Multimed Tools Appl
DOI 10.1007/s11042-016-3264-y

Steganalysis of MP3Stego with low embedding-rate using Markov feature

Chao Jin[1] · Rangding Wang[1] · Diqun Yan[1]

Received: 26 June 2015 / Revised: 14 November 2015 / Accepted: 8 January 2016

Abstract MP3Stego is a typical steganographic tool for MP3 audios, whose embedding behavior disturbs the intrinsic correlation of the quantized MDCT coefficients (refer to as QMDCTs). In this paper, the Markov feature captured this correlation were designed based on the QMDCTs. The feature is sensitive to the subtle alteration caused by MP3stego embedding even at a low embedding-rate. In addition, some work on QMDCT pre-processing, threshold selection and feature optimization were applied to feature construction, which contribute to improving the detection accuracy and reducing the computational complexity of the proposed scheme. Experimental results show that our approach can effectively detect MP3Stego of low embedding-rate and outperforms the prior arts.

Keywords MP3Stego · Steganography · Steganalysis · Low embedding-rate · One-step transition probability · Feature selection

1 Introduction

Steganography is the art and science of hiding secret information into common digital media [8, 18], such as audios, images and videos. Meanwhile, no one can discern the existence of the hidden information except data receivers [5]. The countermeasure to steganography is steganalysis which aims to reveal the presence of the secret data and distinguish steganographic objects from plain carrier [6].

The MPEG-1 Audio Layer 3, commonly known as MP3, has become one of the most popular formats for compressed audios in our daily life. For this very reason, MP3 would be a desirable digital carrier for steganography, and furthermore some steganographic tools of MP3 have been arisen over the last decade, such as MP3stego [7], UnderMP3Cover [11] and MP3stegz [1]. MP3stego, the most widely-used one, embeds compressed and encrypted data

✉ Rangding Wang
wangrangding@nbu.edu.cn

[1] College of Information Science and Engineering, Ningbo University, Zhejiang 315211, China

Published online: 02 February 2016

in MP3 bit stream during the encoding process by introducing slight distortion without attracting attention from the listener.

As MP3 steganography develops, MP3 steganalysis has been escalating over the last few years. To attack MP3Stego, Qiao et al. [12] presented their work on successful steganalysis of MP3Stego. The discriminative features extracted in their work were the moment statistics of the second derivatives, as well as Markov transition features and neighboring joint density of the MDCT coefficients based on each specific frequency band of MP3 audios. In [13], Qiao et al. enriched their classification model by adding GGD shape statistical features, and designed a more optimal feature set by a two-step feature selection approach incorporating ANOVA with SVM-RFE. They asserted that the average testing accuracy can achieve 84.4~87.9 % for detecting the MP3 audios with approximate modification density of 8 %, that is, embedded with the hidden message of 30 Bytes. Another steganalysis method of MP3Stego was put forward by Yan et al. [19]. They designed the standard deviation of the second-order differential sequence from quantization step as the classification feature. Although their algorithm performs better than Qiao's scheme based on the database constructed in our experiment part, the steganalysis ability were not satisfied with the low-bitrate MP3 processed by MP3Stego. Additionally, performance of the aforementioned methods could be improved when the embedding-rate is low. In order to decide whether or not an MP3 file carries contents hidden with MP3Stego of low embedding rate, Westfeld [17] proposed a steganalytic method by using the variance of block length values, which is able to detect 0.001 % of steganographic payload (about 28 Bytes) in MP3 files. Thereafter, Wan et al. [16] presented a steganalysis approach based on Huffman table distribution and recoding scheme. Their method was successful in detecting MP3Stego even at a low embedding rate (28 Bytes of hidden data). However, these two approaches were merely developed for the MP3 audio of 128 kbps, while they may not work well with other bitrate cases due to the fact that different bitrate MP3 audios do not share the identical rule on selecting coding parameter, such as the block length used by Westfeld and the Huffman table utilized by Wan et al.

In this paper, we present a detector to discriminate the cover MP3 audio (without hidden data) from the stego MP3 audio which just bears a small amount of hidden data, and furthermore, this approach is effective for the MP3 audios of 5 common bitrates. Based on the analysis of the embedding principle of MP3Stego and the observation of numerous experiments, we found that the quantized MDCT coefficients (refer to as QMDCTs) of most stego MP3 audios would be altered following a similar way. This alternation affects the correlations between adjacent QMDCTs of cover MP3 audios. Inspired by Sullivan et al. [15] and Shi et al. [14], we think this alteration can be essential to perform a steganalysis for MP3Stego. Moreover, some similar experiments [3, 9, 10, 21] conducted for compressed image steganalysis have testified that the disturbance of correlation will be enlarged by taking the differences of coefficients' magnitudes. Thus, Markov transition probabilities which characterize the correlations of the QMDCTs were considered as the steganalytic features in this work. In order to have a suitable balance between high steganalysis capability and manageable computational complexity, the one-step transition probability was chosen, which refers to the transition probabilities between two neighboring values of the QMDCT differences along the row and column directions. The procedures of QMDCTs pre-processing and feature selection were investigated during our detailed experimentation. Finally, experimental results demonstrate that the proposed scheme can effectively detecting low embedding-rate of MP3Stego and outperform the state-of-the-art algorithms.

The rest of this paper is organized as follows. Section 2 briefly reviews the embedding algorithms of MP3Stego and the alteration of cover audios caused by MP3Stego. Section 3 elaborates on feature construction and the complete process of the proposed steganalysis algorithm. In Section 4, we present the experimental results and performance comparisons. Finally, conclusion is drawn in Section 5.

2 Review of MP3Stego steganography

2.1 Embedding algorithm of MP3Stego

With the original audio of WAV format as input, MP3Stego embeds the data compressed and encrypted by using 3DES during the compression process of MPEG Lay-3 [7].

As the first primary part of Layer III encoding, the filter bank transforms the PCM audio signals into the frequency samples which would be quantized and coded within two nested loops. The inner loop quantizes the input data and increases the quantizer step size until the output data can be coded with the available amount of bits. The outer loop checks the distortion of each scalefactor band and, if the allowed distortion is exceeded, amplifies the scalefactor band and calls the inner loop again [4]. These two loops are an iterative process that strives for optimal trade-off between the compression ratio and the audio quality.

Actually, the hiding process of MP3Stego takes place in the inner loop. Without embedding, the inner loop would be finished when the quantizer step size is increased to fit the available number of bits. But for MP3Stego, the quantizer step size variable of the embeddable granules would be increased continually until the embedded bit equals the Least Significant Bit (LSB) of part2_3_length variable which contains the number of main_data bits used for scalefactors and Huffman code data in the MP3 bit-stream [7]. Simultaneously, only the granules randomly chosen by a pseudo-random bit generator based on SHA-1 would be modified.

2.2 Impact of embedding

As mentioned in section 2.1, MP3Stego embeds data by changing the termination condition of the inner loop during the encoding process [13], which makes the quantizer step size of embeddable granules bigger than that of non-embeddable granules. By increasing the quantizer step size *sz*, the absolute values of QMDCTs decrease according to the quantization formula as follows:

$$ix(i) = nint\left(\left(|xr(i)| \Big/ \sqrt[4]{2}^{sz}\right)^{0.75} - 0.0946\right) \tag{1}$$

where $xr(i)$ is defined as the vector of the magnitudes of the spectral values, namely MDCT coefficients. And the vector of quantized values is denoted by $ix(i)$ that is QMDCTs. The nearest integer operator *nint*() returns the nearest integer value to the real-valued argument.

Based on the above analysis, it can be noticed that QMDCTs of the stego audios generated by MP3Stego must be different from the cover audios' QMDCTs. The embedding operation leads to the change of a big amount of QMDCTs. We think it is more effective and reliable that the features constructed based on the numerous changed data instead of single or several coding parameters for MP3Stego. Therefore, the steganalytic features of our approach were extracted from the QMDCTs.

3 The steganalysis approach

3.1 QMDCT pre-processing

MP3Stego embeds hidden bits in a consecutive manner, and starts at the first frame. So the first 50 frames of MP3 audio were considered in this analysis. Note that for the stereo MP3, two granules constitute a frame and each granule consists of two channels; and for the mono MP3, one granule just has a single channel. In our experiments, the stereo MP3s were used, and their first 50 frames contain 200 channels and each channel has 576 frequency sub-bands. Moreover, 200 frequency sub-bands were chosen because the coefficient values of the rest sub-bands almost equal zeroes. The above pre-processing of QMDCTs would contribute to promoting the effectiveness of the features and reducing the computational complexity.

In order to facilitate the description, the QMDCTs of the front 50 frames and the front 200 frequency sub-bands extracted during the decoding process of MP3audios can be denoted as:

$$S_{QMDCT} = \begin{pmatrix} Q_{11} & \cdots & Q_{1j} \\ \vdots & \ddots & \vdots \\ Q_{i1} & \cdots & Q_{ij} \end{pmatrix} \tag{2}$$

where $i \in \{1,\ 2,\ \ldots,\ 200\}$ is defined as the number of channels (the row number of the matrix S_{QMDCT}), the variable $j \in \{1,\ 2,\ \ldots,\ 200\}$ represents the number of the selected frequency sub-bands (the column number of the matrix S_{QMDCT}).

According to Eq. (2), the differences of the QMDCTs and the differences of the absolute values of QMDCTs can be interpreted as D_{QMDCT} and D_{QMDCT_ABS} respectively:

$$D_{QMDCT} = \begin{pmatrix} QD_{11} & \cdots & QD_{1n} \\ \vdots & \ddots & \vdots \\ QD_{m1} & \cdots & QD_{mn} \end{pmatrix} \tag{3}$$

$$D_{QMDCT\_ABS} = \begin{pmatrix} AQD_{11} & \cdots & AQD_{1n} \\ \vdots & \ddots & \vdots \\ AQD_{m1} & \cdots & AQD_{mn} \end{pmatrix} \tag{4}$$

where $m \in \{1,\ 2,\ \ldots,\ 199\}$, $n \in \{1,\ 2,\ \ldots,\ 200\}$, and the elements of the matrices D_{QMDCT} and D_{QMDCT_ABS} are defined as follows:

$$QD_{m,n} = Q_{m+1,n} - Q_{m,n} \tag{5}$$

$$AQD_{m,n} = \left|Q_{m+1,n}\right| - \left|Q_{m,n}\right| \tag{6}$$

Based on Eqs. (2)–(4), the one-step transition probabilities of the matrices with two directions are calculated according to the following formulae.

Row direction (inter-channels):

$$P_{Inter} = \sum_{i=1}^{199}\sum_{j=1}^{200}\delta\left(Q_{i,j}=x, Q_{i+1,j}=y\right) \Big/ \sum_{i=1}^{199}\sum_{j=1}^{200}\delta\left(Q_{i,j}=x\right) \tag{7}$$

$$P_{D\_Inter} = \sum_{m=1}^{198}\sum_{n=1}^{200}\delta\left(QD_{m,n}=x, QD_{m+1,n}=y\right) \Big/ \sum_{m=1}^{198}\sum_{n=1}^{200}\delta\left(QD_{m,n}=x\right) \tag{8}$$

$$P_{AD\_Inter} = \sum_{m=1}^{198}\sum_{n=1}^{200}\delta\left(AQD_{m,n}=x, AQD_{m+1,n}=y\right) \Big/ \sum_{m=1}^{198}\sum_{n=1}^{200}\delta\left(AQD_{m,n}=x\right) \tag{9}$$

Column direction (intra-channel):

$$P_{Intra} = \sum_{i=1}^{200}\sum_{j=1}^{199}\delta\left(Q_{i,j}=x, Q_{i,j+1}=y\right) \Big/ \sum_{i=1}^{200}\sum_{j=1}^{199}\delta\left(Q_{i,j}=x\right) \tag{10}$$

$$P_{D\_Intra} = \sum_{m=1}^{199}\sum_{n=1}^{199}\delta\left(QD_{m,n}=x, QD_{m,n+1}=y\right) \Big/ \sum_{m=1}^{199}\sum_{n=1}^{199}\delta\left(QD_{m,n}=x\right) \tag{11}$$

$$P_{AD\_Intra} = \sum_{m=1}^{199}\sum_{n=1}^{199}\delta\left(AQD_{m,n}=x, AQD_{m,n+1}=y\right) \Big/ \sum_{m=1}^{199}\sum_{n=1}^{199}\delta\left(AQD_{m,n}=x\right) \tag{12}$$

where $\delta(X=x, Y=y) = \begin{cases} 1, & X=x, Y=y \\ 0, & otherwise \end{cases}$ and $x, y \in [-T \quad T]$. T is a threshold for controlling the model dimensionality of the proposed approach and the related questions of how to select a proper T value would be discussed in section 3.2.

To facilitate examination of the difference of one-step transition probabilities between cover and stego audios, we subtracted the mean P_{Inter} value of 1000 cover audios from that of 1000 stego audio yielding to $x, y \in [-15 \quad 15]$, and converted the P_{Inter} difference into absolute values. Likewise, the absolute P_{Intra}, P_{D_Inter}, P_{D_Intra}, P_{AD_Inter}, and P_{AD_Intra} differences were acquired in the similar way. Taking the absolute P difference of 128 kbps audios for example, Fig. 1a sketches the absolute P_{Inter} difference which consists of $31 \times 31 = 961$ dimensions non-negative values, and Fig. 1d shows the absolute P_{Intra} difference. As can be inferred in this pair of sub-figures, the P_{Inter} of cover audios is distinct from that of stego audios while their P_{Intra} values are almost the same. This phenomenon is caused by the embedding rule of MP3Stego that modified all of the coefficients in one frame at the same time, which also means that the intra-channel distribution is preserved; On the other hand, the inter-channels pattern has been altered across the adjacent channels [12]. When the difference values of QMDCTs along the

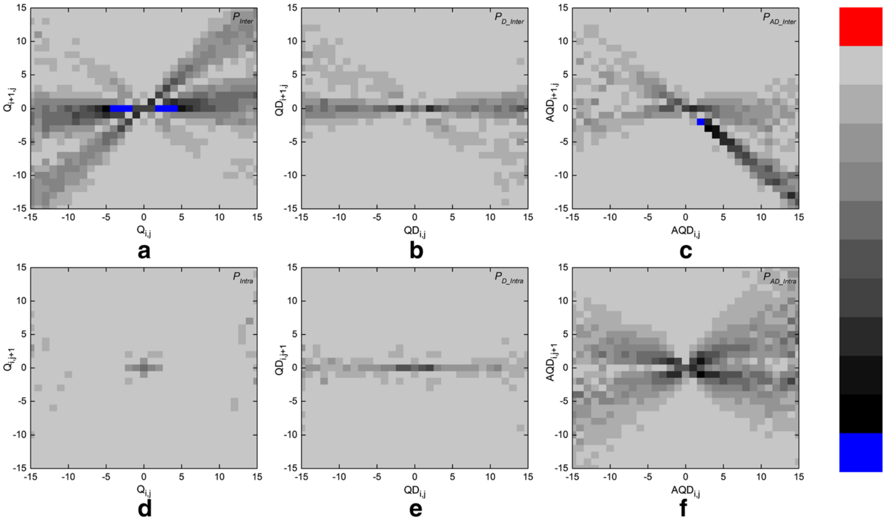


Fig. 1 The absolute P difference values for the matrices S_{QMDCT}, D_{QMDCT} and D_{QMDCT_ABS}. (**a**) Absolute P_{Inter} difference values (**b**) Absolute P_{D_Inter} difference values (**c**) Absolute P_{AD_Inter} difference values (**d**) Absolute P_{Intra} difference values (**e**) Absolute P_{D_Intra} difference values (f) Absolute P_{AD_Intra} difference values

channels were calculated, the P_{D_Inter} and P_{D_Intra} would alter with the similar trend, which are shown in Fig. 1b and e. This indicates the variation between neighboring coefficients in the same channel would be introduced by the subtraction of the corresponding coefficients of adjacent channels [22]. From the observation of Fig. 1c and f, we find that the P_{AD_Inter} and P_{AD_Intra} are both changed obviously after embedding by MP3Stego with low payload. It illustrates that the disturbance caused by the embedding behavior was indeed enlarged by taking the differences of the absolute values of QMDCTs. Based on the above analysis, the one-step transition probabilities of the difference values of the absolute values of QMDCTs with row and column directions were designed as the features for constructing the classification model.

3.2 T Selection

According to Fig. 2a, it can be seen that the distributions of the original QMDCTs of the MP3 audios with different bitrates are not the same, but they all can be approximately modeled as the Laplace distribution [20]. The values of most QMDCTs fall into an interval $[-T\ T]$ where T can be a small integer value and the number of the elements falling into a fixed interval becomes less and less with the bitrate increasing. This phenomenon is consistent with the principle of MP3 quantization, when the bitrate is set bigger for the encoding process, the MDCT coefficients will be quantized finer and more zero values of the quantized data will turn into the nonzero values. So the QMDCT distribution of the MP3 with lower bitrate is tighter, and the distributions do not change much for the difference values of the absolute values of QMDCTs, which are shown in Fig. 2b.

The percentage of the elements in matrix of cover audio and stego audio falling into $[-T\ T]$ for $T = \{4, 5, 6, 7, 8, 9, 10, 15, 20\}$ are shown in Fig. 3a and b, respectively. The probability

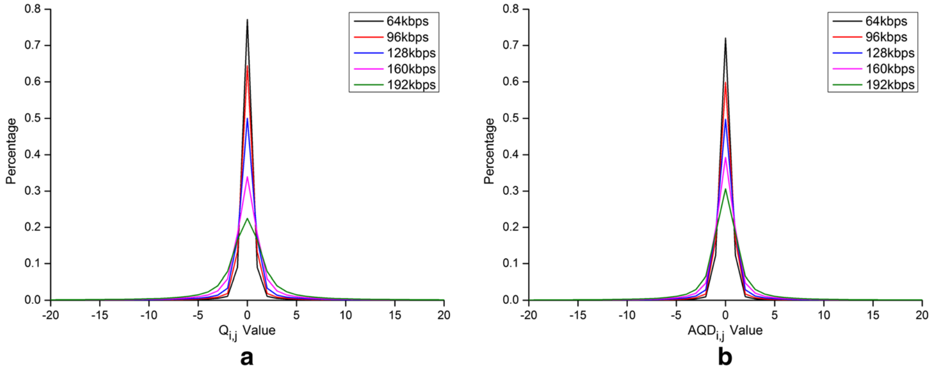


Fig. 2 Distributions of the elements in matrices S_{QMDCT} and D_{QMDCT_ABS} of cover audio. (**a**) Average of the histograms for S_{QMDCT} (**b**) Average of the histograms for D_{QMDCT_ABS}

values were estimated from 1000 audio pairs from the database in section 4.1. It can be seen that the percentages of the fixed intervals are almost concentrated in the range from 0.9 to 1.0; and when T was set bigger that 5, even the smallest vertical value amid all meaningful points exceeds 0.9. Taken the higher detection rates and lower computational complexity into account, the threshold T is set as 6 in the experiments.

3.3 Feature optimization

In the light of the analysis previous sections, the P_{AD_Inter} and P_{AD_Intra} were chosen as the feature set for our scheme. However, as shown in Fig. 4, both the P_{AD_Inter} and P_{AD_Intra} consist of 169 dimension features and most of them do not contribute to enhancing the classification accuracy, instead they would increase the algorithm computational complexity. So we endeavor to eliminate the useless features and keep the most effective ones to balance the high detection accuracy and low feature dimensionality.

Based on the investigation of amounts of experiments, and for the classification models of 5 different bitrates, Eqs. (13) and (14) were applied to pick out the optimal features with more discriminative information out of the P_{AD_Inter} and P_{AD_Intra}, respectively.

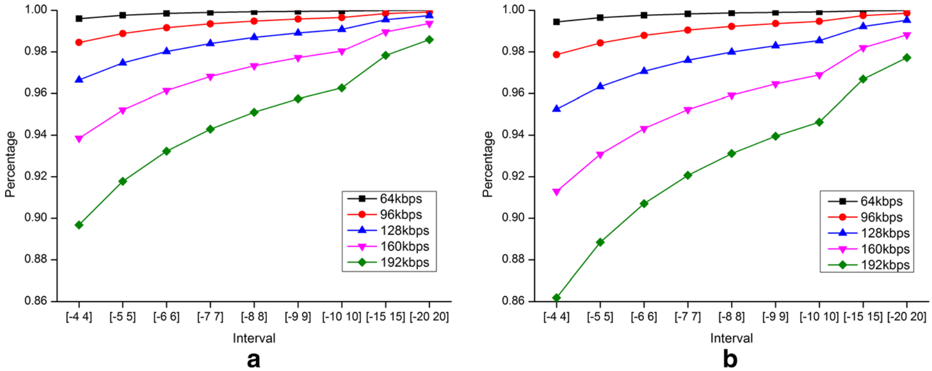
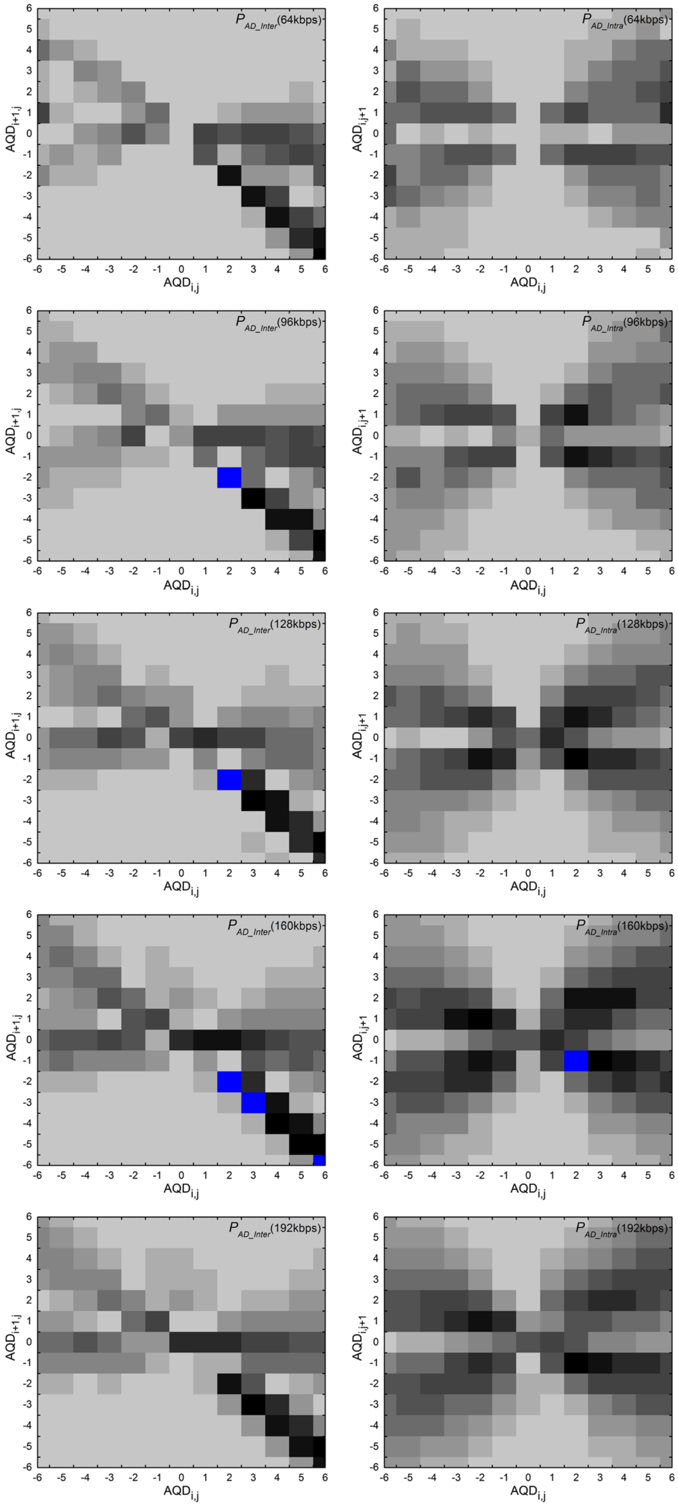


Fig. 3 The percentage of the elements of D_{QMDCT_ABS} falling within $[-T\ T]$ of cover and stego audios. (**a**) Average of the percentages for cover audio (**b**) Average of the percentages for stego audios


P_{AD_Inter}(64kbps)
P_{AD_Intra}(64kbps)
P_{AD_Inter}(96kbps)
P_{AD_Intra}(96kbps)
P_{AD_Inter}(128kbps)
P_{AD_Intra}(128kbps)
P_{AD_Inter}(160kbps)
P_{AD_Intra}(160kbps)
P_{AD_Inter}(192kbps)
P_{AD_Intra}(192kbps)
AQD_{i+1,j}
AQD_{i,j+1}
AQD_{i,j}
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6

◀ **Fig. 4** The absolute P_{AD_Inter} and P_{AD_Intra} differences for 5 different bitrates. The 1th row, Absolute P difference values for 64 kbps. The 2nd row, Absolute P difference values for 96 kbps. The 3rd row, Absolute P difference values for 128 kbps. The 4th row, Absolute P difference values for 160 kbps. The 5th row, Absolute P difference values for 192 kbps

The optimal features of row direction (P_{AD_Inter}):

$$\begin{cases} \left|\mathrm{MEAN}\left(\sum_{i=1}^{N} C_i(P_{AD_{Inter}}(x,y))\right) - \mathrm{MEAN}\left(\sum_{i=1}^{N} S_i(P_{AD_{Inter}}(x,y))\right)\right| > \varepsilon_{inter\_mean} \\ \mathrm{STD}\left(\sum_{i=1}^{N} S_i(P_{AD_{Inter}}(x,y))\right) < \varepsilon_{inter\_std} \end{cases} \tag{13}$$

The optimal features of column direction (P_{AD_Intra}):

$$\begin{cases} \left|\mathrm{MEAN}\left(\sum_{i=1}^{N} C_i(P_{AD_{Intra}}(x,y))\right) - \mathrm{MEAN}\left(\sum_{i=1}^{N} S_i(P_{AD_{Intra}}(x,y))\right)\right| > \varepsilon_{intra\_mean} \\ \mathrm{STD}\left(\sum_{i=1}^{N} S_i(P_{AD_{Intra}}(x,y))\right) < \varepsilon_{intra\_std} \end{cases} \tag{14}$$

where $x, y \in [-6, 6]$, $C_i(\cdot)$ and $S_i(\cdot)$ are defined as the ith cover and stego audios respectively. Herein N is equivalent to 1000, and the threshold values ε_{inter_mean}, ε_{inter_std}, ε_{intra_mean} and ε_{intra_std} are obtained as empirical values and are given in Table 2. Then 5 feature subsets were obtained. In order to facilitate implementation and get more general conclusions, the features whose indexes were occurred in all 5 subsets were selected as the final features for our algorithm, as deduced from Eq. (15):

$$FS = FS_{64} \cap FS_{96} \cap FS_{128} \cap FS_{160} \cap FS_{192} \tag{15}$$

where FS_m means the set of optimal feature indexes of m kbps audio.

In summary, the optimal features of P_{AD_Inter} and P_{AD_Intra} selected by Eqs. (13)–(15) were successful to reduce the computational complexity on the premise of controlling the classifying ability decreased in a reasonable range compared to the intact features.

3.4 Overview of the proposed approach

The proposed approach can be outlined as the following steps, and the entire procedure is depicted in Fig. 5.

1) Extract the matrix S_{QMDCT} of cover and stego audio pairs according to Eq. (2) and the difference matrix of S_{QMDCT} is obtained as D_{QMDCT_ABS}.

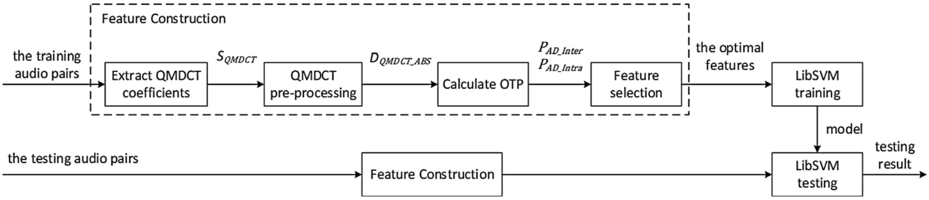


Fig. 5 Diagram of the proposed approach

2) Calculate the one-step transition probabilities (OTP) P_{AD_Inter} and P_{AD_Intra} between two immediately neighboring difference coefficients based on the formulae depicted in Eqs. (9) and (12), respectively.
3) Choose the optimal features according to Eqs. (13)–(15).
4) Take the support vector machine LibSVM [2] as the classifier, 50 % of the feature vectors of cover and stego audio pairs are selected randomly for training and the remaining 50 % feature vectors for testing.

4 Experiments and results

4.1 Experimental setup

In order to effectively evaluate the performance of our proposed steganalysis, a dataset, which consists of 1000 stereo WAV audios with a sample rate of 44.1 kHz and several musical genres of classical, pop, folk, rock, hip-hop and so on, was constructed. Each audio has duration of 10 s, and the file size is 1.68 MB. The whole dataset contains two categories, one is composed of the common music played by human voices and instruments together, and another is only instrumental music. Besides, the common music consists of several language songs, such as Chinese, English, Korean, Spain, and so on. The instrumental music is played by manifold instruments, such as keyboard, string, wind, and percussion, of course include symphonies as well.

In our experiment, 5000 cover MP3 audios and 5000 stego MP3 audios were both obtained based on the audio set mentioned above with five different bitrates of 64, 96, 128, 160 and 192 kbps, as shown in Table 1. In addition, the embedded data carried by any one of the stego audios was randomly selected from the text, audio or image file which has the same size (10 Bytes). It is observed that the maximum embedding-rate for MP3Stego is 2 bits per frame for mono MP3 audios and 4 bits per frame for stereo MP3 audios. So 10 Bytes, namely 80 Bits, about 10 % embedding-rate for the stereo MP3 of 376 frames which were used in the experiments.

Table 1 The audio set of our experiments

	64 kbps	96 kbps	128 kbps	160 kbps	192 kbps
Num. of cover audios	1000	1000	1000	1000	1000
Num. of stego audios	1000	1000	1000	1000	1000

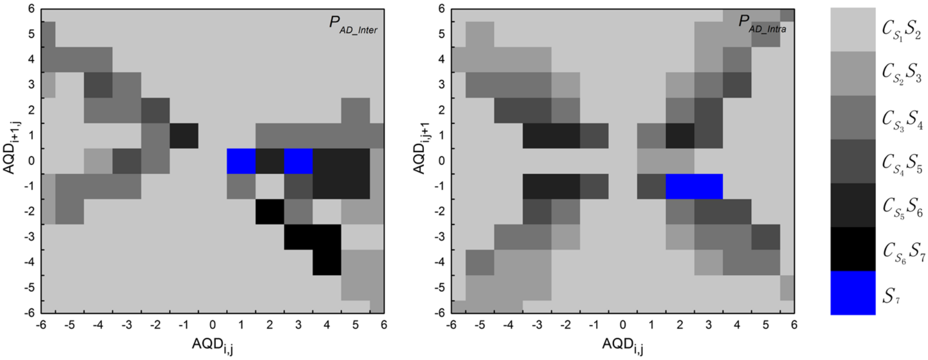


Fig. 6 Distributions of different feature sets

4.2 Comparison of the optimal features and the raw features

Combining with the support vector machine LibSVM, the classification models of MP3 audios for 5 different bitrates were constructed by using the function *svmtrain*(·) whose parameters were set as defaults. Then 50 % of the cover and stego audio pairs were randomly selected from the aforementioned dataset for training and the remaining 50 % pairs for testing.

Accuracy (refer to as *ACC*) is the typical indicator of steganalysis capability, which is defined as:

$$ACC = {}^{(TP+TN)}\Big/_{(P+N)} \tag{16}$$

where *TP* is the number that the stego audio has been classified as stego, and *TN* means the cover audio was sorted as cover; *P* and *N* represent the amount of the actual cover and stego audios, respectively. Each *ACC* value was averaged over 50 times in our experiments.

Figure 6 shows the specific positions of the raw features S_1, as well as the different feature sets S_2–S_7 optimized pursuant to Eqs. (13)–(15) with different ε values which are given in Table 2. For example, the 4 features of S_7 colored in blue are located in P_{AD_Inter} with the coordinates of (1, 0), (3, 0) and P_{AD_Intra} with the coordinates of (2, −1), (3, −1). By the way, $C_{S_6}S_7$ stands for the complementary set of S_7, that is, $C_{s_6}S_7 + S_7 = S_6$. Compared with Fig. 4,

Table 2 The ε values used for selecting different feature sets

$P_{AD\_Inter}+P_{AD\_Intra}$	ε_{inter_mean}	ε_{inter_std}	ε_{intra_mean}	ε_{intra_std}
$S_1(169+169)$	–	–	–	–
$S_2(50+64)$	0.005	0.15	0.005	0.05
$S_3(37+38)$	0.005	0.08	0.01	0.05
$S_4(16+19)$	0.015	0.08	0.02	0.05
$S_5(12+7)$	0.02	0.08	0.025	0.05
$S_6(6+2)$	0.03	0.08	0.03	0.05
$S_7(2+2)$	0.03	0.05	0.03	0.05

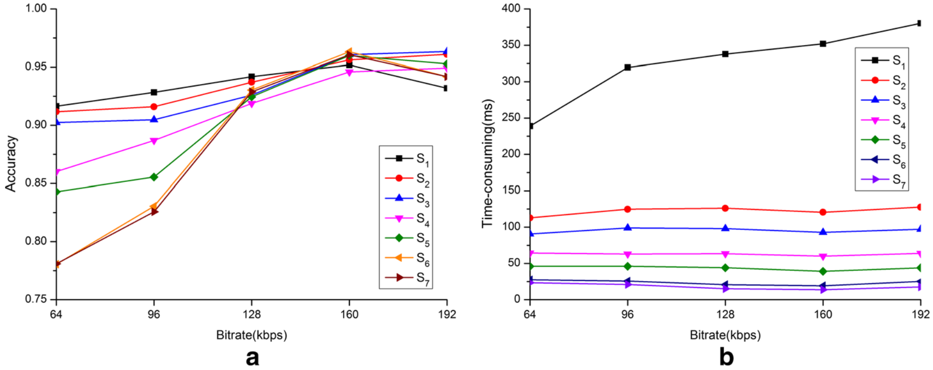


Fig. 7 Performance comparison of different feature sets. (**a**) Detection accuracy comparison for the 7 feature sets (**b**) Time-consuming comparison for the 7 feature sets

these distributions are approximately coincided with each other, which indicates that the feature with better classification ability were correctly selected via the feature optimization.

The detection accuracy of the raw features compared with other 6 optimal feature sets are shown in Fig. 7a. It is observed that there is a significant drop in accuracy as the feature dimension decreasing under low bitrate (64 and 96 kbps) conditions, whereas detection accuracies for the other 3 bitrate cases keep stable. However, the performance of S_1 is the worst one among the 7 feature sets in terms of 192 kbps audios. This is due to the fact that the features with good discriminative ability for 64 and 96 kbps audios were eliminated by the operation of calculating feature intersection for the 5 bitrate cases. On the other hand, the features which are harmful to improving the detection accuracy were discarded for 128 and 160 kbps audios. As expected, the computational complexity are increasing by adding features during the training and testing process, as shown in Fig. 7b.

4.3 Comparison of the proposed algorithm and the state-of-the-art steganalysis methods

For the sake of assessing the proposed method comprehensively, the performance of our approach was compared with the state-of-the-art algorithms proposed in [13, 16, 17, 19] with the same experimental setup in section 4.2. On account of seeking a trade-off between detection accuracy and computational complexity, the feature set S_5 was chosen as the feature vector for our approach.

Table 3 Comparison of the detection accuracy of our algorithm and the state-of-the-art algorithms in paper [13, 16, 17, 19]

	Ours	Qiao et al.'s	Yan et al.'s	Westfeld's	Wan et al.'s
64 kbps	84.3 %	75.2 %	59.8 %	–	–
96 kbps	85.6 %	78.7 %	61.8 %	–	–
128 kbps	92.5 %	83.5 %	78.2 %	93.4 %	79.00 %
160 kbps	96.0 %	82.2 %	92.8 %	–	–
192 kbps	95.3 %	85.6 %	95.1 %	–	–

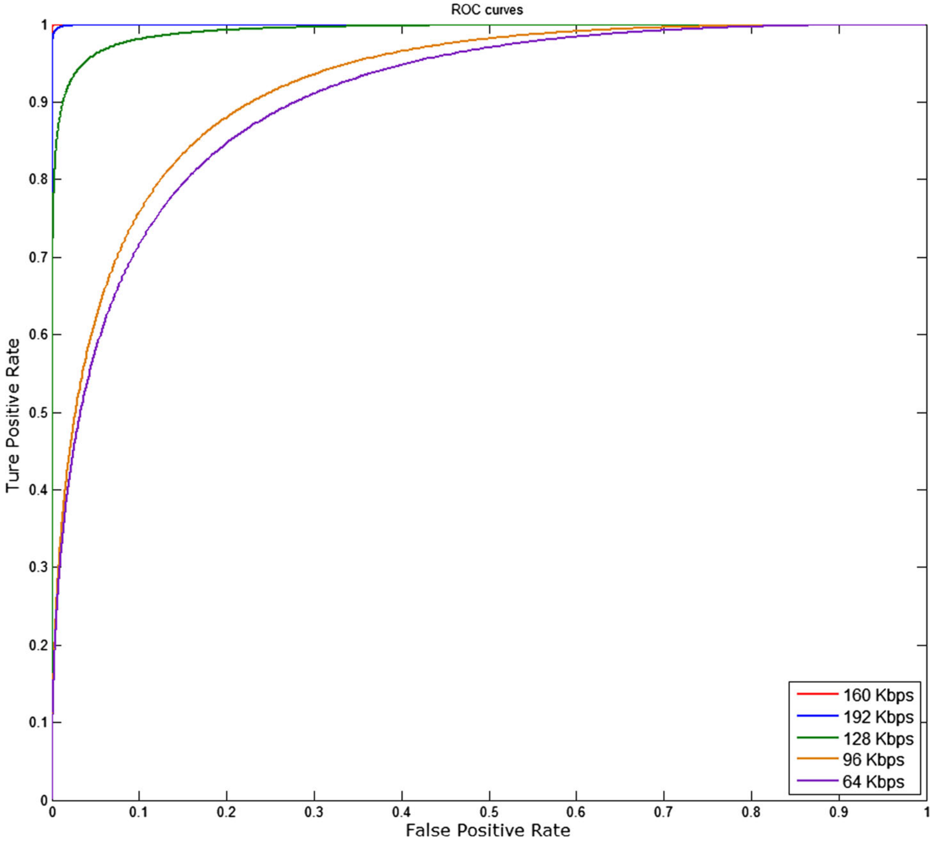


Fig. 8 ROC curves of the proposed scheme

From the observation of Table 3, it can be seen that Qiao et al. and Yan et al. methods can detect MP3Stego steganography in MP3 audios with various bitrates, but the detection accuracy could be improved, especially at the low embedding-rate. On the other hand, Westfeld and Wan et al. schemes are merely appropriate for the default condition of MP3Stego compression, codifying the audio with 128 kbp, while they are invalid for the MP3 audios of other bitrates. Finally, the Receiver Operating Characteristic (ROC) curves of our method for the 5 bitrates are given in Fig. 8. ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The more each curve hugs the left and top edges of the plot, the better the classification accuracy. So the curves in Fig. 8 reveal that the proposed approach is able to detect the MP3Stego steganography with low embedding-rate reliably.

5 Conclusions

In this paper, we presented a scheme for detecting low embedding-rate in the MP3 audio which was processed by MP3Stego. The proposed scheme can effectively detect a small amount of hidden data (about 10 Bytes) in MP3 audios. Based on investigating the embedding principle of MP3Stego and observing the alteration of QMDCTs between the cover and stego audios, one-step transition probabilities of the difference of the absolute values of QMDCTs were utilized for constructing the classification features. Furthermore, the pre-processing of QMDCTs and feature selection were proved to be highly beneficial to reducing the computational complexity of our method. Finally, the experimental results shown that the extracted features have good performance for discriminating between the MP3 audios with and without hidden data.

In our future work, we intend to study whether the feature set proposed in this paper suitable for other typical MP3 steganographic tools. We think that the embedding behavior of MP3 steganographic tools would alter the MDCT coefficients of the audios without embedding, and furthermore affect the distribution and correlation of coefficients.

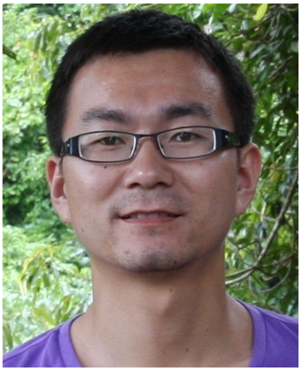
References

1. Achmad Z (2008) MP3Stegz. http://sourceforge.net/projects/mp3stegz. Accessed 25 June 2015
2. Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:27:1–27:27, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
3. Chen CC, Shi YQ (2008) JPEG image steganalysis utilizing both intrablock and interblock correlations. In: Proceedings of the IEEE International Symposium on Circuits and Systems. Seattle, WA, 18–21 May, pp 3029–3032
4. ISO/IEC 11172–3:1993 Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 3: Audio
5. Morkel T, Eloff JHP, Olivier MS (2005) An overview of image steganography. In: Proceedings of the 5th Annual Information Security South Africa Conference on New Knowledge Today. Sandton, South Africa, 29 June–1 July
6. Nissar A, Mir AH (2010) Classification of steganalysis techniques: a study. Digit Signal Process 20:1758–1770
7. Petitcolas FAP (2002) MP3Stego. http://www.petitcolas.net/steganography/mp3stego. Accessed 25 June 2015
8. Petitcolas FAP, Anderson RJ, Kuhn MG (1999) Information hiding—a survey. Proc IEEE 87(7):1062–1078
9. Pevny T, Bas P, Fridrich J (2010) Steganalysis by subtractive pixel adjacency matrix. IEEE Trans Inf Forensics Secur 5(2):215–224
10. Pevny T, Fridrich J (2007) Merging markov and DCT features for multi-class JPEG steganalysis In: Proceedings of SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX. San Jose, CA, USA, 28 January. doi:10.1117/12.696774
11. Platt C (2004) UnderMP3Cover. http://sourceforge.net/projects/ump3c. Accessed 25 June 2015
12. Qiao MY, Sung AH, Liu QZ (2009) Steganalysis of MP3Stego. In: Proceedings of the International Joint Conference on Neural Networks. Atlanta, Georgia, USA, 14–19 June, pp 2566–2571
13. Qiao MY, Sung AH, Liu QZ (2013) MP3 audio steganalysis. Inf Sci 231:123–134
14. Shi YQ, Chen CC, Chen W (2006) A markov process based approach to effective attacking JPEG Steganography. In: Proceedings of the 8th International Workshop on Information Hiding. Alexandria, VA, USA, 10–12 July, 4437:249–264
15. Sullivan K, Madhow U, Chandrasekaran S, Manjunath BS (2006) Steganalysis for Markov cover data with applications to images. IEEE Trans Inf Forensics Secur 1(2):275–287
16. Wan W, Zhao XF, Huang W, Sheng RN (2012) Steganalysis of MP3Stego based on Huffman table distribution and recoding. J Univ Chin Acad Sci 21(1):118–124
17. Westfeld A (2002) Detecting low embedding rates. In: Proceedings of the 5th International Workshop on Information Hiding. Noordwijkerhout, The Netherlands, 7–9 October, 2578:324–339
18. Xu DW, Wang RD, Shi YQ (2014) Data hiding in encrypted H.264/AVC video streams by codeword substitution. IEEE Trans Inf Forensics Secur 9(4):596–606

19. Yan DQ, Wang RD, Yu XM, Zhu J (2013) Steganalysis for MP3Stego using differential statistics of quantization step. Digit Signal Process 23:1181–1185
20. Yu RS, Lin X, Rahardja S, Ko CC (2004) A statistics study on the mdct coefficient distribution for audio. In: Proceedings of the IEEE International Conference on Multimedia and Expo. Taipei, Taiwan, 27–30 June, 2: 1483–1486
21. Zhang H, Ping XJ, Xu MK, Wang R (2014) Steganalysis by subtractive pixel adjacency matrix and dimensionality reduction. Sci Chin Inf Sci 57:048101:1–048101:7
22. Zou DK, Shi YQ, Su W, Xuan GR (2006) Steganlysis based on markov model of thresholded prediction-error image. In: Proceedings of the IEEE International Conference on Multimedia and Expo. Toronto, ON, Canada, 9–12 July, pp 1365–1368

Chao Jin is currently a Ph.D. candidate in College of Information Science and Engineering, Ningbo University, China. His research interests mainly include audio steganalysis, multimedia forensics and audio/speech signal processing.

Rangding Wang received his M.S. degree in the Department of Computer Science and Engineering from the Northwest Polytechnic University, Xian in 1987, and received his Ph.D. degree in the School of Electronic and Information Engineering from Tongji University, Shanghai, China, in 2004. He is now a professor in College of Information Science and Engineering, Ningbo University, China. His research interests mainly include multimedia security, digital watermarking for digital rights management, and data hiding.



Diqun Yan received B.S., M.S., Ph.D. degrees in Circuit and System from Ningbo University, Zhejiang China, in 2002, 2008 and 2012 respectively. He is currently an associate professor in College of Information Science and Engineering, Ningbo University. His research interests include multimedia forensics and security.