A Steganalysis Scheme for AAC Audio Based on MDCT Difference Between Intra and Inter Frame

Yanzhen Ren[1,2](✉), Qiaochu Xiong[2], and Lina Wang[1,2]

[1] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, Wuhan, China
[2] School of Computer Science, Wuhan University, Wuhan, China
renyz@whu.edu.cn, xiongqiaochu_whu@163.com, lnawang@163.com

Abstract. AAC (Advanced Audio Coding), as an efficient audio codec, has been used widely in mobile internet applications. Steganographies based on AAC are emerging and bringing new challenges to information content security. In this paper, an AAC steganalysis scheme to detect the steganographies which embedded secret information by modifying MDCT coefficient is proposed. The modification of MDCT coefficient will cause the statistical characteristic of the difference between inter-frame and intra-frame changed simultaneously. Based on this ideal, we proposed a scheme to extract combination features to classify cover and stego audio. There are 16 groups of sub-features to represent the correlation characteristics between the multi-order differential coefficients of Intra and Inter frame (MDI2), each sub-feature's performance are analyzed in this paper, and an ensemble classifier is used to realize the steganalyzer. Experiment results show that the detection accuracy of the proposed scheme are above 85.34% when the relative embed rate is over 50%, this performance is obviously better than the literatures methods. Due to the similarity of the coding principle of AAC and MP3, the proposed features can be applied into MP3 steganalysis.

Keywords: MDCT · AAC · MP3 · Steganography · Steganalysis

1 Introduction

As an audio coding standard for lossy digital audio compression, AAC has been used widely in mobile internet applications. AAC is designed to be the successor of the MP3 format, and generally achieves better sound quality than MP3 at similar bit rates [1]. AAC is adopted as the default audio format for many popular applications and smart phones, such as YouTube, iPhone, Play Station, Android and BlackBerry etc., and it is also supported by manufacturers of in-dash car audio systems.

Y. Ren—This work is supported by the Natural Science Foundation of China (NSFC) under the grant No. U1536114, No. U1536204, and China Scholarship Council.

C. Kraetzer et al. (Eds.): IWDW 2017, LNCS 10431, pp. 217–231, 2017.
DOI: 10.1007/978-3-319-64185-0_17

With the widely use of AAC, it has became an ideal carrier to hide secret information [2–7]. There are three main embedding domain for AAC compression parameters: MDCT coefficients [2–4], Huffman coding parameter [5,6] and quantization parameter [7]. MDCT coefficients is the mainly parameter of AAC, occupies almost 70% of the AAC coding bit stream. Slightly modifications of MDCT coefficients will not cause the declining of hearing quality. Therefore, the steganography schemes by modifying the MDCT coefficient generally have good concealment and large embedding capability. Wang [2] proposed to divide the MDCT band into different regions: Big Data Region, Small Data Region and Zero Data Region, embeds secret message into Small Data Region. Zhu [3] proposed to embed secret information by using the characteristics of the sign bits of Huffman codeword according to the structure of AAC bitstream. The coefficients which absolute value is small were changed. Zhu [5] divides the Huffman codeword into two similar groups, and embed the secret message according to the corresponding codeword groups. All schemes will cause the change of the MDCT coefficients, and have high capacity and good imperceptibility.

Up to now, there are many steganalysis methods [8–13] against MP3 steganography, but there is a fat lot of steganalysis methods for AAC. Due to the similarity between the coding principle of AAC and MP3, the methods of those MP3 steganalysis schemes can be applied into AAC steganalysis. Most of the steganalysis schemes for MP3 are designed to detect MP3Stego [14], which is an open tool to embed secret message into MP3 bitstream by controlling parity of the block length. The processing mode of MP3Stego will change the value of the MDCT coefficients. There are two types of mainstream steganalysis schemes for MP3: special scheme and general scheme. In special scheme, the features are designed based on direct change of MP3Stego. Yan [8] find that MP3Stego will change of the quantization step, so they proposed to use standard deviation of its second-order sequence as the steganalysis feature. Yan [9] found that the number of bits in the bit reservoir will be changed, they proposed to extract the statistic characters of bit reservoir and use calibration to build the steganalysis features. Yu [10] proposed to extract the statistic features of the big values in side information, used recompression to improve the features sensitivity. Those special schemes do well in MP3Stego, but failed to detect other steganography scheme of MDCT coefficients modification. In general schemes, the difference between MDCT coefficients are analyzed. Qiao [11] proposed to extract the second-order differential QMDCT (quantized MDCT) coefficients of the same sub-band, build four steganalysis feature sets: frequency-based sub-band moment statistical features, accumulative Markov transition features, accumulative neighboring joint density features and the shape parameters of generalized Gaussian; Kuriakose [12] fused Markov feature and accumulative neighboring joint density of the second-order differential QMDCT coefficients from inter-frame, and improved the performance. Based on the first-order differential QMDCT coefficients of the same sub-band, Jin [13] proposed to extract Markov features from its horizontal and vertical direction. The general schemes have good versatility.

However, The main goal of current MP3 steganalysis schemes is to detect MP3Stego. MP3Stego will change the parameters of quantization process, it will disturb the correlations of inter-frame MDCT coefficients obviously. Therefore, the current MP3 steganalysis schemes extract the features based on the correlation between adjacent MDCT coefficients in the same frame. But for AAC, the steganography schemes based on MDCT [2–4] have different influence. Those schemes modified MDCT coefficients from different domain, for example: the values of MDCT [2], the sign of MDCT [3] and the Huffman codeword [5]. The different embedding domain will cause different characters distribution. Since that the current MP3 steganalysis schemes didn't work well for AAC. To improve the generality of the AAC steganalysis, in this paper, based on the structure of AAC frame, multi-order difference between inter-frame and intra-frame are extracted, 16 sets of features are proposed to measure the influence of MDCT modification comprehensively.

The rest of the paper is organized as follows. In Sect. 2, The characters of MDCT in AAC are analyzed. In Sect. 3, our proposed AAC steganalysis scheme based on MDCT Difference between Intra and Inter Frame are introduced. In Sect. 4, the performance of each sub-features are analyzed. Experiments results are shown in Sect. 5, and conclusion is drawn in Sect. 6.

2 The Characteristics of MDCT in AAC

2.1 The Principle of AAC Coding

AAC [15] is a lossy perceptual compression standard which is based on psychoacoustic model. Its encoding principle is showed in Fig. 1. The AAC encoding process mainly includes those modules: Psychoacoustic model, Filter-bank MDCT transform, Spectral processing, MDCT coefficient quantization and Huffman coding. The core encoding process is MDCT coefficient transform and quantization.

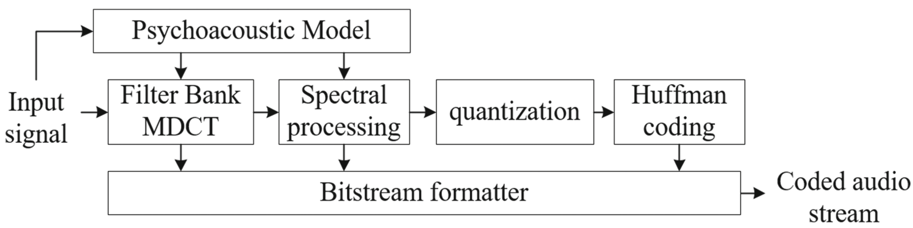


Fig. 1. Encoding procedure of AAC

The psychoacoustic model calculates the maximum distortion energy which is masked by the input signal energy, and figure up a set of signal-to-mask ratios and thresholds, labels the block type of the MDCT with long, start, stop or short types. The filter bank takes the appropriate block of time samples, modulates them by the appropriate window function, and performs the MDCT process. Each block of input samples is overlapped by 50% with the immediately preceding block and the following block. Quantization module is subdivided into three levels which are called frame iteration loop, outer iteration loop and inner iteration loop. The inner iteration loop quantizes the MDCT coefficients until they can be coded with an available number of bits. Outer iteration loop checks the distortion of each sub band, and if the allowed distortion is exceeded, attenuates the sub-band and calls the inner iteration loop again. Huffman coding is done inside the inner iteration loop, it is part of an iterative process, 1024 quantized MDCT coefficients are clipped into several sub-bands, and section coding is used. A section may include several sub-bands, each section use one codebook to encode. In each sub-band, four or two quantized MDCT coefficients are grouped and encoded as one Huffman word.

MDCT coefficients are the mainly coding parameters of AAC. According to AAC coding principle, there are two types of AAC audio frame: long frame and short frame. If the audio signal is smoothly and steadily, long frame is used, and there are 1024 MDCT coefficients in each long frame. Short frame is used when the signal is acutely. The MDCT coefficients are divided into 8 uniform short window, and each short window includes 128 MDCT coefficients. Each window is coded by MDCT transition, quantization and Huffman coding separately, which is called short frame. The structure of AAC frames is shown in Fig. 2, there are long frames and short frames, they are encoded separately.

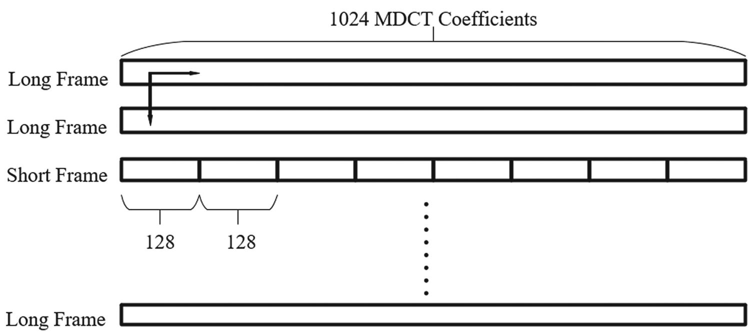


Fig. 2. The construction of AAC frame

2.2 Influences of MDCT Embedding

MDCT coefficients are the mainly coding parameter of AAC, a slightly modifications of MDCT coefficients wont bring in obviously hearing perception, so steganography schemes of AAC MDCT domain [2,3,5] usually have a good concealment and large embedding space. However, due to the short-term relationship of the audio in the time domain and frequency domain, the MDCT coefficients in

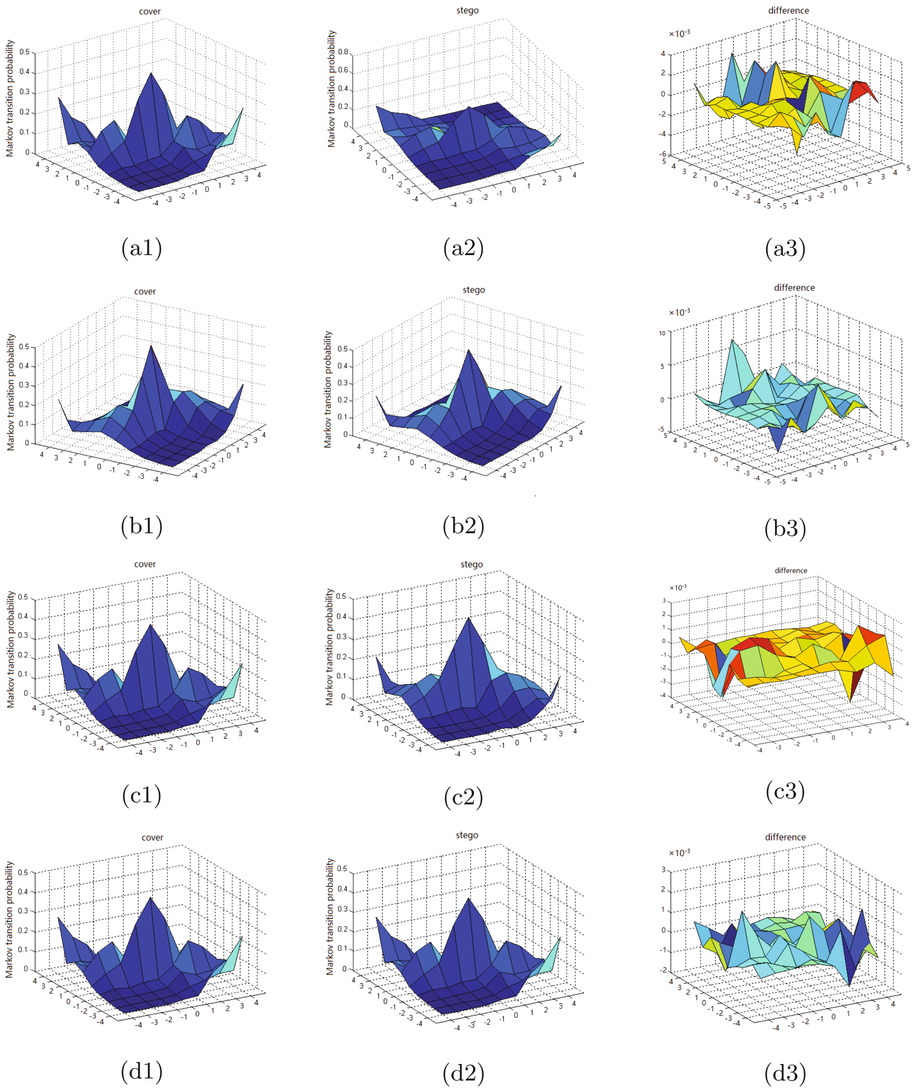


Fig. 3. Markov transition probability of first and second order differential MDCT coefficients. *a1–a3* and *b1–b3* are the first order of features distribution from intra-frame and inter-frame separately. *c1–c3* and *d1–d3* are the second order of features distribution from intra-frame and inter-frame separately. *a1, b1, c1, d1* are cover audios. *a2, b2, c2, d2* are stego audios. *a3, b3, c3, d3* are the difference between cover and stego audios.

the same frame and adjacent frames will keep certain degree of correlation. So the modification of MDCT coefficients will disturb those correlations. Based on this ideal, we proposed to design universal steganalysis features which can represent the intrinsic correlation of MDCT coefficients, distinguish the difference of the correlation distribution of cover and stego audio. Due to the obvious difference of long frame and short frame, we deal with them independently. The correlation between the adjacent MDCT coefficients in the same frame (intra-frame) and the correlation between the MDCT coefficients which has same frequency in the adjacent frame (inter-frame) and evaluated, multi-order differences are analyzed, include first order and second order, and the main statistical correlations include Markov transition probability and accumulative neighboring joint density.

Figure 3 shows the Markov transition probability of 100 audio samples' long frames, includes the cover, stego and their difference. There are four kinds data: first and second order difference of inter-frame, first and second order difference of intra-frame. All the stego audios are generated by Zhu [3], 96 kbps bitrate and the relative embedding rate (RER) is 100%. From Fig. 3 we can find that the modification of MDCT coefficients will make those correlations of MDCT coefficients changed. The changes in intra-frame are much more sensitive than the changes in inter-frame.

3 AAC Steganalysis Scheme Based on MDI2

3.1 Proposed Scheme

It is an effective method to improve the generality of the steganalysis method to extract the rich features from different angles [16–20]. In this paper, based on the analysis of the correlation of MDCT coefficient of AAC frames, we proposed an AAC steganalysis scheme based on MDCT difference between Intra-frame and Inter Frame (MDI2). Figure 4 shows the procedure to extract the feature sets in our proposed scheme. All AAC frames are divided into two groups by their block type: long frame and short frame, combined separately, and extracted the features independently. For each kind of frame, the correlation of MDCT coefficients between intra-frame and inter-frame are analyzed. In each analysis domain of intra-frame or inter-frame, two degrees of adjacent relation are measured, including first order and second order difference between adjacent MDCT coefficients in their analysis domain. For each adjacent relations, two correlation metrics are measured, including Markov transition probability and accumulative neighboring joint density. Therefore, there are 16 sets of sub-features in MDI2. Based on the analysis as follow, the range of MDCT coefficients is from -4 to 4, the number of each sub-feature is $9 * 9 = 81$, and the dimensionality of MDI2 is $16 * 81 = 1296$. Due to the good performance of ensemble classifier [21] on high dimensional features, ensemble classifier was adopted in our schemes to evaluate the classification performance of our proposed features. The key methods in this schemes are described in detail as follows.

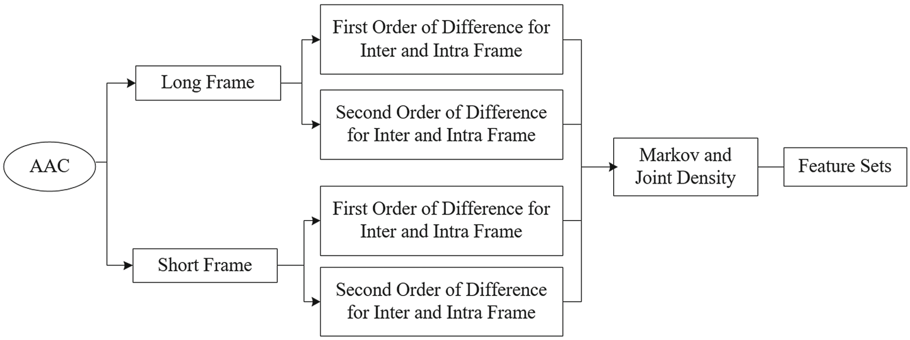


Fig. 4. Feature construction method of the proposed scheme

3.2 Differential Matrix of Inter-frame and Intra-frame

The differential operation of signal sequence can be used to distinguish the isolated point effectively. In order to get a better view of the difference between cover and stego audio, we extract features from differential MDCT coefficients. In addition, considering the limited detecting performance of independent intra-frame or inter-frame feature, we have thought of both inter-frame and intra-frame features separately.

A MDCT matrix: $M_{N*1024} = \{f_1, f_2, \cdots, f_i, \cdots, f_N\}$ is formed. N means the number of frames, f_i is consisted by 1024 MDCT coefficients of the i_{th} frame (for short frame, the number of MDCT coefficients is 128 here). Secondly, first and second order differential MDCT coefficients of intra-frame and inter-frame are calculated, which are showed in Eqs. (1) to (4). r and c are the index of MDCT matrix, so we can get four differential matrix M_{order}^{type}, in which *order* stands for the order of differential matrix, *type* stands for feature constructing direction, and *inter* is inter-frame and *intra* is intra-frame.

$$M_1^{inter} = \{x_{r,c} | x_{r,c} = M(r+1, c) - M(r, c)\} \tag{1}$$

$$M_2^{inter} = \{x_{r,c} | x_{r,c} = M(r, c) + M(r+2, c) - 2 * M(r+1, c)\} \tag{2}$$

$$M_1^{intra} = \{x_{r,c} | x_{r,c} = M(r, c+1) - M(r, c)\} \tag{3}$$

$$M_2^{intra} = \{x_{r,c} | x_{r,c} = M(r, c) + M(r, c+2) - 2 * M(r, c+1)\} \tag{4}$$

In order to reduce the computational complexity, a thresholding technique is applied to $\mathbf{D}$ ($D \in \{M_1^{inter}, M_2^{inter}, M_1^{intra}, M_2^{intra}\}$) as Eq. (5). For an arbitrary $x_{r,c}, \{x_{r,c} | x_{r,c} \in D\}$, r and c is the index of $\mathbf{D}$. If $x_{r,c}$ is either larger than a predefined threshold T or smaller than $-T$, it will be represented by T or $-T$ correspondingly. In our experiment part, T is set to 4.

$$x_{r,c} = \begin{cases} T & x_{r,c} \geqslant T \\ x_{r,c} & -T < x_{r,c} < T \\ -T & x_{r,c} \leqslant -T \end{cases} \tag{5}$$

3.3 Sub-feature of Differential Matrix

Markov transition probability is a good representation to analysis the correlation between data, and is a very effective steganalysis feature in other domains, such as image [22,23] and videos [24,25]. Since that, Markov transition probability is adopted to analysis the correlation of differential sequence in our scheme. Accumulative neighboring joint density is a probability distribution that gives the probability that each evaluated data falls in any particular range or discrete set of values specified for that variable, it shows the correlation of two data from the angle that the two data appear simultaneously, it will be a good complementary feature to Markov transition probability. Therefore, the accumulative neighboring joint density is also considered as an assistant feature set in our proposed scheme.

Markov transition probability of differential matrix $\mathbf{D}$ is calculated by Eq. (6). And accumulative neighboring joint density is calculated by Eq. (7). We extract features of intra-frame in horizontal direction while the features of inter-frame in vertical direction.

$$IM(m,n) = \frac{\sum \delta(x_{r,c} = m, x_{r+k_1,c+k_2} = n)}{\sum \delta(x_{r,c} = m)} \tag{6}$$

$$INJ(m,n) = \frac{\sum \delta(x_{r,c} = m, x_{r+k_1,c+k_2} = n)}{(N_r - k_1) * (N_c - k_2)} \tag{7}$$

N_r and N_c are the rows and columns of $\mathbf{D}$, and $m, n \in [-4, 4]$. $\delta(\cdot)$ is a mathematical operator, its value is set to 1 when it is agreed otherwise it is set to 0. $k_1 = 1$ and $k_2 = 0$ when inter features are constructed, otherwise, $k_1 = 0$ and $k_2 = 1$.

3.4 Sub-feature Fused Based on AAC Coding Principles

According to AAC encoding Principle, there are long and short frames in AAC audio. Long frame is used when the signal is steadily and short frame is used when the signal is acutely. AAC frames will be divided into two groups: long frame group F_l and short frame group F_s. The MDCT matrix M_l of long frames and M_s of short frames are constructed respectively. For F_l, the 1024 MDCT coefficients are put into M_l as a row, and for F_s, we divide each frame into 8 parts to form a new 8*128 sub-matrix, then the sub-matrix is put into M_s. Therefore, 16 sub-features can be obtained totally.

4 Analysis of Sub-Feature

The proposed scheme is a rich feature sets and include many kinds of features. To evaluate each sub-feature's performance, we analyzed those sub-features separately, extract each sub-feature from cover and stego audios to train the classification model, then use the model to detect the test audio samples.

In this experiment, ensemble classifier was adopted, and the metric to measure the performance of the sub-feature is AER (Average Error Rate), which means the average detecting error of the sub-feature, and is calculated as Eq. (8).

$$AER = \frac{n_1 + n_2}{n} \tag{8}$$

n is the total number of cover and stego. n_1 is the number of audios that belong to cover but wrongly classified to stego. n_2 is the number of audios that belong to stego but wrongly classified to cover. In order to describe more clearly, Those sub-features are named as the rule descried in Eq. (9):

$$name = \{frame\_type\}\,\{relation\_type\}\,\{order\}\,\{direction\} \tag{9}$$

$frame\_type$ is the type of frame: L stands for F_l while S stands for F_s. $relation\_type$ is the type of features: M stands for Markov transition probability while J stands for accumulative neighboring joint density. $order$ is the order of differential matrix: 1 stands for first-order while 2 stands for second-order. direction is the direction of feature constructing: B stands for inter-frame while I stands for intra-frame.

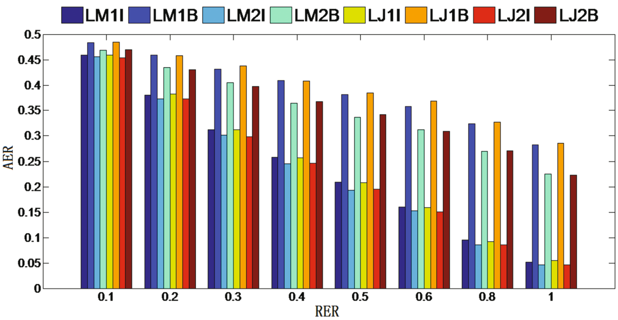


Fig. 5. sub-features' AER of F_l, 96 kbps, Zhu [3]

Figure 5 shows the sub-features'AER of F_l from stego AAC audio encoded with Zhu [3], bitrate is 96 kbps, and Fig. 6 shows the sub-features' AER of F_s from the same audio. Because of Zhu [3]'s better undetectability than Wang [2] and Zhu [5], we only give the AER analysis result of Zhu [3] as an example.

From Figs. 5 and 6 we can find that all the $AERs$ of the 16 sub-features are lower than 50% for all kind of testing RER. Especially, the AER of each sub-feature is lower than 40% when the RER is 50%, which means that all sub-features have ability to distinguish cover and stego audios, so all sub-features can be fused into our final feature sets. In addition, in both long frame and short frame, features of intra-frame outperform the features of inter-frame, it means that in AAC, the correlation between the MDCT coefficient in same frame is more tightly than in the adjacent frame. Comparing with the existed steganalysis schemes for MP3, In our scheme, the introduction of features of intra-frame is very necessary for AAC steganalysis.

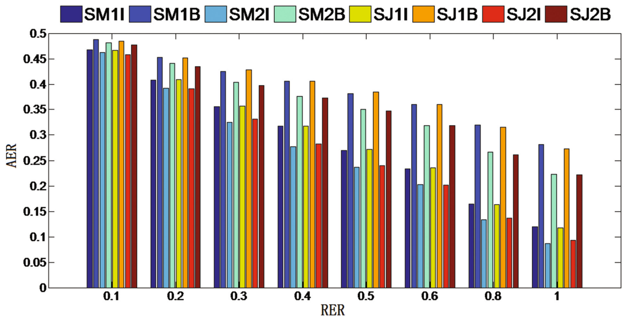


Fig. 6. sub-features' AER of F_s, 96 kbps, Zhu [3]

5 Experiments

To evaluate the performance of proposed scheme, there are two experiments are carried out. Experiment I is to test the performance of proposed scheme to analysis the contribution of the feature of inter-frame. There are two kinds of features set are evaluated: MDI and MDI2. MDI only has the intra-frame features and MDI2 is the whole feature sets which include both the feature of intra-frame and inter-frame. Experiment II is to test the universality of proposed features, and comparing with the existing steganalysis schemes. In this experiment, a mixed model is built to detect three existing steganography schemes.

5.1 Experimental Setup

Audio Database. The AAC audio database contains several types of music, such as jazz, rock and natural. The samples has a duration of 20 s, which contains 1000 frames. WAV audio set consists of 1575 44.1 kHz 16-bit quantization in uncompressed, PCM coded clips which are downloaded from network. The WAV audio set is encoded by AAC encoder [26] into M4A files with the 44.1 kHz sample ratio and bitrate of 96 kbps and 128 kbps.

Steganography Schemes. Wang [2], Zhu [3] and Zhu [5] are implemented to generate the stego AAC audios, for each encoding bitrate described as above, secret information generated by pseudo-random sequence is embedded in the WAV audio samples during the encoding process, at a RER of 30%,50%,80% and 100%.

Contrastive Steganalysis Schemes. There are three contrastive steganalysis features, which are showed in experiment result talbes as Qiao Markov [11], Qiao Joint [11] and Kuriakose [12]. The features of those three schemes are extracted from MDCT domain and used to detect MP3Stego. Qiao Markov [11] extracted Markov transition probability of second order differential QMDCT coefficients from inter-frame. Qiao Joint [11] extracted accumulative neighboring joint density of second order differential QMDCT coefficients from inter-frame. Kuriakose [12] fuses the two feature set (Qiao Markov and Qiao Joint) and obtained a combinated features. In our experiment, we have implemented those three steganalysis features in AAC, and compared the performance of them with proposed scheme.

5.2 Experiments Results

To evaluate the performance of steganalysis scheme, True positive rate (TPR) and true negative rate (TNR) are used as metric. TPR stands for the proportion of stego audios correctly classified, and TNR stands for the proportion of cover audios correctly classified.

Results of Experiment I. To detect the steganography schemes: Wang [2], Zhu [3] and Zhu [5], classified model are trained separately, and the performance of MDI and MID2 are analyzed. The experiment results are showed in Table 1. We can find that the performance of MDI2 is better than MDI, it means that the feature of the inter-frame is necessary. And for different steganagraphy scheme, Wang [2] is easy to detect, for MDI2, even for the RER 30%, the TPR is bigger than 96.67%. And for Zhu [3], the detecting performance of MDI2 under RER 30% is 61.56 %, it means that this kind of steganagraphy scheme is more safer than Wang [2].

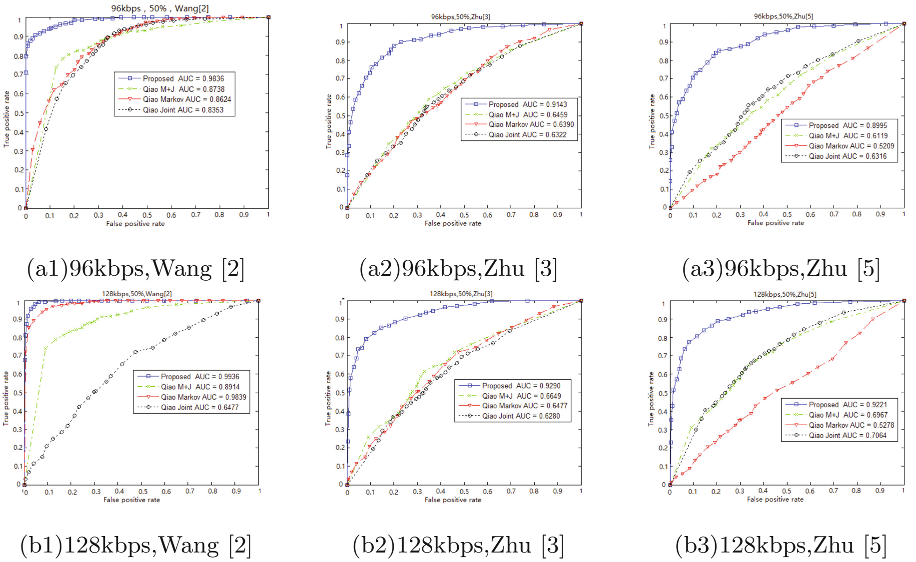
Table 1. Steganalysis performance of different scheme

Bitrate	Embedding Schemes	Feature Sets	TPR				TNR
			30%	50%	80%	100%	cover
96 kbps	Wang [2]	MDI	92.25%	96.31%	99.93%	100%	95.31%
		MDI2	96.67%	99.92%	100%	100%	99.83%
	Zhu [3]	MDI	50.33%	82.67%	94.44%	96.73%	91.21%
		MDI2	61.56%	90.33%	98.56%	100%	95.44%
	Zhu [5]	MDI	86.89%	93.01%	97.89%	98.97%	91.20%
		MDI2	92.58%	98.37%	100%	100%	94.83%
128 kbps	Wang [2]	MDI	94.37%	97.31%	99.36%	100%	96.49%
		MDI2	96.55%	99.98%	100%	100%	98.46%
	Zhu [3]	MDI	63.48%	84.67%	94.65%	98.29%	89.46%
		MDI2	71.82%	89.34%	99.67%	100%	92.37%
	Zhu [5]	MDI	85.34%	90.76%	94.14%	99.68%	91.45%
		MDI2	88.46%	92.16%	99.52%	100%	96.29%

Results of Experiment II. In this experiment, to test the universality of the proposed scheme, a mixed model of MDI2 feature is built to detect all of three steganography schemes. At the same time, the performance of existing steganalsyis schemes are compared, including Qiao Markov [11], Qiao Joint [11] and Kuriakose [12]. The experiment result is showed in Table 2. From Table 2, we can find that the proposed scheme outperforms the other three contrastive schemes under all encoding bitrate. The TNR of MDI2 is above 82% while the

Table 2. Universality of steganalysis performance with MDI2 feature sets

Bitrate	Steganalysis Schemes	Embedding Schemes	TPR 30%	TPR 50%	TPR 80%	TPR 100%	TNR cover
96 kbps	MDI2	Wang [2]	89.63%	96.54%	98.94%	100%	82.18%
		Zhu [3]	55.85%	83.78%	97.61%	100%	
		Zhu [5]	62.33%	80.32%	96.01%	97.87%	
		Average	**69.27%**	**86.88%**	**97.52%**	**99.29%**	
	Kuriakose [12]	Wang [2]	56.91%	88.83%	97.34%	98.94%	66.49%
		Zhu [3]	46.01%	55.32%	66.22%	71.54%	
		Zhu [5]	50.00%	49.47%	51.06%	53.99%	
		Average	**50.97%**	**64.54%**	**71.54%**	**74.82%**	
	Qiao Markov [11]	Wang [2]	41.76%	99.47%	100%	100%	69.68%
		Zhu [3]	47.34%	50.53%	61.70%	69.68%	
		Zhu [5]	33.51%	34.57%	35.90%	39.10%	
		Average	**40.87%**	**61.52%**	**65.87%**	**69.59%**	
	Qiao Joint [11]	Wang [2]	12.77%	96.28%	100%	100%	68.35%
		Zhu [3]	43.35%	48.67%	61.44%	61.17%	
		Zhu [5]	59.57%	62.50%	61.97%	61.17%	
		Average	**38.56%**	**69.15%**	**74.47%**	**74.11%**	
128 kbps	MDI2	Wang [2]	94.41%	99.73%	99.73%	99.47%	83.78%
		Zhu [3]	60.64%	85.64%	97.87%	99.47%	
		Zhu [5]	64.84%	84.84%	96.54%	98.67%	
		Average	**73.30%**	**90.07%**	**98.05%**	**99.20%**	
	Kuriakose [12]	Wang [2]	24.20%	88.30%	99.73%	100%	70.48%
		Zhu [3]	44.95%	52.93%	68.09%	76.33%	
		Zhu [5]	55.59%	56.91%	57.98%	59.57%	
		Average	**41.58%**	**66.05%**	**75.27%**	**78.63%**	
	Qiao Markov [11]	Wang [2]	50.53%	91.49%	99.47%	100%	60.90%
		Zhu [3]	49.73%	56.38%	65.43%	70.74%	
		Zhu [5]	41.49%	41.67%	41.67%	44.95%	
		Average	**47.25%**	**63.18%**	**68.86%**	**71.90%**	
	Qiao Joint [11]	Wang [2]	41.22%	89.36%	99.47%	100%	65.96%
		Zhu [3]	46.01%	54.79%	63.03%	69.15%	
		Zhu [5]	54.79%	55.32%	53.19%	53.99%	
		Average	**47.34%**	**66.49%**	**71.90%**	**74.38%**	



(a1)96kbps,Wang [2] (a2)96kbps,Zhu [3] (a3)96kbps,Zhu [5]

(b1)128kbps,Wang [2] (b2)128kbps,Zhu [3] (b3)128kbps,Zhu [5]

Fig. 7. ROC curve of different bitrate and steganography method from different steganalysis scheme (RER: 0.5)

compared steganalysis schemes are all below 70%. The TPR of MDI2 is above 80% when the RER is above 50%. The proposed scheme outperforms the compared stenalalysis schemes under different embedding rate obviously.

The ROC curve of the proposed scheme and the contrastive schemes are showed in Fig. 7, which shows the proposed scheme has a better detecting accuracy than the contrastive schemes. The detecting accuracy increases with the compressive bitrate increases.

6 Conclusion

In this paper, we proposed an AAC steganalysis schemes to detect the steganographies which embedded secret information by modifying MDCT coefficient. The contribution of this paper is that a rich feature sets about the MDCT correlation in audio compression stream are proposed. In this paper, we consider the intrinsic relationship of MDCT coefficient in intra-frame and inter-frame, proposed to extract the steganalysis feature sets of multi-order differential coefficients of Intra and Inter frame (MDI2). According to the structure of AAC frame, long frame and short frame are analyzed separately, and totaly there are 16 groups of sub-features are extracted. In this paper, the detecting performance of each sub-feature are analyzed, the results show that the correlation in intra-frame is more important than that of inter-frame. The comprehensive experiment results show that the detection accuracy of the proposed scheme are above 85.34% when the relative embed rate is over 50%, which is obviously better than the literatures schemes under the same experiment condition. Due to the similarity of the coding principle of AAC and Mp3, the proposed features can be used to detect MP3 steganography. From the development trend of the steganalysis, high dimensional feature will achieve good generalization ability. Therefore, in our next work, we will attempt to use the thinking of deep learning, fuse with the character of the steganography, to obtain the universal steganalysis features.

References

1. Brandenburg, K.: MP3, AAC explained. In: Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding. Audio Engineering Society (1999)
2. Wang, Y.J., Guo, L., Wang, C.P.: Steganography method for advanced audio coding. J. Chin. Comput. Syst. **32**(7), 1465–1468 (2011)
3. Zhu, J., Wang, R., Yan, D.: The sign bits of Huffman codeword-based steganography for AAC audio. In: International Conference on Multimedia Technology, pp. 1–4 (2010)
4. Wang, Y., Guo, L., Wei, Y., Wang, C.: A steganography method for AAC audio based on escape sequences. In: International Conference on Multimedia Information NETWORKING and Security, pp. 841–845 (2010)
5. Zhu, J.: The research on information hiding in MPEG-2/4 advanced audio coding (AAC). Master's thesis, Ningbo University (2012)
6. Zhu, J., Wang, R.D., Li, J., Yan, D.Q.: A Huffman coding section-based steganography for AAC audio. Inf. Technol. J. **10**(10), 1983–1988 (2011)
7. Shuzheng, X.U., Peng, Z., Wang, P., Yang, H.: Performance analysis of data hiding in MPEG-4 AAC audio. Tsinghua Sci. Technol. **14**(1), 55–61 (2009)
8. Yan, D., Wang, R., Xianmin, Y., Zhu, J.: Steganalysis for MP3Stego using differential statistics of quantization step. Digit. Signal Proc. **23**(4), 1181–1185 (2013)
9. Yan, D., Wang, R.: Detection of MP3Stego exploiting recompression calibration-based feature. Multimed. Tools Appl. **72**(1), 865–878 (2014)
10. Yu, X., Wang, R., Yan, D.: Detecting MP3Stego using calibrated side information features. J. Softw. **8**(10), 2628–2636 (2013)
11. Qiao, M., Sung, A.H., Liu, Q.: MP3 audio steganalysis. Inf. Sci. **231**(9), 123–134 (2013)
12. Kuriakose, R., Premalatha, P.: A novel method for MP3 steganalysis. In: Jain, L.C., Patnaik, S., Ichalkaranje, N. (eds.) Intelligent Computing, Communication and Devices. AISC, vol. 308, pp. 605–611. Springer, New Delhi (2015). doi:10.1007/978-81-322-2012-1_65
13. Jin, C., Wang, R., Yan, D., Ma, P., Yang, K.: A novel detection scheme for MP3Stego with low payload. In: IEEE China Summit & International Conference on Signal and Information Processing, pp. 602–606 (2014)
14. MP3Stego. http://www.petitcolas.net/fabien/steganography/mp3stego/
15. Watkinson: Introduction to Digital Audio Coding and Standards. WatkinsonIntroduction
16. Fridrich, J., Kodovsky, J.: Rich models for steganalysis of digital images. IEEE Trans. Inf. Forensics Secur. **7**(3), 868–882 (2012)

17. Tang, W., Li, H., Luo, W., Huang, J.: Adaptive steganalysis against wow embedding algorithm. In: ACM Workshop on Information Hiding and Multimedia, Security, pp. 91–96 (2014)
18. Denemark, T., Sedighi, V., Holub, V., Cogranne, R.: Selection-channel-aware rich model for steganalysis of digital images. In: IEEE Workshop on Information Forensic and Security, pp. 48–53 (2014)
19. Tang, W., Li, H., Luo, W., Huang, J.: Adaptive steganalysis based on embedding probabilities of pixels. IEEE Trans. Inf. Forensics Secur. **11**(4), 1–1 (2015)
20. Holub, V., Fridrich, J.: Phase-aware projection model for steganalysis of jpeg images. In: SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics, vol. XVII, pp. 94090T–94090T-11 (2015)
21. Kodovsky, J., Fridrich, J., Holub, V.: Ensemble classifiers for steganalysis of digital media. IEEE Trans. Inf. Forensics Secur. **7**(2), 432–444 (2012)
22. Pevny, T., Fridrich, J.: Merging Markov and DCT features for multi-class JPEG steganalysis. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 6505, pp. 650503–650503-13 (2007)
23. Pevn, T., Fridrich, J.: Multi-class blind steganalysis for JPEG images. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 6072, pp. 257–269 (2006)
24. Wang, K., Han, J., Wang, H.: Digital video steganalysis by subtractive prediction error adjacency matrix. Multimed. Tools Appl. **72**(1), 313–330 (2014)
25. Da, T., Li, Z.T., Feng, B.: A video steganalysis algorithm for H.264/AVC based on the Markov features. In: Huang, D.-S., Jo, K.-H., Hussain, A. (eds.) ICIC 2015. LNCS, vol. 9226, pp. 47–59. Springer, Cham (2015). doi:10.1007/978-3-319-22186-1_5
26. Audiocoding.com (2011). http://www.audiocoding.com/