A Shift-Recompression-Based Approach to AAC Audio Forgery Detection

Qingzhong Liu¹*, Andrew H. Sung², Lei Chen¹, Zhongxue Chen³, Ming Yang⁴, Jing Zhang⁵# and Yanxin Liu¹,+  

¹Department of Computer Science, Sam Houston State University, Huntsville, TX 77341, USA.  
²School of Computing, University of Southern Mississippi, Hattiesburg, MS 39406, USA  
³School of Public Health, Indiana University Bloomington, Bloomington, IN 47045, USA  
⁴Department of Information Technology, Kennesaw State University, Marietta, GA 30060, USA  
⁵School of Electronic Information Engineering, Tianjin University, Tianjin, 200072, China  

Abstract  
Advanced Audio Coding (AAC), a standardized lossy compression scheme for digital audio, which was designed to be the successor of the MP3 format, generally achieves better sound quality than MP3 at similar bit rates. While AAC is also the default or standard audio format for many devices and AAC audio files may be presented as important digital evidences, the authentication of the audio files is highly needed but relatively missing. In this paper, we propose a scheme to expose tampered AAC audio streams that are encoded at the same encoding bit-rate. Specifically, we design a shift-recompression based method to retrieve the differential features between the re-encoded audio stream at each shifting and original audio stream, learning classifier is employed to recognize different patterns of differential features of the doctored forgery files and original (untouched) audio files.  Experimental results show that our approach is very promising and effective to detect the forgery of the same encoding bit-rate on AAC audio streams. Our study also shows that shift recompression-based differential analysis is very effective for detection of the MP3 forgery at the same bit rate.  

Keywords: these are the keywords, these are the keywords, these are the keywords.  

1. Introduction  
In multimedia forensics, steganalysis and forgery detection are two interesting areas with broad impact to each other. While multiple promising and well-designed steganalysis methods have been proposed and several steganographic systems have been successfully steg-analyzed [3, 5, 6, 7, 10, 11, 12, 15], it seems that the advance in forgery detection falls behind.  
While we enjoy huge volumes of digital multimedia, our traditional confidence in the integrity via our eyes and ears has also been undermined since doctored pictures, video clips, and audio streams are easily manipulated. Generally, tampering manipulation in digital media involves several basic operations and the detection of these fundamental manipulations and relevant forgery has been well studied [1, 2, 4, 8, 9, 13, 14, 16]. To our knowledge, most study of multimedia forgery detection is focused on digital images. Some works have been presented to detect the forgery or related manipulation in audio streams, including MPEG-1 Audio Layer 3, also known as MP3. For example, if two MP3 audio streams encoded at different bit-rates are selected in part and composited together and encoded in MP3 format, such forgery manipulation undergoes double...
MP3 compression. While we will be able to reveal the behavior of double MP3 compression, we may catch the forged part in MP3 audio streams. However, if two MP3 audio streams encoded at the same bit-rate and composited together and encoded in MP3 format with the same bit-rate, the method of detecting double MP3 compression does not work [15].

Advanced Audio Coding (AAC), a lossy audio compression scheme, standardized by ISO and IEC, which was designed to be the successor of the MP3 format, generally obtains better sound quality than MP3 at similar bit rates. AAC is supported on iPhone, iPod, iPad, Nintendo DSi, iTunes, DivX Plus Web Player, PlayStation 3, PlayStation Portable, Wii, Sony Walkman MP3, Sony Ericsson, Nokia, Android, Blackberry, and webOS-based mobile phones [18]. While AAC audio files widely spread, to our knowledge, the literature of the forgery detection of AAC audio files is still missing to this date. Inspired by the method to detect MP3 forgery by checking offset [16] and the method to detect misaligned recompression-based forgery in JPEG images [13], in this paper, we propose a scheme to detect the forgery with the same bit-rate in AAC audio streams, by designing a shift-recompression-based differential analysis with learning classifier. We describe background and relevant work in section 2, and propose our detection method in section 3. Experiments are presented in section 4, followed by our conclusion in section 5.

2. BACKGROUND AND AAC TAMPERING

AAC is a standardized, lossy digital audio compression scheme. It was developed with the cooperation and contributions of companies mainly including Dolby, Fraunhofer (FhG), AT&T, Sony and Nokia, and was officially declared an international standard by the Moving Pictures Experts Group in April of 1997. AAC was promoted as the successor to MP3 for audio coding at medium to high bitrates. It follows the same basic coding paradigm as Layer-3 including high frequency resolution filter bank, non-uniform quantization, Huffman coding, iteration loop structure using analysis-by-synthesis but improves on Layer-3 in a lot of details and uses new coding tools for improved quality at low bit-rates. Its popularity is currently maintained by it being the default iTunes codec, the media player which powers iPod, the most popular digital audio player on the market. Furthermore, the iTunes Music Store, whose sales account for 85% of the market for legal online downloads, sells AAC-encoded songs (encapsulated with FairPlay Digital Rights Management). Compared to MP3, AAC improves the following aspects:

1) More sample frequencies (from 8 kHz to 96 kHz) than MP3 (16 kHz to 48 kHz);
2) Up to 48 channels (MP3 supports up to two channels in MPEG-1 mode and up to 5.1 channels in MPEG-2 mode);
3) Arbitrary bit-rates and variable frame length. Standardized constant bit rate with bit reservoir;
4) Higher efficiency and simpler filter bank (rather than MP3’s hybrid coding, AAC utilizes a pure MDCT);
5) Higher coding efficiency for stationary signals (AAC uses a block size of 1024 or 960 samples, allowing more efficient coding than MP3’s 576 sample blocks);
6) Higher coding accuracy for transient signals (AAC uses a block size of 128 or 120 samples, allowing more accurate coding than MP3’s 192 sample blocks);
7) Can use Kaiser-Bessel derived window function to eliminate spectral leakage at the expense of widening the main lobe;
8) Much better handling of audio frequencies above 16 kHz;
9) More flexible joint stereo (different methods can be used in different frequency ranges);
10) Adds additional modules (tools) to increase compression efficiency: temporal noise shaping (TNS) [20], Backwards Prediction, Perceptual Noise Substitution (PNS) [21], etc.
11) Improved Huffman Coding: In AAC, coding by quadruples of frequency lines applied more often. In addition, the assignment of Huffman code tables to coder partitions can be much more flexible.

Figure 1 shows the general processing flow in MPEG-2 AAC encoding.

Figure 1. AAC flowing chart [22]
removing some voice signals, and the modified audio signal is encoded to AAC audio format. If the bitrate of inserted audio signal fragment is different from the bitrate of original AAC audio signal, or the bitrate used for encoding modified audio signal is different from the bitrate of original AAC audio signal, the tampering may be exposed by revealing the different bitrates once used in the control process. However, if the bitrates are the same, the detection becomes difficult.

Our task targets on the detection of the AAC audio forgery encoded at the same bitrate used for the original AAC audio stream.

3. SHIFT-RECOMPRESSION-BASED APPROACH

3.1. Related Study in Image Forgery Detection

Our proposed approach is based on our previous work in detecting image forgery, therefore, we described the feature extraction for image forgery detection below [13].

1) Decode an JPEG image under examination to spatial domain, which is denoted by matrix $S(i,j)$ ($i=1,2,...,M; j=1,2,...,N$);
2) Shift the matrix $S(i,j)$ by $d_1$ rows and $d_2$ columns in the spatial domain, $(d_1, d_2) \in \{ (0,1), ..., (7,0), (1,0), ..., (7,7) \}$ and generate a shifted spatial image $S'(d_1, d_2) = S(i- d_1, j- d_2)$ ($i=1,2,...,M; j=1,2,...,N$), then compress the shifted spatial image $S'(d_1, d_2)$ to JPEG format at the same quantization matrix;
3) Decode the shifted JPEG image to spatial domain, denoted by a matrix $S''(d_1, d_2)$;
4) Calculate the difference $D(d_1, d_2) = S'(d_1, d_2) - S''(d_1, d_2)$;
5) Shift-recompression based Reshuffle Characteristic features (SRSC) on the region of interest $R$, SRSC$_R$ are defined by:

$$SRSC_R(d_1, d_2) = \frac{\sum |D_R(d_1, d_2)|}{\sum |S_R(d_1, d_2)|},$$

Where $(d_1, d_2) \in \{(0,1), ..., (0,7), (1,0), ..., (7,7)\}$. There are a total of 63 features for each $R$.

If an image was cropped under the misalignment by $p$ rows and $q$ columns, mod $(p,8) \neq 0$ or mod $(q,8) \neq 0$, $0 \leq p \leq 8$, $0 \leq q \leq 8$, and then recompressed at the same quantization matrix to the original JPEG image, we expect that the SRSC features will be distinct due to the misalignment, and the values of $p$ and $q$ can be determined by the SRSC features. Figure 2 shows an original JPEG image (a) and a cropped image with misalignment $p=4$ and $q=4$ (b). The SRSC features from original image and the misaligned recompressed image ($p=4, q=4$), respectively, are shown in (c), and (d). The circle highlights the major differences of SRSC features between the original and manipulated image. It shows that SRSC features may be effective in exposing the misaligned operations that are encoded with the same quantization matrix.
3.2. Proposed Approach to Detecting AAC Audio Forgery

As pointed out in [13], JPEG compression generally generates block artifacts. We surmise that similar to JPEG compression, AAC audio compression also introduces block (frame) artifacts. While two AAC audio files are manipulated together, or some part is removed from an AAC audio file in temporal domain, and doctored audio data are re-encoded in AAC format at the same bit rate, the original block artifacts are generally undermined, in other words, original block switching structure will be reshuffled with a part of the neighbor blocks. By revealing such reshuffling manipulation, we may locate the doctored areas in the AAC audio forgery that was encoded at the same bit rate. According to our previous work in image forgery detection [13], we propose a shift-recompression-based differential analysis to detect the forgery in AAC audio streams with the same compression bit rate, described as follows.

Shift-Recompression-based Differential Analysis

1. Decode the examined AAC audio stream to temporal domain, denoted by a matrix S(i,j) (i=0,1,2, ..., M; j indicates the number of channel of the audio signal).
2. For AAC stationary audio signal, each frame contains 1024 time-domain samples. Allegedly, each frame stands alone and does not depend on previous frames (whereas many perceptual audio codecs such MP3 overlap data with the previous frame).
3. For t=1:1023
   a) Shift the matrix S(i,j) by t samples in the temporal domain (t=1,2,...,1023). A shifted temporal WAV signal S′(i, j, t) is produced. S′(i, j, t) = S(i-t, j), i=t, t+1, t+2, ..., M;
   b) Encode the shifted temporal signal S′(i, j, t) to AAC format at the same bit rate;
   c) Decode the encoded audio signal from the above step to temporal domain, denoted by S″(i, j, t);
   d) Calculate the difference D(i, j, t) = S″(i, j, t) - S′(i, j, t);
   e) Shift-recompression based reshuffle characteristic features (SRSC) are given by:

\[
SRSC(t) = \sum_{(i,j)} D(i, j, t) \sum_{(i,j)} D′(i, j, t)
\]

(2)

We obtain 1023 SRSC features for a stationary AAC audio file since t = 1, 2, ..., 1023.

While AAC audio stream are tampered in temporal domain and original frame structures are generally broken, by checking the SRSC feature under each different shift-recompression, we surmise that untouched SRSC features and tampered SRSC features are different, especially at the corresponding shift. As a result, the manipulation may be exposed.

4. EXPERIMENTS AND DISCUSSIONS

4.1. Detection of Cropping and Recompression

To verify our proposed shift-recompression-based differential analysis, we select 1000 never compressed WAV files; each file is in the length of 20 seconds. These WAV files are compressed in AAC format by using FAAC encoder, which is based on the original ISO MPEG reference code [19]. To simulate the shift-recompression of AAC audio forgery manipulation, AAC audio files are decoded into temporal domain and cropped by different samples at the beginning of the audio signals, then re-encoded in AAC format at the same bit rate. In our study, we tried to produce the cropping database at each possible cropping, or the number of samples removed is set from 1 to 1023, however, the time-consuming is too high to complete. Therefore, in our experiments, the numbers of the samples cropped are only set as 5, 50, 200, 400, 480, 512, 750, 900, and 1000, respectively. 1023 SRSC features are extracted from 1000 untouched AAC audio files and from the nine categories of doctored AAC audio files. Figure 3 shows the SRSC features extracted from an untouched AAC audio file and from the cropping by 50 samples and the cropping by 900 samples individually and recompressed versions.

We apply a popular SVM technique LibSVM [17] with a linear kernel for training and testing. One hundred experiments are performed for training and testing. In each experiment, 60% feature sets from each category are randomly selected for training and the remainders are used for testing. The mean testing of the confusion matrix over 100 experiments are shown by Table 1.

4.2. Detection of AAC audio forgery

In this type of experiments, we randomly select 200 AAC audio files, and remove a few audio samples in the middle, with the block switch offset by 100, 300, 500, 700, and 900 samples, then encode the doctored audio signals into AAC format at the same bit rate. There are total of 1000 doctored AAC audio files. We apply shift-recompression-based differential analysis to each audio file (including untouched and doctored audio files), each audio file is equally divided into six segments, as a result, 1200 untouched segments and 3000 touched segments are obtained. SRSC features are extracted from each segment, in order to discriminate the
doctored audio files from untouched files, and identify the
doctored areas. The experimental design is the same to the
process described in III.A. Table 2 shows the confusion
matrix with the experimental results over 100 experiments.

Figure 4 shows the SRSC features extracted from the six
parts of an original AAC audio file (first row) and from the
six part of the doctored AAC audio file with the forgery
taking place on the middle. The comparison show that the
first three parts of the original audio and doctored audio are
similar, but the pattern of the SRSC features from the last
three parts are different, which approximately reveals the
forged area in doctored AAC audio stream in the middle.

4.3. Discussions

Although our experiments presented above do not examine
all possibility of AAC audio forgery due to the very high
computational cost (it is very time-consuming to examine of
all possible forgery shifted by 1023 positions), by simulating
part cases of AAC audio forgery at a reasonable
computational cost, our experimental results do verify the
effectiveness of our proposed shift-recompression-based
approach to detecting AAC audio forgery of the same bit
rate. The detection accuracy is very promising.

It is worth noting that shift-recompression-based method
is effective not only for detecting AAC audio forgery, but
also for detecting MP3 audio forgery on the same bit rate.
Figure 5 shows an example of SRSC features extracted from
an untouched MP3 audio file and a doctored MP3 file. The
difference is that each MP3 frame contains 576 time-domain
samples, therefore, the feature set only consists of 575
features, not 1023 features.

Our previous study shows that image complexity and
audio signal complexity [11, 15, 23] may play important
role in evaluating the detection performance. In the future,
we will exam the detection performance under different
audio signal complexity, and explore new audio forgery
detection methods at a lower computational cost.

5. CONCLUSIONS

Inspired by our previous study in image forgery detection, in
this paper, we propose a shift-recompression-based
differential analysis to detecting AAC audio forgery with
learning classifiers. Although our method is pretty
straightforward, experimental results show that our approach
is quite promising and effective.

---

Table 1. Confusion matrix on testing sets (mean values, %) by using LibSVM with linear kernel over 100 experiments.

<table>
<thead>
<tr>
<th>Truth</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>untouched</td>
</tr>
<tr>
<td>untouch</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.4</td>
</tr>
<tr>
<td>50</td>
<td>1.8</td>
</tr>
<tr>
<td>200</td>
<td>2.1</td>
</tr>
<tr>
<td>400</td>
<td>1.5</td>
</tr>
<tr>
<td>480</td>
<td>1.2</td>
</tr>
<tr>
<td>512</td>
<td>2.3</td>
</tr>
<tr>
<td>750</td>
<td>2.0</td>
</tr>
<tr>
<td>900</td>
<td>2.4</td>
</tr>
<tr>
<td>1000</td>
<td>2.2</td>
</tr>
</tbody>
</table>
Table 2. Confusion matrix on testing sets (mean values, %) by using LibSVM with linear kernel over 100 experiments.

<table>
<thead>
<tr>
<th>Truth</th>
<th>Prediction untouched</th>
<th>100</th>
<th>300</th>
<th>500</th>
<th>700</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>untouched</td>
<td>98.8</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>100</td>
<td>1.2</td>
<td>98.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>300</td>
<td>0.8</td>
<td>0.1</td>
<td>99.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>0.6</td>
<td>0</td>
<td>0.0</td>
<td>99.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>700</td>
<td>0.8</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>99.2</td>
<td>0.0</td>
</tr>
<tr>
<td>900</td>
<td>1.2</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td>98.6</td>
</tr>
</tbody>
</table>

Figure 4. The SRSC features extracted from an original AAC audio file and from manipulated AAC audio file (doctoring in the middle).

(a) Original MP3

(b) MP3 cropped by 550 samples and encoded at 128 kbps
A Shift-Recompression-Based Approach to AAC Audio Forgery Detection

the same bit rate

Figure 5. SRSC features of original (a) and manipulated MP3 files (b)

Acknowledgements.
The support for this study from the US National Institute of Justice under the award No. 2010-DN-BX-K223 and from the US National Science Foundation under the award CCF-1318688 is greatly appreciated.

References