STEGANALYSIS OF AAC USING CALIBRATED MARKOV MODEL OF ADJACENT CODEBOOK

Yanzhen Ren^{*†} Qiaochu Xiong^{*†} Lina Wang^{*†}

* School of Computer Science of Wuhan University

[†] Key laboratory of aerospace information security and trusted computing Ministry of education

ABSTRACT

AAC(Advanced Audio Coding) is the most popular audio compression standard and used widely in recent years. The steganography schemes of AAC emerged gradually. This paper presents a novel steganalysis method to attack the steganography of Huffman codebook, which hide information by modifying the codebook of each scale factor band(SFB), and have good imperceptivity and security. Based on the correlation of neighboring SFBs' codebook, the paper proposes to extract the Markov transition probability of adjacent SFBs' codebook as steganalysis feature, and adopt calibration to improve the accuracy. Extensive experiments demonstrate the effectiveness of the proposed methods. To the best of our knowledge, this piece of work is the first one to detect AAC steganography of Huffman codebook.

Index Terms— AAC, Huffman, codebook, steganalysis, Markov

1. INTRODUCTION

Steganography is a technique of secure communication by hiding secret message in a public media, such as image, audio, video etc. [1]. To the contrary, steganalysis is the art and science of detecting the present of hidden message in suspicious digital media to prevent the illegal use of steganography [2]. In recent years, with the development of multimedia and internet technology, AAC (Advanced Audio Coding) [3], as an effective audio compression standard, has been used widely. AAC is the default or standard audio format for many mainstream mobile phone manufacturers and Internet service providers, such as YouTube, iPhone, Android, and some manufacturers of in-dash car audio systems.

With the widespread of AAC, many kinds of steganography about AAC have been proposed [4–10]. There are three main embedding domains of AAC: MDCT coefficient [4–6], quantization parameter [7, 8], and Huffman coding [9, 10]. The steganography of MDCT coefficient [4–6] modifies the MDCT coefficient to embed secret message. The steganography of quantization parameter [7, 8] is to modify the quantization parameter in AAC encoding procedure, such as quantization factor. Both of them will change the MDC-T coefficient of the audio, and bring distortion to the audio quality. Steganography schemes [9,10] of Huffman codebook are proposed in recent years, they modifies the codebook of each SFBs to embed secret information. Zhu [9] hides the secret messages by adjusting the SFB's codebook within a range of section to match the steganography rules. Tang [10] embeds secret message by controlling the choice strategy of each SFB's codebook. Although this change will affect the compression efficiency, but it won't change the value of MD-CT coefficients. Thus the steganography schemes of Huffman codebook will not introduce any distortion to audio quality and have good imperceptivity.

Compared with the efforts devoted to other compressed audio steganalysis, such as MDCT coefficient and quantization parameter for AAC and MP3 [11-17], steganalysis against Huffman coding for AAC and MP3 remains unexplored. Alomost all the steganalysis schemes [11-17] are seeking the differences of MDCT coefficients between cover and stego audios. While the steganography of Huffman codebook will not change the value of MDCT coefficients, the existing steganalysis schemes will fail on it. In the process of AAC Huffman coding, the choice strategy of cookbook for each SFB is to make the frame's bit number least, thus AAC encoder will try to make the adjacent SFB have the same codebooks to reduce the bit number, and the modification of SFB's codebook will damage the correlation between adjacent SFB's codebook. Markov transition probability was used frequently in image steganalysis field, and this feature can described the correlation and continuity between adjacent or neighbouring block in a quantized method. CC-PEV feature set [18] is classical steganalysis method using Markov feature in image field, and it can reach a satisfactory detection effect. Based on the analysis above, the paper proposed to extract the feature of Markov transition probability of the adjacent SFB's codebook to detect AAC audios.

The rest of the paper is organized as follows. Section 2 introduces the steganography of Huffman codebook. Section 3 shows the statistical properties of the steganography and

This work was supported by the National Natural Science Foundation of China (NSFC) under the grant No. U1536114, U1536204, 61373169, and China Scholarship Council. email: renyz@whu.edu.cn; xiongqiaochu_whu@163.con; lnawang@163.com

illustrates our proposed steganalysis feature in detail. Our staganalysis scheme is shown in Section 4. The performance of our proposed method is analyzed in Section 5. Finally, the conclusions and further works are drawn in Section 6.

2. AAC STEGANOGRAPHY OF HUFFMAN ENCODING

2.1. Principle of AAC encoding

AAC is an audio coding standard for lossy digital audio compression. Fig.1 shows the block diagram of a typical AAC encoder. The process of Huffman coding is to encode MD-CT coefficients lossless. There are 1024 quantized MDCT coefficients for each frame, they are divided to several SF-Bs according to the psychoacoustic model and threshold of human's hearing. For each SFB, Based on the largest absolute value (LAV) of MDCT coefficients, AAC encoder will adopt several Huffman codebooks, try to encode MDCT coefficients, and select the optimal codebook which needs the least number of bits as the SFB's codebook. To reduce the bit number further, Sectioning process is carried out, which merges several SFBs into one section using single codebook to make the bit number of the whole section least.



Fig. 1 Block diagram of AAC encoder

2.2. AAC steganography of Huffman Codebook

Steganography algorithms [9, 10] embed the secret bits by modifying the SFB's codebook. Zhu [9] segments the SF-Bs of each frame into several groups, each group have 2 or 3 SFBs according to the key. They set the rule that if all SFBs in a group has same codebook, the bit embedde in this group is 0, otherwise 1. If the group doesn't match the secret message, one codebook of a SFB will be changed randomly to compliance with the embedding rule. Tang [10] embeds the secret message based on the relationship between selection strategy of Huffman codebook and statistic properties of MD-CT coefficients. For the same MDCT coefficients, the length of bit stream is different while choosing different codebook. AAC encoder will select one which has least bit number as the optimal codebook. It is feasible to embed secret message by controlling the available codebook if two codebooks have same number of bits after encoding. For example, if there are two codebook h_i and h_j (*i* is Huffman codebook index, and i < j) have the same least bit number after Huffman coding for a MDCT sequence, 0 can be represented by codebook h_i , and 1 can be represented by h_j .

3. THE DISTRIBUTION CHARACTERISTICS OF THE CODEBOOK

In the process of AAC Huffman coding, the choice strategy of cookbook for each SFB is to make the length of the frame's bit stream shortest. To achieve it, the AAC encoder will do sectioning operation to make the adjacent SFB have the same codebooks to reduce bit number. Thus the modification of SFB's codebook [9, 10] will change the correlation between adjacent SFB's codebook. The codebook of adjacent SFB in cover audios should be strongly related, and for stego audios, due to the randomness of secret message, the continuity and correlation will be weakened. Based on the above considerations, we measure the continuity and correlation between adjacent SFB's codebook by Markov transition probability matrix of adjacent codebook feature (MAC) to design effective steganalysis scheme.

Some notation is defined to describe the MAC feature clearly at first. The codebooks of each SFB in a frame are denoted by $S = \{c_1, \dots, c_j, \dots, c_N\}$, while j is index of SFB sorted by time sequence, N is the number of SFBs in the frame, and c_i is the j_{th} index of codebook. The codebook in AAC encoder ranges from 0 to 11, and each codebook indicates the information described in [3], which include the LAV, dimension and signed(or unsigned) for MDCT values. There are two special codebooks, zero codebook(index of codebook is 0, all MDCT coefficients are equal to 0) and ESC codebook(index of codebook is 11, the LAV of MDCT coefficients are larger than or equal to 16). If zero and ESC codebook are replaced, the value of MDCT coefficients will be modified extensively, so those two codebooks will not be used in stenography, thus $c_i \in [1, 10]$. The feature of MAC can be calculated by Formula 1. $Pr_{\alpha|\beta}$ is probability that the index of $(j+1)_{th}$ codebook is α while the j_{th} index is β .

$$Pr_{\alpha|\beta} = \frac{Pr(c_{j+1} = \alpha, c_j = \beta)}{Pr(c_j = \beta)}, (\alpha, \beta \in [1, 10])$$
(1)

For S, a Markov transition probability matrix A can be generated by Formula 2. the dimension of A is 10^2 . The element $e_{\alpha\beta}$ of A is the transition probability that codebook of j_{th} SFB equal to β , and codebook of $(j + 1)_{th}$ equal to $\alpha(\alpha, \beta \in [1, 10])$. Obviously, Matrix A reflects the correlation characteristics of neighboring SFB's codebook.

$$A = \begin{pmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,10} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,10} \\ \vdots & \vdots & \ddots & \vdots \\ e_{10,1} & e_{10,2} & \cdots & e_{10,10} \end{pmatrix}$$
(2)

Fig.2 shows the MAC feature of cover and stego audio with 96kbps. The stego audio is produced by steganography method in [9], and the relative embedding rate (RER) is 100%. From Fig.2, we can find the distinguish between cover and stego audio is obvious, especially at the diagonal of MAC feature. But for steganalysis in real world, there is no cover audio to compare, thus if the MAC feature of cover could be estimated, it will be useful to increase the accuracy of steganalysis method. At the same time, the codebook of each SFB is depending on the MDCT coefficients and related to the audio's content closely. In the literature [19, 20], in order to recover the feature of cover image, different quantization matrixes were adopted to recompress JPEG image, extracting difference of testing and double-compressed images as detection feature. Based on the analysis above, we adopt recompression calibration strategy to estimate the feature of original cover audio to decrease the influence of variable audio content and make the MAC feature more stable. The experiment result is shown as Fig.2 and Fig.3.



Fig. 3 The MAC feature of recompressed audios

Due to that the most fundamental purpose of Huffman coding in AAC is to minimize the bit number, different encoders have same target. For AAC audios, whether cover or stego, the recompression will recover the correlation of adjacent SFB's codebook to the cover's, thus we adopt recompression to get the calibrated audio. Fig.3 shows the MAC feature of cover and stego in Fig.2 after recompression, it shows that recompression can recover the MAC feature of the cover audio. So we define Calibrated MAC(C-MAC) feature to classify the cover and stego. C-MAC is the difference between the MAC feature of testing audio and its calibrated one. Fig.4 shows the C-MAC feature of cover and its stego audio, the samples are same as Fig.2. Compared with MAC feature, the distinction of C-MAC feature between cover and stego audio is more obvious, which means C-MAC feature is more sensitive to the change of SFB's codebook.



Fig. 4 The C-MAC feature of cover and stego audios

4. AAC STEGANALYSIS SCHEME BASED ON C-MAC

Through the above basic experiments and analysis, it is shown that C-MAC feature is effective to distinguish cover and stego. Thus, we design a steganalysis scheme based on C-MAC. Fig.5 summarizes the feature extraction process of the C-MAC features.



Fig. 5 Steganalysis algorithm based on C-MAC feature

For the testing AAC segement M1 with encoding bitrate R, MAC features of the testing audio M1 and its calibrated audio M2 are extracted by Formula 2 respectively to get MAC-F1 and MAC-F2. C-MAC feature is the difference between MAC-F1 and MAC-F2. Our proposed steganalysis algorithm adopts a supervised machine learning method. To verify the effectiveness of our proposed features, we adopt LibSVM with its default parameters and RBF kernel as classifier in the experiment.

5. EXPERIMENT

In order to evaluate the detection performance of proposed method, several experiments are carried out. In this section, we will analysis the performance of MAC and C-MAC feature respectively. In addition, two different AAC encoders (FAAC [21] and Nero [22]) are used in calibration to analysis the influence of different encoders.

5.1. Experiment Steup

There are 1,500 WAV audio with 44.1 kHz, 16-bit encoding mode in audio sample set, including several types (blue, jazz, etc.) and different languages (Chinese, English, Korean,

etc). The length is 3 seconds. The WAV audios are encoded by FAAC [21]with the bitrate of 64kbps, 96kbps, 128kbps and 152kbps, the total number of cover AAC audios is 6,000. The steganography described in [9] and [10] are implemented. Secret information is embedded in WAV with a RER of 30%, 50%, 80% and 100% for each bitrate. In the process of calibration, FAAC and Nero [22] encoder are adopted respectively. There are 48,000 stego audios totally.

5.2. Experiment Result

The results of 3 experiments are shown in Table 1, Table2 and Table 3 respectively. While TPR (true positive rate) means the proportion that stego audios are judged as stego, and TNR (true negative rate) means the proportion that cover audios are judged as cover.

5.2.1. Experiment result using MAC feature

To evaluate the performance of MAC feature, we design the experiment to use MAC as the classify feature. FAAC encoder is adopted as compressed encoder. The experiment result is shown in Table 1. It demonstrates that MAC feature has a well performance of steganalysis. TNR is higher than 88% and TPR can reach 85% while the embedding bit rate is 30% or above.

Table 1 The performance of MAC feature

BitRate	stenography	TPR				TNR
		30%	50%	80%	100%	Cover
64kbps	Zhu [9]	87.63%	90.54%	94.13%	95.12%	88 33%
	Tang [10]	85.67%	89.67%	93.99%	94.67%	00.5570
96kbps	Zhu [9]	90.42%	93.28%	94.34%	95.71%	00 87%
	Tang [10]	89.92%	90.67%	92.54%	94.27%	90.87 /0
128kbps	Zhu [9]	92.67%	93.67%	95.03%	96.31%	02 020%
	Tang [10]	91.40%	92.99%	93.93%	94.67%	92.93%
152kbps	Zhu [9]	92.99%	94.63%	95.03%	96.33%	02 33%
	Tang [10]	91.71%	92.99%	93.93%	95.71%	92.3370

5.2.2. Experiment result using C-MAC feature

To evaluate the performance of C-MAC feature, we design the experiment to use C-MAC as the classify feature. FAAC encoder is adopted as the recompressed encoder, which is same with the encoder of the testing audios(FAAC). The experiment result is shown in Table 2. TNR is higher than 95% and TPR can reach 95% when the embedding bit rate is 30%. Table 2 shows that the performance of C-MAC feature is better than MAC feature obviously, and it means that C-MAC feature is more efficient to detect cover and stego.

5.2.3. Experiment result using C-MAC under different decoder

In order to analysis the influence of different encoder to calibration, we design the experiment to use C-MAC as the clas-

Table 2 The performance of C-MAC feature with FAAC

BitRate	stenography	TPR				TNR
		30%	50%	80%	100%	Cover
64kbps	Zhu [9]	96.28%	99.13%	98.93%	100%	07 33%
	Tang [10]	99.13%	99.33%	100%	100%	91.55 10
96kbps	Zhu [9]	98.67%	99.33%	99.93%	100%	00 87%
	Tang [10]	99.13%	99.67%	99.93%	100%	99.01%
128kbps	Zhu [9]	99.13%	99.67%	100%	100%	100%
	Tang [10]	99.33%	99.93%	100%	100%	100 //
152kbps	Zhu [9]	99.54%	100%	100%	100%	100%
	Tang [10]	99.67%	99.93%	100%	100%	100%

Table 3 The performance of C-MAC feature with Nero

		TPR				TND
BitRate	stenography					INK
		30%	50%	80%	100%	Cover
64kbps	Zhu [9]	96.54%	97.13%	98.13%	100%	00 33%
	Tang [10]	98.28%	99.33%	100%	100%	JJ.35 10
96kbps	Zhu [9]	99.13%	99.98%	99.93%	100%	00 87%
	Tang [10]	99.33%	100%	100%	100%	99.01 /0
128kbps	Zhu [9]	99.33%	100%	100%	100%	100%
	Tang [10]	99.67%	100%	100%	100%	100 /0
152kbps	Zhu [9]	99.93%	100%	100%	100%	100%
	Tang [10]	99.93%	100%	100%	100%	100 //

sify feature, and Nero encoder is adopted as the recompressed encoder, which is different with the encoder of the testing audios (FACC). The result of experiments is shown in Table 3. Table 3 shows that C-MAC feature will not be affected by AAC encoder at the process of calibration. The C-MAC feature is robust to different encoders.

6. CONCLUSION

In this paper, we propose a novel steganalysis method to detect AAC steganography of Huffman codebook. The main contributions of our work are as follows: firstly, according to our knowledge, this is the first piece of work on detecting on AAC Huffman steganography. It extends the research topics of steganalysis. Secondly, this work expounds that the modification of AAC Huffman codebook can cause significant diversity between cover and stego. Via the analysis and experiments, we find that C-MAC feature is useful and effective to attack the steganography of Huffman codebook. Experimental results show that the performance of our proposed method is perfect. TNR and TPR of C-MAC feature are al-1 above 97%. Another advantage of the proposed method is that it has a strong robustness and is not influenced by encoders. And due to the similarity between MP3 and AAC, the proposed method can be applied to MP3. In the future, we plan to improve the universality of our scheme by obtaining the essential characteristic of cover audio, and will adopt semi-supervised or non-supervision classification method to practice it.

7. REFERENCES

- Qingzhong Liu, Andrew H Sung, and Mengyu Qiao, "Derivative-based audio steganalysis," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 7, no. 3, pp. 18, 2011.
- [2] Rajarathnam Chandramouli, Mehdi Kharrazi, and Nasir Memon, "Image steganography and steganalysis: Concepts and practice," in *Digital Watermarking*, pp. 35–49. Springer, 2004.
- [3] ISO/IEC, "Information technology -generic coding of moving pictures and associated audio information-part 7: Advanced audio coding (aac)," *Doc. nr.: ISO/IEC*, pp. 13818–7, 1994.
- [4] Jie Zhu, Rangding Wang, and Diqun Yan, "The sign bits of huffman codeword-based steganography for aac audio," in 2010 International Conference on Multimedia Technology, 2010, pp. 1–4.
- [5] Yujie Wang, Li Guo, Yifang Wei, and Cuiping Wang, "A steganography method for aac audio based on escape sequences," in *Multimedia Information Networking and Security (MINES), 2010 International Conference on.* IEEE, 2010, pp. 841–845.
- [6] Yu-Jie Wang, Li Guo, and Cui-Ping Wang, "Steganography method for advanced audio coding," *Journal of Chinese Computer Systems*, vol. 32, no. 7, pp. 1465– 1468, 2011.
- [7] Shuzheng Xu, Peng Zhang, Pengjun Wang, and Huazhong Yang, "Performance analysis of data hiding in mpeg-4 aac audio," *Tsinghua Science & Technology*, vol. 14, no. 1, pp. 55–61, 2009.
- [8] Yifang Wei, Li Guo, and Yujie Wang, "Controlling bitrate steganography on aac audio," in *Image and Signal Processing (CISP)*, 2010 3rd International Congress on. IEEE, 2010, vol. 9, pp. 4373–4375.
- [9] Jie Zhu, Rang-Ding Wang, Juan Li, and Di-Qun Yan, "A huffman coding section-based steganography for aac audio," *Information Technology Journal*, vol. 10, no. 10, pp. 1983–1988, 2011.
- [10] Bu Tian Tang, Li Guo, and Zhen Hua Liu, "An information hiding method in advanced audio coding(aac)," *Technical Acoustics*, vol. 27, no. 4, pp. 533–538, 2008.
- [11] Wangyu Jie, *Rearch on Mechanism and key Technologies of Audio Steganalysis*, Ph.D. thesis, University of Science and Technology of China, 2011.
- [12] Mengyu Qiao, Andrew H Sung, and Qingzhong Liu, "Mp3 audio steganalysis," *Information Sciences*, vol. 231, pp. 123–134, 2013.

- [13] Diqun Yan, Rangding Wang, Xianmin Yu, and Jie Zhu, "Steganalysis for mp3stego using differential statistics of quantization step," *Digital Signal Processing*, vol. 23, no. 4, pp. 1181–1185, 2013.
- [14] Chao Jin, Rangding Wang, Diqun Yan, Pengfei Ma, and Kaiyun Yang, "A novel detection scheme for mp3stego with low payload," in *Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit & International Conference on*, 2014, pp. 602–606.
- [15] Rinu Kuriakose and P. Premalatha, A Novel Method for MP3 Steganalysis, Springer India, 2015.
- [16] Diqun Yan and Rangding Wang, "Detection of mp3stego exploiting recompression calibration-based feature," *Multimedia Tools & Applications*, vol. 72, no. 1, pp. 865–878, 2014.
- [17] Honggang Guo, Diqun Yan, Rangding Wang, Zhangyan Wang, Lin Wang, and T. U. Linqiang, "Mp3 steganalysis based on difference statistics," *Computer Engineering and Applications*, 2015.
- [18] Toma Pevny and Jessica Fridrich, "Merging markov and dct features for multi-class jpeg steganalysis," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 6505, pp. SPIE; IS and T – Society for Imaging Science and Technology –, 2007.
- [19] Tomas Pevny and Jessica Fridrich, "Multi-class blind steganalysis for jpeg images," *Proc Spie*, vol. 6072, pp. 257–269, 2006.
- [20] Jan Kodovsky and Jessica Fridrich, "Calibration revisited," *Proceedings of Acm Multimedia & Security Workshop*, pp. 63–74, 2009.
- [21] "Audiocoding.com," http://www.audiocoding. com/, 2011.
- [22] "Neroaaccodec," http://www.nero.com/s, 2014.