A Perception-based Scalable Encryption Model for AVS Audio

Lan Juan[1,2,3], Huang Tie-Jun[1], Qu Jun-Hua[2]

[1] The Institute of Digital Media, Peking University, Beijing 100101, China
[2] Dept. of Computer Science and Technology, North China Electric Power University, Beijing 102206, China
[3] Patent Examination Cooperation Center of SIPO, Beijing 100083, China

ABSTRACT

Audio Video coding Standard (AVS) is China's second-generation source coding/decoding standard with fully Intellectual Properties. As the sixth part of AVS standard, AVS-DRM aims to offer the universal and open interoperable standard for various requirements of digital rights management (DRM) in digital media industry. In its first version, AVS DRM committee document (CD) did not take into account the encryption modes on AVS audio. However, audio content protection also plays an important role in many digital media applications. Therefore, this paper proposes a perception-based scalable encryption approach for AVS audio, which specifies different audio encryption modes by utilizing the perception classification of the audio bitstream and provides multiple security levels by encrypting different audio coding layers. Moreover, the paper incorporates the proposed scalable encryption method in AVS audio fine-granularity scalable codec technique and then implements an AVS audio trusted coder/decoder. Several objective and subjective tests were performed on a set of 12 critical stereo excerpts provided by AVS audio group. The experimental results show that the proposed scalable encryption approach is able to provide a feasible and effective audio protection scheme for a wide range of applications with different DRM requirements. Currently, the scalable encryption scheme has been partly accepted by AVS-DRM final committee document (FCD).

1. INTRODUCTION

With the widespread infusion of digital technologies and the ensuing ease of digital content transport over the Internet, Digital Rights Management (DRM) of multimedia data have therefore become of critical concern. In order to meet different applications, Audio Video coding Standard working group is specifying the AVS-DRM standard. In its first version, AVS-DRM CD did not take into account the encryption modes on AVS audio. However, more and more recent conflicts in MP3 copyrights indicate that audio content protection also plays an important role in many digital media applications. Therefore, the paper focuses mainly on how to protect AVS audio efficiently and then proposes several audio encryption modes to AVS-DRM Group.

Partial encryption or selective encryption has been used to protect the video and audio content by only encrypting a part of data. As a kind of selective encryption, scalable encryption aims to provide different protection levels according to different application. With the exponential growth of multimedia applications in wired and wireless communication environments, many scalable encryption methods have been developed for compressed image and video data to provide multilevel protection, such as in [1], [2]. However, audio scalable encryption methods have to our knowledge not been studied much. Some previous works on MP3 are always based on the frequency-selective method that selects different frequency coefficients to be encrypted [3][4]. While for speech, scalable encryption methods first classify audio data according to the importance of different parameters, and then select the corresponding parameters to be encrypted [5].

Generally speaking, the bits of an audio bitstream are not perceptually equally important. Bit errors can have vastly different perceptual impact, depending on specifically which bit is corrupted [5]. Moreover, since AVS audio uses Fine Granularity Scalability (FGS) codec technique, it is natural to choose the content-dependent multiple layers encryption method for AVS audio protection. Therefore, this paper proposes a perception-based scalable encryption approach for AVS audio, which specifies different audio encryption modes by utilizing the perception classification of the audio bitstream and provides multiple security levels by encrypting different audio coding layers. The main advantage of the proposed approach is that the users are able to trade or degrade the security level in order to meet other appropriate demands.

The paper is organized as follows. The format of AVS audio is presented in Section 2, and the proposed scalable encryption approach is presented in Section 3. Thereafter, the objective and subjective experiments and results are described. Finally, we conclude this paper.

2. AVS AUDIO FORMAT

AVS audio defines two formats, namely Audio Storage Format (AASF) for storage and Audio Transport Format (AATF) for transmission [6]. The paper takes AASF as the example, but the proposed encryption approach can be easily extended to AATF.

Fig.1 shows bitstream structure of AASF audio which consists of an AASF header and several *raw_data_block*() sequences. Each *Raw_data _block*() is the data unit which can be decoded directly, and consists of main audio data such as 1024 audio samples, related information and other parameters. Moreover, each *Raw_data_block*() has one base layer, denoted by *cbc_base_element*(), and multiple enhancement layers. The base layer is the basic quality code layer, while an enhancement layer is the one to improve audio quality. In AASF, the *cbc_layer_element*() is the basic decoding unit. In general, the base layer often contains several *cbc_layer_element*(), while each enhancement layer contain only one *cbc_layer_element*().

Fig. 2 describes AVS audio FGS mode. To realize FGS, context-dependent bitplane coding (CBC) has been adopted as entropy coding scheme for China AVS audio.

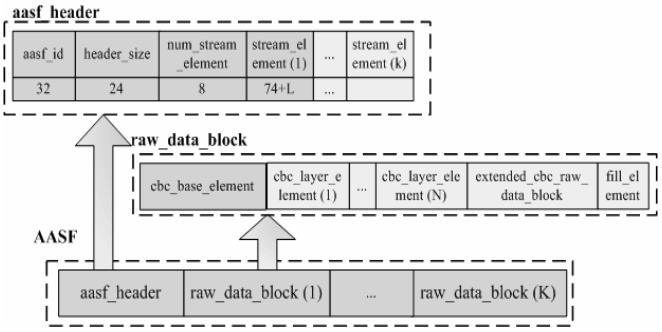

Fig.1. Bitstream structure of AASF audio.

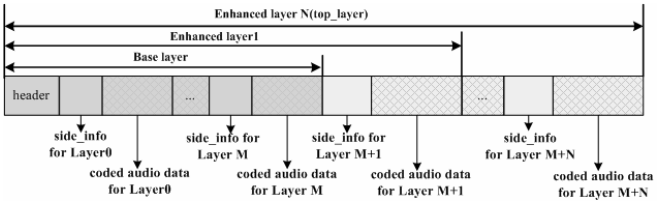

Fig.2. AATF FGS mode.

3. SCALABLE ENCRYPTION

3.1 Perception-based scalable principles

The bits of a compressed speech bitstream are not perceptually equally important. The unequal bit sensitivity can be exploited to design different schemas for different protection levels. Exactly using the unequal bit sensitivity of speech, [5] studied the selective encryption on ITU G.729 speech. By extending this idea, the paper concludes three basic principles on perception-based scalable encryption for audio as follows.

Principle I (Design Goal): The design goal of perception-based scalable encryption for audio is not to preserve perceptual quality above given levels after transmission and decoding, but to cause the desired degree of signal content degradation in terms of naturalness and intelligibility [5].

Principle II (Content Selection): Perception-based scalable encryption approach should perceptually partition the audio bitstream. It should select different data subsets to be encrypted according to different protection goals.

Principle III (Scalability): The perception-based scalable encryption approach should support various kinds of scalability for both FGS and non-FGS audio stream, such as scalability via single encrypted stream, loss-resilient scalability, and flexible accessibility to encrypted audio content.

3.2 Perception-based scalable encryption for AVS audio

Based on AVS audio format, we design the AVS audio scalable encryption approach as follows:

(1) To reduce intelligibility, we need to encrypt the basic layer data. In this way, the key data is *cbc_layer_element*[1…M], where M is the number of sub-layers of the base layer.

(2) To reduce naturalness, we need to protect the enhancement layer data. In this way, the key data is *cbc_layer_element*[M+1…M+N], where N is the number of sub-layers of the enhancement layers.

It should be noted that for perceptual quality degradation, mode M+N is equal to the full encryption. This has been validated in our experiments. Fig. 3 shows the data partitioning for AVS audio scalable encryption approach.

3.3 AVS audio trusted coder/decoder

Fig. 4 depicts the AVS audio trusted coding/decoding procedures. Compared with the standard AVS audio coding/decoding procedures, we add the scalable encryption

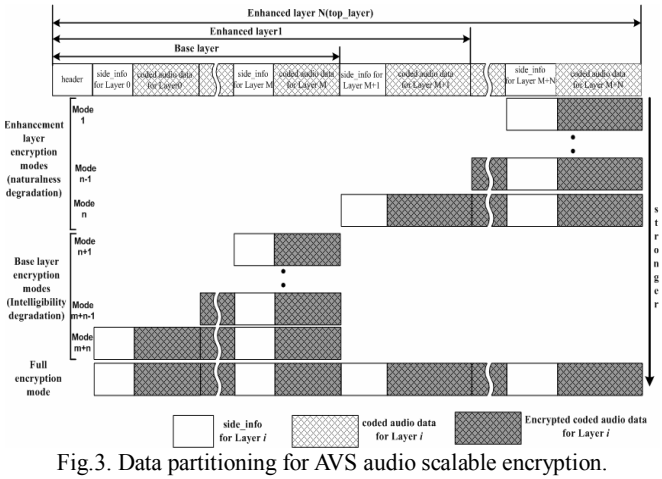

Fig.3. Data partitioning for AVS audio scalable encryption.

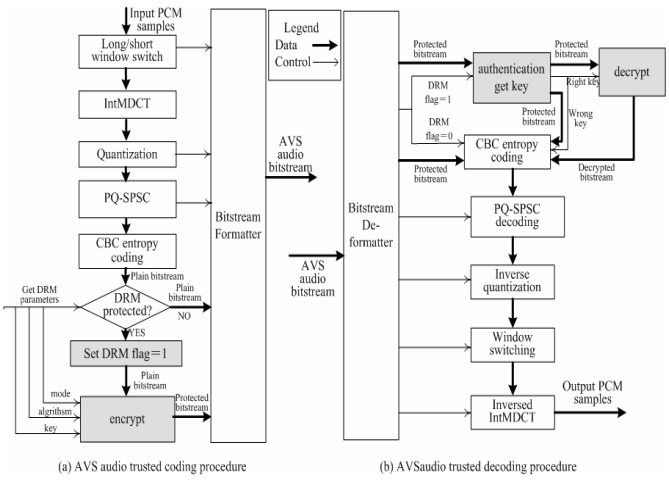

Fig.4. AVS audio trusted coding/decoding procedures.

and authentication steps.

3.4 Security analysis

Full encryption mode changes all bits related to comprehensive data. Its security only depends on the encryption algorithms. Therefore, full encryption mode is far more secure than mode M+N. We can see that in the audio data encrypted by using the proposed approach, there still remain some comprehensible bitsteam structures which might leakage some information for attackers to reduce security. However, the security of the proposed approach also depends mainly on the security of the used cipher algorithms (e.g., DES, AES). In our experiments, it can effectively resist known- and chosen-plaintext attacks, and also has the strong ability to resist brute-force attacks. Thus, we can preliminarily conclude that the proposed approach can be used in many applications with scalable security levels.

4. EXPERIMENTS

We design seven encryption modes according to the proposed method:

- Mode0(M0): plaintext mode;
- Mode1(M1): full encrypted mode;
- Mode2(M2): high enhancement-layer encryption mode with the higher 1/3 enhancement-layers to be encrypted;
- Mode3(M3): low enhancement-layer encryption mode with the lower 1/3 enhancement-layers to be encrypted;
- Mode4(M4): all enhancement-layer encryption mode;
- Mode5(M5): high base-layer encryption mode with the higher 1/3 base-layers to be encrypted;
- Mode6(M6): low base-layer encryption mode with the lower 1/3 base-layers to be encrypted;
- Mode7(M7): all base-layer encryption mode.

Under all modes, the original file format should not be changed so that the file can be still decoded without correct key. Accordingly, the size of audio file doesn't change.

In the experiments, we used the latest coder/decoder and test samples provided by AVS audio group. We also chose 12 original samples with 128 bitrate and 44100Hz sample frequency, and then encoded them in AASF.

2.1 Objective Experiments

Wave pictures can reflect the changes of audio signals directly. Thus we first compare the generated wav files in all modes. Limited by the paper space, here we only give the results of sample *es01*. And we only depict the right channel wave pictures here, where time region is 00:00:00.4-00.00.01.6. The wave picture software is Goldwave V5.10.

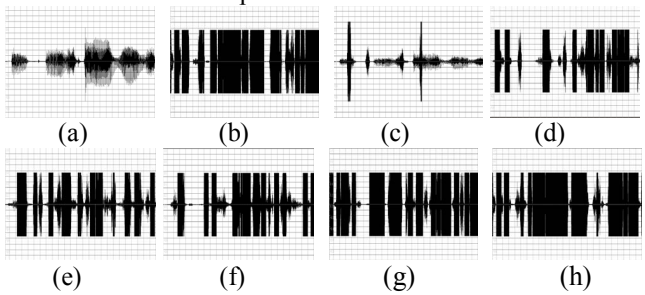

Fig.5. Wave pictures of sample es01 in different DRM modes

(a: M0, b:M1, c:M2, d: M3, e: M4, f: M5, g: M6, h: M7)

From Fig5, we can see that compared with original wave 'a', wave 'b', 'e', 'f', 'g', 'h' have been totally distorted. It can be concluded that M1, M5, M6 and M7 almost have the same effects on audio. There are two reasons. On the one hand, encrypted data length is longer, the degradation will be more seriously. On the other hand, base-layer has the basic quality. Any base-layer data to be encrypted will result in obviously quality degradation. We can also find that wave 'c' is alike 'a' with little distortion. To make sure whether the enhancement-layer has the best scalability, we will study the enhancement-layer mode more deeply in the following experiments.

2.2 Subjective Experiments

AVS audio group recommends Comparison Mean Opinion Score (CMOS) as audio testing criterion [4]. Because of emphasizing on testing audio degradation, we reference Degradation Category Rating (DCR) to examine it and give the Degradation Mean opinion score (DMOS).

Here, we chose results of three samples for testing. Based on objective experiments, we studied M2 much more deeply. So, we encrypted high 8/9 layers (M2-91), high 7/9 layers (M2-92) and high 6/9 layers (M2-93) of the enhancement-layers. Besides that, we also chose each of rest modes into our testing group. Thus, we would generate nine DRM files under such seven modes each group. For AVS audio group hasn't provided AV3 player yet, we chose the decoding wav file to perform the testing. There are twelve people in our lab took part in the experiment. We collected group data and achieved the DCR values, and then computed the DMOS values according formula (1) in table1.

DMOS values according formula (1) in table1. And we drew the degradation picture of AVS audio based on it.

$$DMOS = 1*DCR_1 + 2*DCR_2 + 3*DCR_3 + 4*DCR_4 + 5*DCR_5 \quad (1)$$

Table.1. DMOS values.

Mode	DCR1	DCR2	DCR3	DCR4	DCR5	DMOS
M1	100%	0	0	0	0	**1**
M2-91	36.1%	41.7%	22.2%	0	0	**1.86**
M2-92	25%	30.6%	22.2%	22.2%	0	**2.42**
M2-93	8.3%	30.6%	19.4%	27.8%	13.9%	**3.08**
M3	69.4%	30.6%	0	0	0	**1.31**
M4	72.2%	27.8%	0	0	0	**1.28**
M5	75%	25%	0	0	0	**1.25**
M6	100%	0	0	0	0	**1**
M7	100%	0	0	0	0	**1**

We can draw conclusions from above experiments as follows.

(1) The number of encrypted enhancement-layers is less, DMOS value is higher. And high DMOS means better audio quality.

(2) M3's DMOS is 1.31. It is larger than M1's DMOS. In such mode, audio content can still be understood. So it has more applications than M1.

(3) Effect of encrypting base-layer is the same as encrypting all data. But they are all unacceptable for ordinary audience.

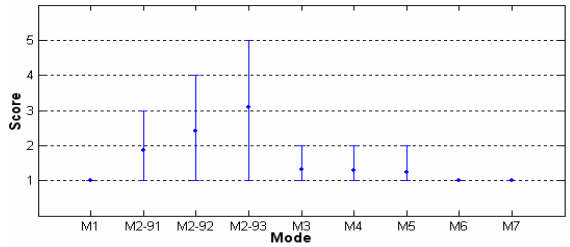

Fig.6. DCR test results of AV3 DRM audio.

(4) Effect of encrypting high base-layers and encrypting low enhancement-layers is also alike.

Besides, different people have different feelings when suffered cacophony. They may give a low score for a good audio after several degradable audio files. So, average values of DMOS can reflect the durable degree of ordinary people. In true application, we may consider to discard the encrypted enhancement layers directly to provide low quality audio without cacophony, which may attract potential users to buy audio.

Based on objective and subjective experiments, together with AVS audio encryption motivation, we categorized proposed modes in table2.

Table.2. categorized the encryption modes by effects.

Degradation	Naturalness	Intelligibility
Seriously	M1、M6、M7	M1、M6、M7
Middle	M3、M4、M5	M4、M5
Slightly	M2	M2、M3

5. CONCLUSION

The paper proposed a perception-based scalable encryption approach for AVS audio, which specifies different audio encryption modes by utilizing the perception classification of the audio bitstream and provides multiple security levels by encrypting different audio coding layers. Based on it, we suggested three additional modes to AVS-DRM CD. Currently, the scalable encryption scheme has been partly accepted by AVS-DRM FCD.

6. ACKNOWLEDGEMENT

This work was supported by National Key Technologies R&D Program of China under Grand No. 2006BAH02A10.

7. REFERENCES

[1] Lindskog S., Strandbergh J., Hackman M. and Jonsson E. A Content-Independent Scalable Encryption Model. In: A. Lagan`a et al. (Eds.) *Proceedings of ICCSA 2004, LNCS 3043*, Heidelberg: Springer-Verlag, pp.821-830, 2004.

[2] Zhu B.B., Swanson M.D., Li, S. Encryption and Authentication for Scalable Multimedia: Current State of the Art and Challenges. In: J. R. Smith, T. Zhang, S. Panchanathan (Eds.) *Proceedings of the SPIE, Internet Multimedia Management Systems*, vol. 5601, pp.157-170, 2004.

[3] Gang L.T., Akansu A.N., Ramkumar M. and Xie XF. On-line Music Protection and MP3 Compression. In: *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing*, Hong Kong. pp.13-16, May 2001.

[4] Sewetti A., Testa C., De Martin J. C. Frequency-Selective Partial Encryption of Compressed Audio. In: *Proceedings of ICASSP 2003*, vol.V, pp.668-671, 2003.

[5] Servetti A. De, Martin J. C. Perception-Based Partial Encryption of Compressed Speech, *IEEE Trans. on Speech and Audio Processing*, vol. 10, pp.637-643, Nov 2002.

[6] Ai H.-J., Chen S.-X., Hu R.-M. Introduction to AVS audio. *Journal of Computer Science and Technology*, vol.21, pp.360-365. May 2006.