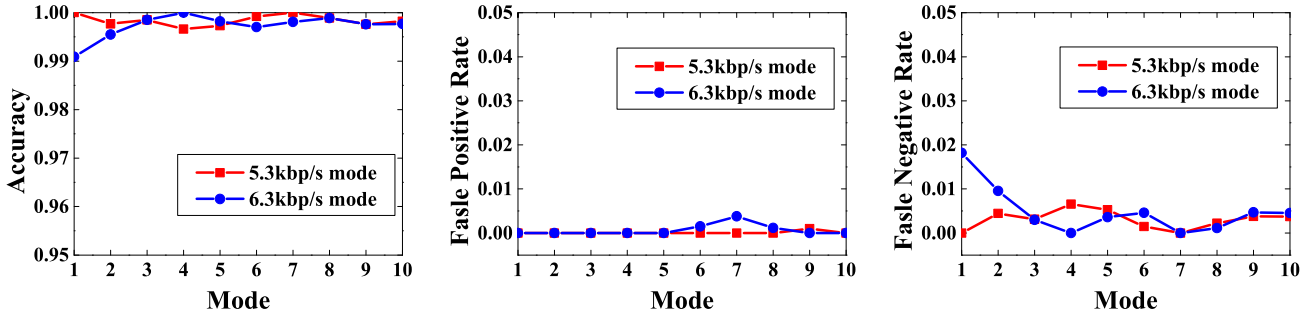
$$FPR = \frac{N_{FP}}{N_{FP} + N_{TN}}, \tag{19}$$

FNR is calculated as the ratio between the number of positives wrong categories as negatives and the total number of actual

TABLE 1. Ten test modes.

Mode	The total number of employed samples	The size of training set	The size of test set
1	440	220	220
2	880	440	440
3	1320	660	660
4	1760	880	880
5	2200	1100	1100
6	2640	1320	1320
7	3080	1540	1540
8	3520	1760	1760
9	3960	1980	1980
10	4400	2200	2200

Note: in the i-th (i =1, 2, …, 10) mode, the employed N = 440 * i speech samples include 220 * i cover sample and their steganographic versions, where the 110 * i cover samples with odd indices and their steganographic versions are used as the training set, and the remainder 220 * i samples (i.e., the 110 * i cover samples with even indices and their steganographic versions) are used as the test set.



(a) Statistical result of ACC (b) Statistical result of FPR (c) Statistical result of FNR

FIGURE 9. The detection performance for different test modes.

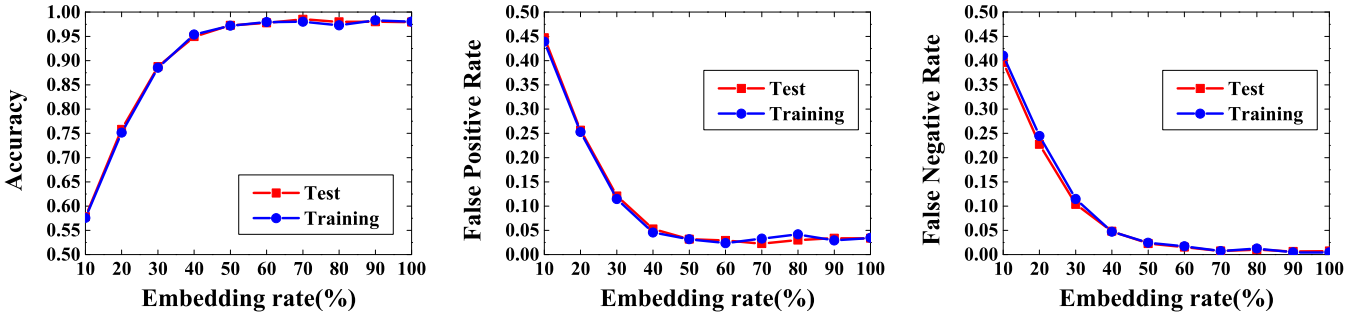
positives, which is expressed as

$$FNR = \frac{N_{FN}}{N_{TP} + N_{FN}}, \tag{20}$$

B. BASIC PERFORMANCE ANALYISIS

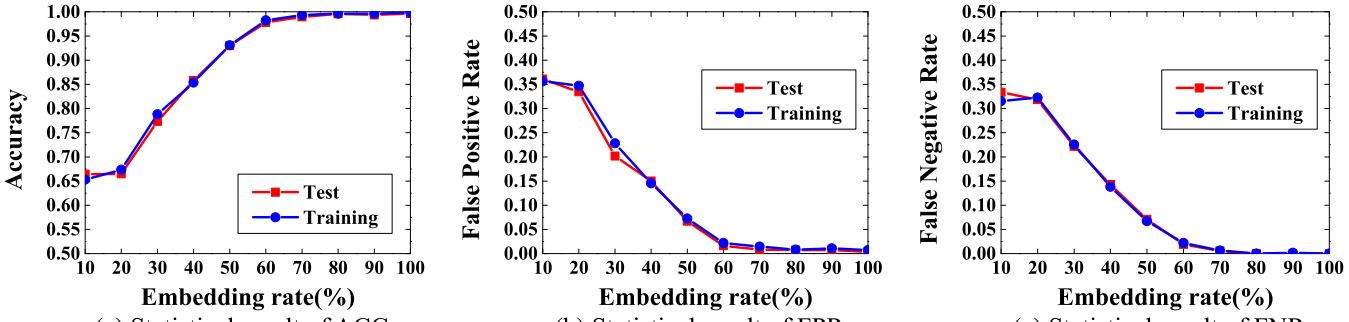
To verify the effect of the proposed feature set, we define ten test modes through changing the sample numbers of training and test sets, as shown in TABLE 1. For each mode, we carry out the experiments with the ten-second speech samples respectively encoded at 5.3 kbps mode and 6.3 kbps mode. All the steganographic samples are produced at the embedding rate of 100%. Figure 9 shows the experimental results for the ten test modes. From the results, we can learn that the detection accuracies are larger than 99% in any cases, indicating that the presented steganalysis feature set is highly effective. Moreover, even with small numbers of training samples, we can obtain good classifier model. In the following experiments, however, to obtain the best classifier model as well as achieve the most reliable detection results, we carry out each steganalysis experiment with the mode 10.

In addition, to evaluate the performance of the presented scheme, we compare the detection performance between the training sets and test sets at various embedding rates (from 10% to 100%), as shown in Figures 10 and 11. From the experimental results, we can learn the following facts: First,



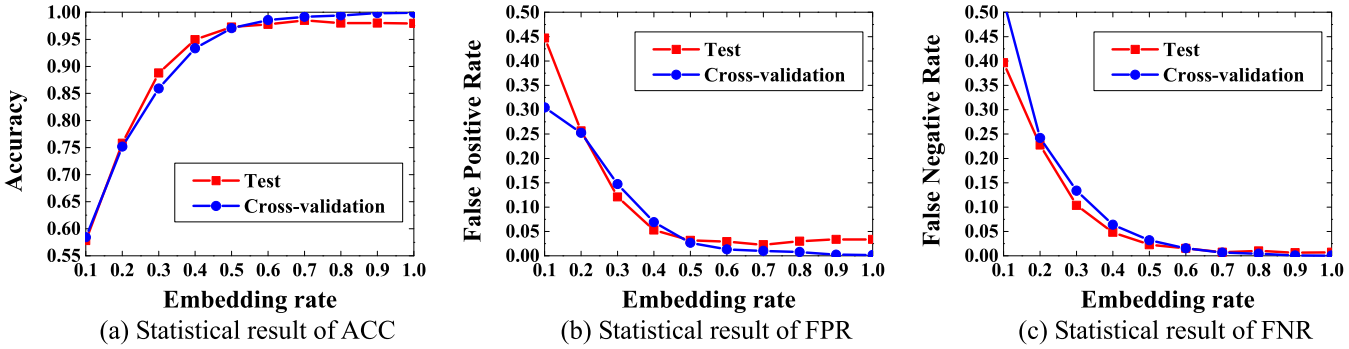
(a) Statistical result of ACC (b) Statistical result of FPR (c) Statistical result of FNR

FIGURE 10. The contrast of performances on training and testing sets at 5.3 kbps mode.



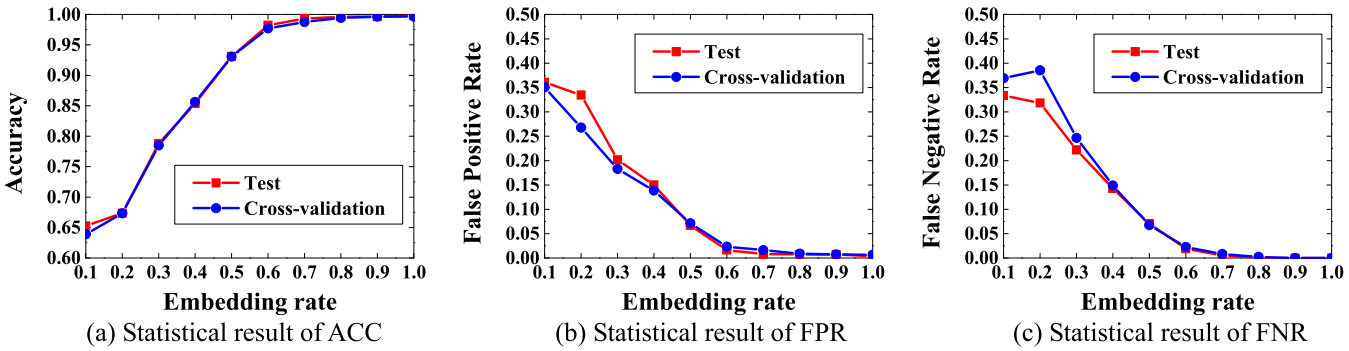
(a) Statistical result of ACC (b) Statistical result of FPR (c) Statistical result of FNR

FIGURE 11. The contrast of performances on training and testing sets at 6.3 kbps mode.



(a) Statistical result of ACC (b) Statistical result of FPR (c) Statistical result of FNR

FIGURE 12. The performance contrast of test and cross validation at 5.3 kbps mode.



(a) Statistical result of ACC (b) Statistical result of FPR (c) Statistical result of FNR

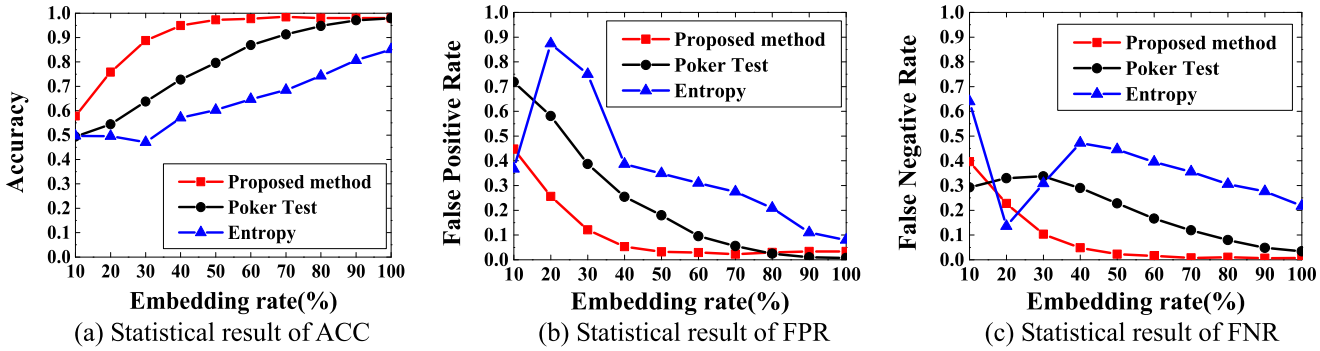
FIGURE 13. The performance contrast of test and cross validation at 6.3 kbps mode.

there is almost no difference between the detection results for the training sets and test sets, meaning that there is no overfitting in all the training processes. Second, at the cases of embedding rates not smaller than 30%, the detection accuracies are larger than 80%, indicating the classifier models in these cases are relatively good. However, at the cases of very low embedding rate (e.g., smaller than 20%), there is much room to improve the detection accuracies, meaning the classifier models in the cases are somewhat underfitting. That is to say, improving the detection performance at the cases of very low embedding rates is still a challenge and deserves further study.

To further evaluate the generalization ability of the proposed model, we conduct 5-fold cross-validation at various embedding rates (from 10% to 100%). Specifically, for each embedding rate, the training set, including 1100 cover samples and 1100 corresponding steganographic versions, are randomly divided into 5 subsets. In each experiment, four subsets are used to train the model, and the remaining one is employed for test. The average ACC, FPR and FNR for all the five test subsets are considered as the final results of the 5-fold cross-validation. The experimental results for various embedding rates are shown in Figures 12 and 13. From the charts, we can learn that there are only very slight differences between the results of 5-fold cross-validation and those for the test sets, indicating that the proposed model has a good generalization ability.

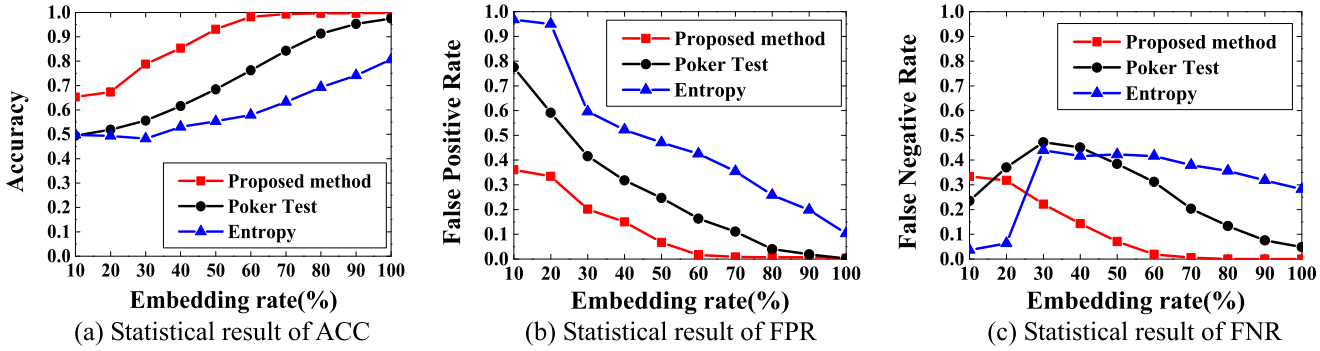
C. PERFORMANCE COMPARISON WITH PREVIOUS METHODS

We first compare the proposed method with the entropy-based and poker test-based methods using the speech sample at various embedding rates (from 10% to 100%). Figures 14 and 15 show the experimental results for the speech samples respectively encoded at 5.3 kbps mode and those encoded at 6.3 kbps mode, from which we can learn the following facts: First, for all the three detection methods for steganography in inactive frames, the accuracy increases with the embedding rate of the steganographic samples, which means that the detection performance has a positive correlation with the adopted embedding rate of the steganographic methods. Second, for both Huang et al.'s method [11] performed on the speech samples encoded at 6.3 kbps mode and Lin's method [28] performed on the speech samples encoded at 5.3 kbps mode, the proposed steganalysis method can achieve better detection performance than the entropy-based and poker test-based methods, particularly at the cases of relatively low embedding rates. For example, for detecting Huang et al.'s method, the accuracy of the proposed method is higher than 85% when the embedding rate is only 40%, while the poker test-based method achieves the similar accuracy rate when the embedding rate is larger than 70%, and the entropy-based method cannot achieve this accuracy rate even if the embedding rate is 100%; for detecting Lin's method, the accuracy of the proposed method is higher than 88%



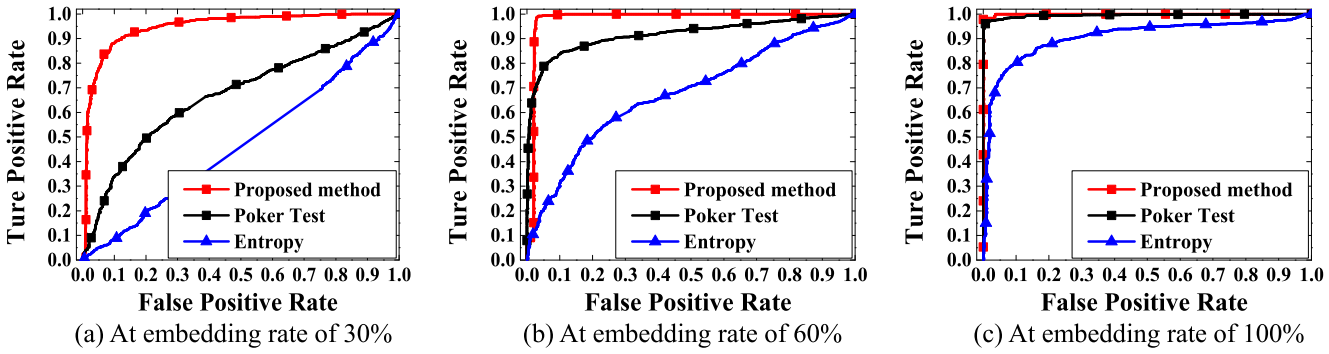
(a) Statistical result of ACC (b) Statistical result of FPR (c) Statistical result of FNR

FIGURE 14. Performance comparison of the proposed method and state-of-the-art methods at various embedding rates at 5.3 kbps mode.



(a) Statistical result of ACC (b) Statistical result of FPR (c) Statistical result of FNR

FIGURE 15. Performance comparison of the proposed method and state-of-the-art methods at various embedding rates at 6.3 kbps mode.



(a) At embedding rate of 30% (b) At embedding rate of 60% (c) At embedding rate of 100%

FIGURE 16. The ROC curves of the proposed method and state-of-the-art methods at 5.3 kbps mode.

when the embedding rate is only 30%, while the poker test-based method achieves the similar accuracy rate when the embedding rate is larger than 60% and even if the embedding rate is 100% the entropy-based method cannot achieve this accuracy rate. To further evaluate the performance of the three steganalysis methods, the receiver-operating-characteristic (ROC) curves for detecting the existing two steganographic methods at typical embedding rates of 30%, 60% and 100%, are shown in Figures 16 and 17. The results demonstrate once again that the proposed method significantly outperforms the entropy-based and poker test-based methods in detection performance, particularly at the relatively low embedding rates.

In addition, we evaluate the performance of the three steganalysis methods for detecting the existing steganographic methods at the embedding rate of 100%, using various small quantities (from 1 to 10) of inactive frames. Figures 18 and 19 show the experimental results for the speech samples respectively encoded at 5.3 kbps mode and those encoded at 6.3 kbps mode, from which we can learn the following facts: First, for all the three steganalysis methods, the detection performance has a positive correlation with the number of the used inactive frames. Overall, the more the used inactive frames, the better the detection performance. Second, the proposed steganalysis method can achieve much better detection performance than the entropy-based and poker

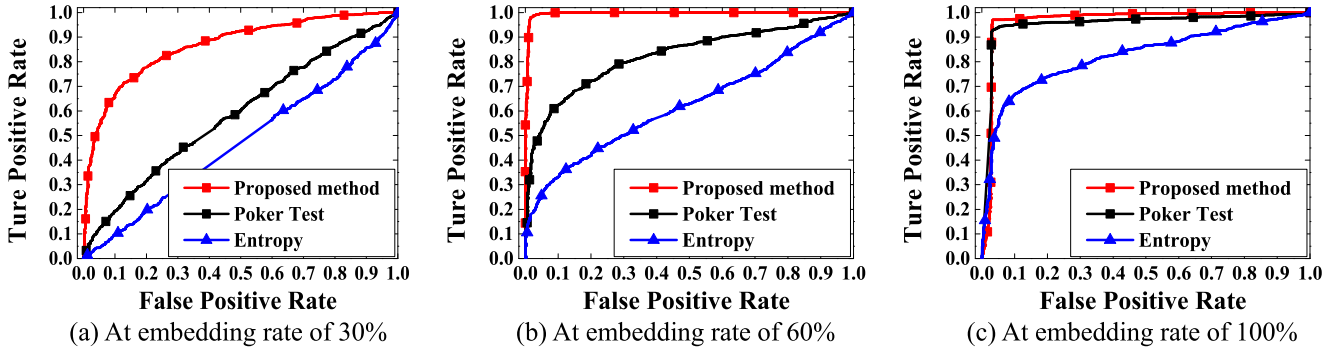


FIGURE 17. The ROC curves of the proposed method and state-of-the-art methods at 6.3 kbps mode.

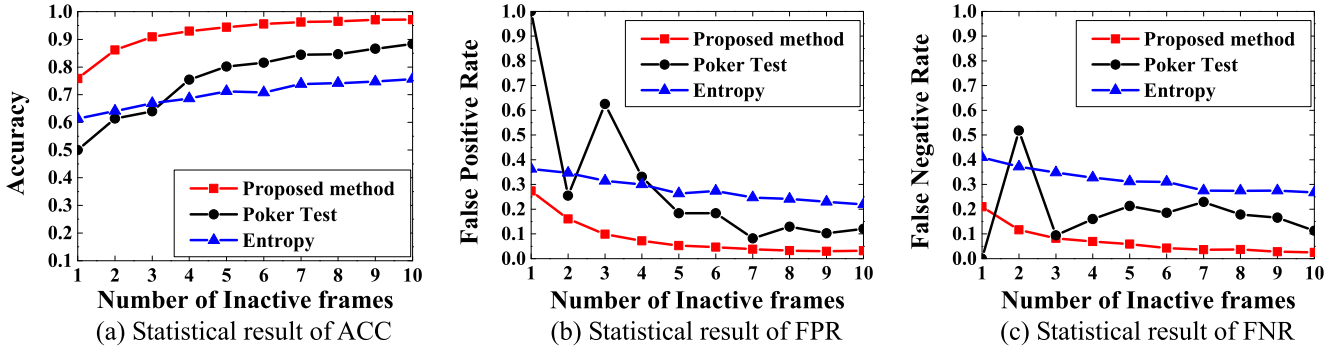


FIGURE 18. Performance comparison of the proposed method and state-of-the-art methods using various quantities of inactive frames at 5.3 kbps mode.

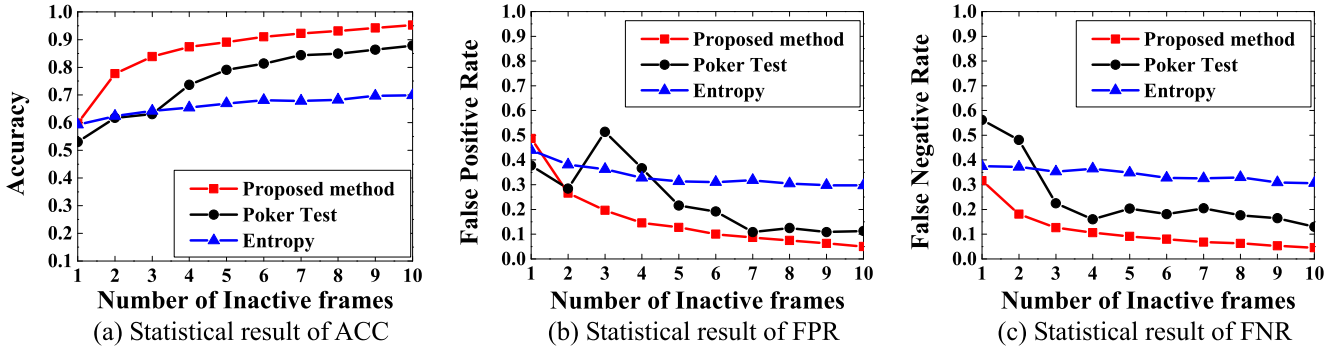


FIGURE 19. Performance comparison of the proposed method and state-of-the-art methods using various quantities of inactive frames at 6.3 kbps mode.

test-based methods in all cases. For example, for detecting Huang et al.'s method, the accuracy of the proposed method is higher than 83% using only three inactive frames, while entropy-based method cannot achieve this accuracy rate even using ten inactive frames, and the poker test-based method needs at least seven inactive frames to achieve this similar accuracy rate; for detecting Lin's method, the accuracy of the proposed method is higher than 86% using only two inactive frames, while entropy-based method cannot achieve this accuracy rate even using ten inactive frames, and the poker test-based method needs at least nine inactive frames to achieve this similar accuracy rate. To sum up, the experimental results demonstrate again that, for the case of detecting the steganography in the same small quantity of inactive frames, the proposed steganalysis method outperforms the existing methods in detection performance. Particularly, the proposed steganalysis method can effectively detect the existing steganographic methods even using very small quantities of inactive frames.

V. CONCLUSION

Steganography in inactive speech frames is a new effective technique of covert communication based on VoIP, which can achieve large steganographic capacity while maintaining

excellent embedding transparency. However, its illegitimate use by terrorists and lawbreakers would facilitate cybercrimes and pose a serious threat to cybersecurity. Thus, in this paper, we aim to develop an efficient steganalysis technique to detect this type of steganography. Differing from the existing entropy-based and poker test-based methods, we employ the statistics for ZCC, including the average ZCC of inactive frames, the ratio between the average ZCC of inactive frames and that of all frames, and the difference between the average ZCC of inactive frames and their calibrated versions, to characterize the frame-level dynamic characteristic of speech signals; moreover, we utilize the average values of MFCCs to represent the invariant characteristic of inactive frames. Further, an SVM-based steganalysis for inactive speech frames is presented. The proposed steganalysis method is evaluated with a great quantity of ITU-T G.723.1 encoded speech samples, and compared with the existing methods. The experimental results show that the proposed method significantly outperforms the previous ones in detection performance for any given embedding rates or using the same number of inactive frames. In particular, the proposed method can render accurate results for detecting the existing steganographic methods only using very small quantity of inactive frames, and thereby be adopted to detecting potential inactive-frame steganography behaviors in real-time speech streams.

REFERENCES

[1] X. Duan, K. Jia, B. Li, D. Guo, E. Zhang, and C. Qin, "Reversible image steganography scheme based on a U-Net structure," *IEEE Access*, vol. 7, pp. 9314–9323, Jan. 2019.

[2] T. Rabie and M. Baziyad, "The pixogram: Addressing high payload demands for video steganography," *IEEE Access*, vol. 7, pp. 21948–21962, Feb. 2019.

[3] H. Ghasemzadeh, "Multi-layer architecture for efficient steganalysis of UnderMp3Cover in multi-encoder scenario," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 186–195, Jan. 2019.

[4] M. T. Ahvanooey, Q. Li, J. Hou, H. D. Mazraeh, and J. Zhang, "AITSteg: An innovative text steganography technique for hidden transmission of text message via social media," *IEEE Access*, vol. 6, pp. 65981–65995, Aug. 2018.

[5] E. T. Affonso, R. D. Nunes, R. L. Rosa, G. F. Pivaro, and D. Z. Rodríguez, "Speech quality assessment in wireless VoIP communication using deep belief network," *IEEE Access*, vol. 6, pp. 77022–77032, Oct. 2018.

[6] W. Mazurczyk, "VoIP steganography and its detection: A survey," *ACM Comput. Surv.*, vol. 46, pp. 20:1-20:21, Mar. 2018.

[7] H. Tian, J. Qin, Y. Huang, Y. Chen, T. Wang, J. Liu, and Y. Cai, "Optimal matrix embedding for voice-over-IP steganography," *Signal Process.*, vol. 117, pp. 33–43, Dec. 2015.

[8] H. Tian, J. Qin, S. Guo, Y. Huang, J. Liu, T. Wang, Y. Chen, and Y. Cai, "Improved adaptive partial-matching steganography for voice over IP," *Comput. Commun.*, vol. 70, pp. 95–108, Oct. 2015.

[9] H. Tian, J. Sun, C.-C. Chang, J. Qin, and Y. Chen, "Hiding information into voice-Over-IP streams using adaptive bitrate modulation," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 749–752, Jan. 2017.

[10] Y. Jiang, S. Tang, L. Zhang, M. Xiong, and Y. J. Yip, "Covert voice over Internet protocol communications with packet loss based on fractal interpolation," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 12, pp. 54:1–54:20, Aug. 2016.

[11] Y. F. Huang, S. Tang, and J. Yuan, "Steganography in inactive frames of VoIP streams encoded by source codec," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 2, pp. 296–306, Jan. 2011.

[12] H. Tian, R. Guo, J. Lu, and Y. Chen, "Implementing covert communication over voice conversations with windows live messenger," *Adv. Inf. Sci. Service*, vol. 4, pp. 18–26, Mar. 2012.

[13] H. Tian, Y. Wu, C.-C. Chang, Y. Huang, J. Liu, T. Wang, Y. Chen, and Y. Cai, "Steganalysis of analysis-by-synthesis speech exploiting pulse-position distribution characteristics," *Secur. Commun. Netw.*, vol. 9, no. 15, pp. 2934–2944, Feb. 2016.

[14] H. Tian, Y. Wu, C.-C. Chang, Y. Huang, Y. Chen, T. Wang, Y. Cai, and J. Liu, "Steganalysis of adaptive multi-rate speech using statistical characteristics of pulse pairs," *Signal Process.*, vol. 134, pp. 9–22, Nov. 2016.

[15] Z. Lin, Y. Huang, and J. Wang, "RNN-SM: Fast steganalysis of VoIP streams using recurrent neural network," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1854–1868, Feb. 2018.

[16] W. Mazurczyk and J. Lubacz, "LACK—A VoIP steganographic method," *Telecommun. Syst.*, vol. 45, pp. 153–163, Dec. 2018.

[17] Y. Huang, J. Yuan, M. Chen, and B. Xiao, "Key distribution over the covert communication based on VoIP," *Chin. J. Electron.*, vol. 20, no. 2, pp. 357–360, Apr. 2011.

[18] Y. F. Huang, S. Tang, and Y. Zhang, "Detection of covert voice-over Internet protocol communications using sliding window-based steganalysis," *IET Commun.*, vol. 5, no. 7, pp. 929–936, Jun. 2011.

[19] P. Liu, S. Li, and H. Wang, "Steganography integrated into linear predictive coding for low bit-rate speech codec," *Multimedia Tools Appl.*, vol. 76, pp. 2837–2859, Jan. 2017.

[20] Y. Huang, C. Liu, S. Tang, and S. Bai, "Steganography integration into a low-bit rate speech codec," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1865–1875, Dec. 2012.

[21] B. Xiao, Y. Huang, and S. Tang, "An approach to information hiding in low bit-rate speech stream," in *Proc. IEEE GLOBECOM IEEE Global Telecommun. Conf.*, Nov./Dec. 2008, pp. 1–5.

[22] B. Geiser and P. Vary, "High rate data hiding in ACELP speech codecs," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar./Apr. 2008, pp. 4005–4008.

[23] H. Miao, L. Huang, Z. Chen, W. Yang, and A. Al-Hawbani, "A new scheme for covert communication via 3G encoded speech," *Comput. Elect. Eng.*, vol. 38, no. 6, pp. 1490–1501, Nov. 2012.

[24] S. Yan, G. Tang, and Y. Chen, "Incorporating data hiding into G.729 speech codec," *Multimedia Tools Appl.*, vol. 75, no. 18, pp. 11493–11512, Sep. 2016.

[25] A. Janicki, "Pitch-based steganography for speex voice codec," *Security Commun. Netw.*, vol. 9, pp. 2923–2933, Feb. 2016.

[26] S. Yan, G. Tang, Y. Sun, Z. Gao, and L. Shen, "A triple-layer steganography scheme for low bit-rate speech streams," *Multimedia Tools Appl.*, vol. 74, no. 24, pp. 11763–11782, Dec. 2015.

[27] A. Nishimura, "Data hiding in pitch delay data of the adaptive multi-rate narrow-band speech codec," in *Proc. 5th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process*, Sep. 2009, pp. 483–486.

[28] R. S. Lin, "High capacity information hiding scheme using VAD algorithm," in *Proc. IEEE Int. Conf. Consum. Electron.-Taiwan (ICCE-TW)*, May 2016, pp. 1–2.

[29] S. Li, Y. Jia, and C.-C. J. Kuo, "Steganalysis of QIM steganography in low-bit-rate speech signals," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 5, pp. 1011–1022, May 2017.

[30] J. Yang and S. Li, "Steganalysis of joint codeword quantization index modulation steganography based on codeword Bayesian network," *Neurocomputing*, vol. 313, pp. 316–323, Jun. 2018.

[31] H. Miao, L. Huang, Y. Shen, X. Lu, and Z. Chen, "Steganalysis of compressed speech based on Markov and entropy," in *Digital-Forensics and Watermarking*. Berlin, Germany: Springer, 2014, pp. 63–76.

[32] Y. Ren, T. Cai, M. Tang, and L. Wang, "AMR steganalysis based on the probability of same pulse position," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 9, pp. 1801–1811, Sep. 2015.

[33] H. Tian, J. Sun, C.-C. Chang, Y. Huang, and Y. Chen, "Detecting bitrate modulation-based covert voice-over-IP communication," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1196–1199, Jun. 2018.

[34] Y. Ren, J. Yang, J. Wang, and L. Wang, "AMR steganalysis based on second-order difference of pitch delay," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 6, pp. 1345–1357, Dec. 2016.

[35] J. Liu, H. Tian, C.-C. Chang, T. Wang, Y. Chen, and Y. Cai, "Steganalysis of inactive voice-Over-IP frames based on poker test," *Symmetry*, vol. 10, p. 336, Aug. 2018.

[36] C. Kraetzer and J. Dittmann, "Mel-cepstrum based steganalysis for VoIP steganography," in *Security, Steganography, and Watermarking of Multimedia Contents*. Bellingham, WA, USA: SPIE, 2007, pp. 650505-1–650505-12.

[37] Q. Liu, A. H. Sung, and M. Qiao, "Temporal derivative-based spectrum and Mel-Cepstrum audio steganalysis," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 3, pp. 359–368, Sep. 2009.
[38] H. Malik, K. P. Subbalakshmi, and R. Chandramouli, "Nonparametric steganalysis of QIM steganography using approximate entropy," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 418–431, May 2012.
[39] J. J. Harmsen and W. A. Pearlman, "Steganalysis of additive noise modelable information hiding," *Proc. SPIE*, vol. 5020, pp. 131–142, Feb. 2003.
[40] H. Ghasemzadeh and M. H. Kayvanrad, "Comprehensive review of audio steganalysis methods," *IET Signal Process.*, vol. 12, pp. 673–687, Mar. 2018.
[41] P. Harremoes, "Binomial and Poisson distributions as maximum entropy distributions," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 2039–2041, Jul. 2001.
[42] T. Terence, "Sumset and inverse sumset theory for Shannon entropy," *Combinatorics, Probab. Comput.*, vol. 19, no. 4, p. 37, Jul. 2010.
[43] S. Cecchi, L. Romoli, and F. Piazza, "Multichannel double-talk detector based on fundamental frequency estimation," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 94–97, Jan. 2015.
[44] L. Romoli, S. Cecchi, P. Peretti, and F. Piazza, "A mixed decorrelation approach for stereo acoustic echo cancellation based on the estimation of the fundamental frequency," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 2, pp. 690–698, Feb. 2012.
[45] L. Romoli, S. Cecchi, and F. Piazza, "A voice activity detection algorithm for multichannel acoustic echo cancellation exploiting fundamental frequency estimation," in *Proc. 9th Int. Symp. Image Signal Process. Anal. (ISPA)*, 2015, pp. 244–249.
[46] J. Lorenzo-Trueba and N. Hamada, "Noise robust voice activity detection for multiple speakers," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, 2010, pp. 1–4.
[47] B. Kedem, "Spectral analysis and discrimination by zero-crossings," in *Proc. IEEE*, vol. 74, no. 11, pp. 1477–1493, Nov. 1986.
[48] A. K. H. Al-Ali, D. Dean, B. Senadji, V. Chandran, and G. R. Naik, "Enhanced forensic speaker verification using a combination of DWT and MFCCfeature warping in the presence of noise and reverberation conditions," *IEEE Access*, vol. 5, pp. 15400–15413, Dec. 2017.
[49] S. Nakagawa, L. Wang, and S. Ohtsuka, "Speaker identification and verification by combining MFCC and phase information," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 4, pp. 1085–1095, May 2012.
[50] H. Wu, Y. Wang, and J. Huang, "Identification of electronic disguised voices," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 3, pp. 489–500, Mar. 2014.
[51] T. Zhang, Y. Wu, Y. Shao, M. Shi, Y. Geng, and G. Liu, "A pathological multi-vowels recognition algorithm based on LSP feature," *IEEE Access*, vol. 7, pp. 58866–58875, Jan. 2019.
[52] S. Chatterjee and W. B. Kleijn, "Auditory model-based design and optimization of feature vectors for automatic speech recognition," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 6, pp. 1813–1825, Sep. 2011.
[53] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, Jan. 2011.

HUI TIAN received the Ph.D. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 2010. He is currently a Full Professor and an Associate Dean of the College of Computer Science and Technology, National Huaqiao University, Xiamen, China. He is also a Guest Professor with the State Key Laboratory of Information Security of China, Chinese Academy of Sciences, China. He has published more than 90 articles in refereed proceedings of conferences, journals and books, and got nine patents. His current research interests include network and information security, cloud computing security, steganography, and digital forensics. He is a member of the Technical Committee on the Internet of CCF, a member of the Technical Committee on Information Storage of CCF, and a Senior Member of China Computer Federation (CCF).

JIE LIU received the B.Sc. degree in electronic information science and technology from the Hubei University of Technology, Wuhan, China, in 2017. He is currently pursuing the M.Sc. degree in computer science with National Huaqiao University, Xiamen, China. He is a Guest Research Assistant with the State Key Laboratory of Information Security of China, Chinese Academy of Sciences, China. His research interests are in the areas of cybersecurity with current focus on steganography and steganalysis for voice-over-IP.

CHIN-CHEN CHANG received the Ph.D. degree in computer engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1982. Prior to joining Feng Chia University, he was an Associate Professor with Chiao Tung University, a Professor with National Chung Hsing University, and the Chair Professor with National Chung Cheng University. His current title is the Chair Professor with the Department of Information Engineering and Computer Science, Feng Chia University, since February 2005. He is a Guest Professor with Hangzhou Dianzi University, China. His current research interests include computer cryptography and information security, cloud computing, data engineering, and database systems. He has over 850 publications in major journals and international conferences in these areas. He is also a Fellow of IEE, U.K. Since his early years of career development, he consecutively received the Outstanding Youth Award of Taiwan, Outstanding Talent in Information Sciences of Taiwan, the AceR Dragon Award of the Ten Most Outstanding Talents, the Outstanding Scholar Award of Taiwan, the Outstanding Engineering Professor Award of Taiwan, the Chung-Shan Academic Publication Awards, the Distinguished Research Awards of National Science Council of Taiwan, and Top Fifteen Scholars in Systems and Software Engineering of the *Journal of Systems and Software*.

YONGFENG HUANG received the Ph.D. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 2000. He is currently a Full Professor with the Department of Electronic Engineering, Tsinghua University, Beijing, China. He has published five books and over 150 research articles on computer network and multimedia communications. His current research interests include cloud computing, cloud storage, network security, and the next-generation Internet. He is a member of the Technical Committee on Internet of CCF and a Senior Member of China Computer Federation (CCF).

YIQIAO CAI received the Ph.D. degree from Sun Yat-sen University, Guangzhou, China, in 2012. He is currently an Associate Professor with the College of Computer Science and Technology, National Huaqiao University, Xiamen, China. He has published more than 50 articles in refereed conferences and journals. His research interests include machine learning, differential evolution, multi-objective optimization, and other evolutionary computation techniques.