Detecting Double-compressed MP3 with the Same Bit-rate

Pengfei Ma, Rangding Wang
College of Information Science and Engineering, Ningbo University, Ningbo 315211, China
Email: mpfabc@126.com, wangrangding@nbu.edu.cn

Diqun Yan, Chao Jin
College of Information Science and Engineering, Ningbo University, Ningbo 315211, China
Email: yandiqun@nbu.edu.cn, auking@126.com

***Abstract*—The MPEG-1 Audio Layer 3 can be recorded as archive and lawful evidence. However, this MP3 audio may often be forged by audio forgers for their own benefits in some significant events, which will cause double MP3 compression. In this paper, the statistical features based on scale factors under long window application in the iterative loop are extracted, and a Support Vector Machine is applied for classification to detect double MP3 compression. Experimental results demonstrate that the proposed method is accurate and effective for double MP3 compression detection at the same-bitrate condition. To the best of our knowledge, it is the first time to include other basic features in addition to MDCT coefficients in the scope of double audio compression detection.**

***Index Terms*—double MP3 compression, scale factor, same-bitrate**

I. INTRODUCTION

With the development of the multimedia information technology, varieties of information on events occurring all over the world affect people's normal life according to the content of audios, images and other recordings. However, the sources of them are not always authentic information, such as the well-known Watergate political scandal, which depends on whether a magnetic audio tape had been forged to decide the accused person culpable or not in the court. While the solution in that time is fading nowadays in digital audio [1]. As the current network's most popular music format, MP3 audio can be downloaded from the Internet or recorded as digital files, meanwhile more and more people start to use smart mobile terminals to record sounds as evidences in court. Therefore, research of MP3 audio forensics, which is of great significance, becomes the present and future problem to be solved.

The process of MP3 tampering must take double MP3 compression. As can be seen in Figure 1, an audio forger must decompress the MP3 audio into WAV audio, then manipulate the WAV audio with inserting, deleting, stitching, cutting or other operations in the temporal domain, finally recompress the WAV audio into MP3 audio. So the doctored audio undergoes double MP3 compression, and checking whether the MP3 audio has been suffered double compression can speculates whether the MP3 audio has been manipulated or which position have been tampered.

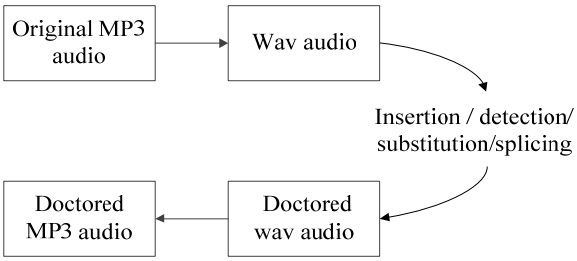


Figure 1. The procedures of double MP3 compression.

While double image and video compression detection have attracted many scholars' attention, research on double MP3 compression detection is rare, especially in the condition of the same bitrate. Yang et al. [2, 3] considered that the number of the MDCT coefficients between -1 and 1 in double compressed audio is less than that in single compressed audio. The proposed method can detect the MP3 audio transcoded from high bitrate to lower bitrate, but the results show low accuracy rate for the opposite situation. Liu and Qiao et al. [4, 5] observed the statistical characteristics of the MDCT coefficients which exceed a certain threshold, and a support vector machine is applied for classification. Their experiment results demonstrated that the algorithm can correctly detect whether the MP3 audio is transcoded from high to lower bitrate or from low to higher bitrate.

In summary, although many researchers have been devoted to double MP3 compression detection under different bitrate, the detection of the double MP3 compression at the same bitrate is still a blank in audio forensics nowadays. In this case, the encoding parameters between the single and double compression are identical, and the difference of the missed information between them is extremely small. Therefore it will be hard for

Corresponding author：Rangding Wang.
Email：wangrangding@nbu.edu.cn

doi:10.4304/jsw.9.10.2522-2527

detecting the double compression with same bit-rate, which is the major work in this paper.

The rest of this paper is organized as follows. In section 2, the procedures of the MP3 encoding are reviewed, which is the preliminary work for feature selection. The statistical features of the scale factors are observed between single and double compressed MP3 audios in section 3. Experiments of detecting double MP3 compression are implemented based on the extracted features in section 4. Finally, the conclusions are summarized in section 5.

II. PRELIMINARY WORK

Figure 2 shows the block diagram of the typical encoder of MP3 audios. MP3 encoding is in units of frame. Each frame contains two granules, and each granule contains two channels in stereo music. Each channel creates 576 frequency lines after MDCT transformation, and the MDCT coefficients are given by the vector xr .

$$xr = \left[xr_0 \cdots xr_i \cdots xr_{575} \right] \tag{1}$$

Psychoacoustic model extracts the frequency information from the FFT transformation prior to the MDCT, then provides information on that which window type should be applied in each channel. If the signal changes smoothly, the long window type is applied to enhance frequency characteristic. On the contrary, if the frame signal changes more violent than normal, the short window type is used to enhance frequency domain resolution.

In the case of the short window application, the 576 MDCT coefficients in each channel are separated into three consecutive packets of 192 coefficients, and each packet is grouped by 12 intervals [6]. In this paper, we concentrated on the case of the long window application. In this situation, each channel is separated into 21 intervals, and each interval corresponds to one scale-factor band. The grouped MDCT coefficients can be shown as the following formula:

$$xr = \left[xr_{[0]} \cdots xr_{[s]} \cdots xr_{[21]} \right] \tag{2}$$

After the MDCT transformation, the most significant work is the iterative loop where the noise and other high frequency information that people can't hear are removed. The inner loop aims to quantify the MDCT coefficients by adjusting the quantization step size to meet the requirement bits of the output bit-stream after Huffman encoding. The quantization formula holds:

$$ix(i) = n\text{int}\left(\left(\frac{|xr(i)|}{2^{\frac{stepsize}{4}}} \right)^{\frac{3}{4}} - 0.0946 \right) \tag{3}$$

where the $ix(i)$ is the MDCT coefficients after quantization, and the $n\text{int}()$ means taking the nearest integer value.

The inner loop would be ended under two following conditions. If both of the conditions are not met, the quantization step size would plus by one and the inner loop would be restarted until either of the end conditions is satisfied.

(1) All the quantitative values are in the scope which the Huffman codebooks can express.

(2) The number of bits after Huffman encoding for the quantitative values does not exceed the number of bits allowed by the encoding parameters.

The outer loop compares the quantization error caused by inner loop with the masking threshold. The scale factors of the sub-band are adjusted and the inner loop is recalled if the quantization error is bigger than the masking threshold. When the number of bits which is determined by encoding parameters can't control the quantization error below the masking threshold, the distortion will be generated [7]. The quantization formula changes as follows.

$$ix(i) = n\text{int}\left(\left(\frac{|xr(i)|}{2^{\frac{stepsize - 2sf\,(1+sf\_scale)}{4}}} \right)^{\frac{3}{4}} - 0.0946 \right) \tag{4}$$

Where the sf means the adjusted scale factor, and the sf_scale is 0 if the scale factor is smaller than 15, otherwise the value is assigned by 1 to increase the range represented by the scale factor. So increasing the quantization step size is equivalent to reducing the scale factor indirectly [8]. Generally, the scale factor does not need to adjust in the case of high bitrate, because of the minor compression ratio and the little quantization error. However, in low bitrate, the outer loop will adjust the scale factor several times [9].

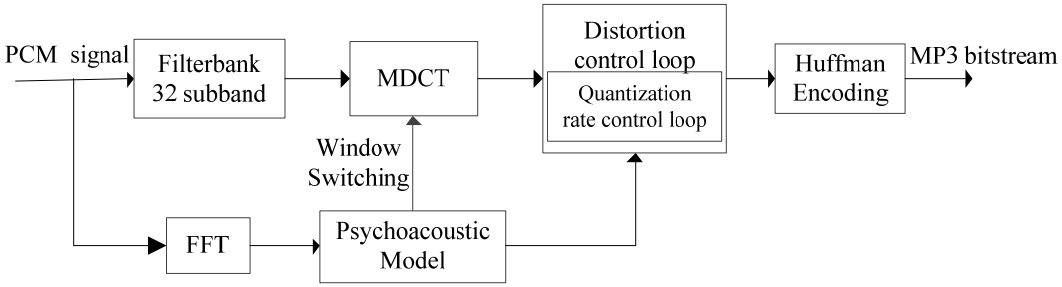


Figure 2. Block diagram of MP3 encoding.

III. Feature Extraction

A. Feature Analysis

The high-frequency components in the original audio maybe the noise that human beings can't heard, so this components accounting for a little proportion would be eliminated during the process of MP3 encoding. Experiments prove that more than ninety percent frames apply long window before MDCT in the single compressed MP3 audio. With the times of the compression increasing, the high-frequency components become fewer and fewer [10], so the percentage of the long window will be bigger in the secondry compression.

The energy of the high-frequency components is so small that the step size must be small enough to meet the precision of the high-frequency information. After the secondary compression, most of the high-frequency components will be quantized to zero, and the step size in double compression will be bigger than that in single compression. However, the situation of the scale factors is just the opposite. The quantization step size and the scale factors must be the significant features for the double compressed MP3 detection duo to the important alteration features between the single and double MP3 compression. Quantization step size responses the overall feature in one frame, while the scale factors can further react the changes of each frame in detail. Therefore, the scale factors are selected as the characteristics to detect double MP3 compression in this paper.

Due to the different encoding rules between long window and short window application, the division of the scale factor in one frame is distinct too. In order to analysis the changes of the scale factors in a certain frame between the primary and secondary compression, the scale factors of the same positions in primary, secondary, and third compression are extracted in this work. In fact, there is no obvious regularity among them. Afterwards the mean differences of the scale factors between this time and the second time compression are calculated in the corresponding channels. The algorithm flowchart shows in Figure 3.

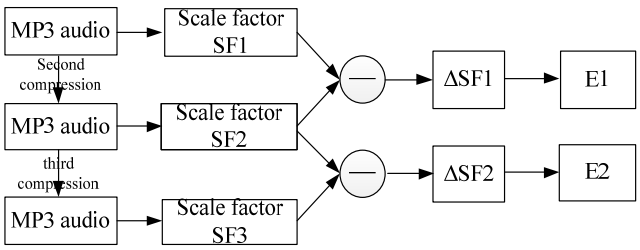


Figure 3. The features of E1 and E2 are extracted.

Wherein $E1$ represents the mean differences between the primary and the secondary compression, and $E2$ means the mean differences between the secondary and the third compression. They represent the single MP3 and the double MP3 compressed characteristics respectively. If it is not specified, the scale factors for the rest of the paper are referred as the scale factors corresponding to the frame where the long window is applied.

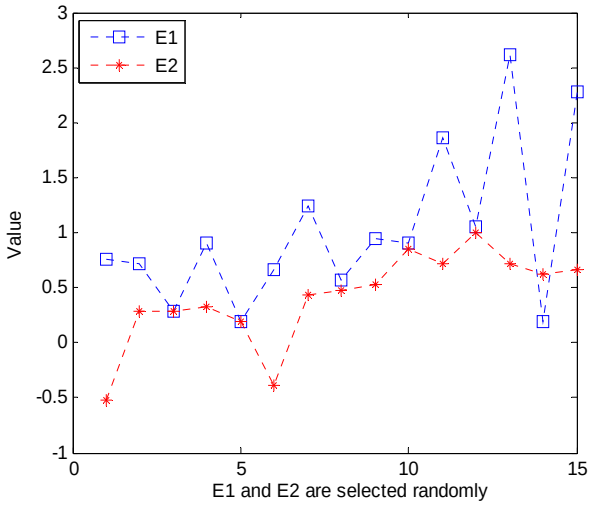


Figure 4. The fifteen mean differences of the scale factors are selected randomly.

As can be seen in Figure 4, in addition to the individual values, $E1$ is bigger than $E2$ as a whole, which means that the scale factors in the primary compression are bigger than that in the secondary and the third compression. In order to observe the features' overall effect, the sum of the $E1$ and $E2$ in every frame are calculated in the corresponding classified features, so one feature in each audio sample is obtained. The results can be shown in Figure 5, wherein the blue curve represents the weighted sum of $E1$ between the primary and the secondary compression, and the red curve shows the weighted sum of $E2$ between the secondary and the third compression.

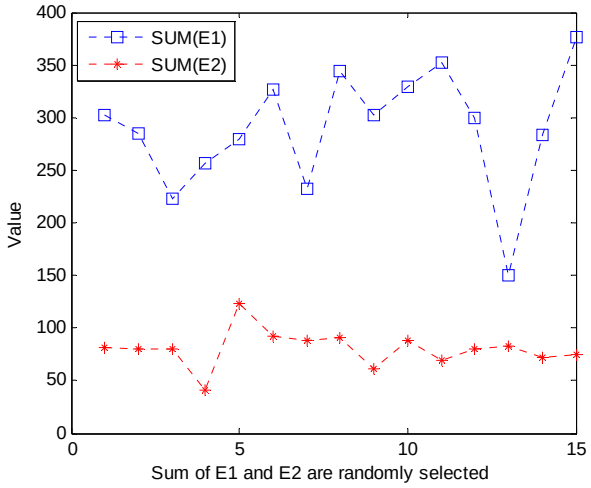


Figure 5. The weighted sum of E1 and E2 in fifteen audio samples.

As can be seen from the Figure 5, the weighted sum of $E1$ is much bigger than that of $E2$, which leaves a significant clue. The $E1$ and $E2$ may be distinguished from the global view. However, the characteristics of the sum in each audio for the detection are so few that the $E1$ and $E2$ are the most important characteristics to be selected for the double MP3 compressed detection.

Support Vector Machine (SVM) [11, 12], a learning machine based on the principle of finding the hyper-planes in a multidimensional space, can effectively separate the features of each type in different classes. So the $E1$ and $E2$ are treated as the two classified vectors, and the SVM is used for classification.

B. Feature Extraction

Each vector contains 250 consecutive values selected in each sample in the following steps.

<1> Long-window frame is defined as the current coding frame which is applied the long window type before MDCT transformation. The WAV audio is obtained after decoding the MP3 audio. The positions of every long-window frame are obtained to combine a column matrix $array = \begin{bmatrix} L_1 & L_2 & \cdots & L_I \end{bmatrix}^T$ during the decoding process. The scale factor matrix of all frames is extracted and defined as equation (5) and equation (6), for the mono MP3 audio and dual-channel MP3 audio, respectively.

$$sf_s = \begin{bmatrix} S_{(2q-1,1)} & S_{(2q-1,2)} & \cdots & S_{(2q-1,n)} \\ S_{(2q,1)} & S_{(2q,2)} & \cdots & S_{(2q,n)} \end{bmatrix} \tag{5}$$

$$sf_s = \begin{bmatrix} S_{(4q-3,1)} & S_{(4q-3,2)} & \cdots & S_{(4q-3,n)} \\ S_{(4q-2,1)} & S_{(4q-2,2)} & \cdots & S_{(4q-2,n)} \\ S_{(4q-1,1)} & S_{(4q-1,2)} & \cdots & S_{(4q-1,n)} \\ S_{(4q,1)} & S_{(4q,2)} & \cdots & S_{(4q,n)} \end{bmatrix} \tag{6}$$

wherein $q = 1, 2, \cdots, I$, $n = 21$, the I donates the number of the long-window frames in the MP3 audio sample.

<2> The scale factor matrix of the long-window encoding frames in the MP3 audio sample is picked up in sequence, and the matrix can be expressed as sf_a. Then the WAV audio will be encoded again with the same bitrate. The scale factor matrix sf_b is extracted using the counts of channels and the array of the long-window positions obtained in step <1>. So the difference matrix can be explained as $\Delta sf = sf_a - sf_b$.

$$\Delta sf = \begin{bmatrix} \Delta S_{(1,1)} & \Delta S_{(1,2)} & \cdots & \Delta S_{(1,n)} \\ \Delta S_{(2,1)} & \Delta S_{(2,2)} & \cdots & \Delta S_{(2,n)} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta S_{(m,1)} & \Delta S_{(m,2)} & \cdots & \Delta S_{(m,n)} \end{bmatrix} \tag{7}$$

wherein the $m = 2I$ when the MP3 audio sample is mono, or $m = 4I$ for the stereo audio.

<3> The mean value of each long-window frame is computed using the equation (8) in Δsf, where $\overline{\Delta sf_{frame}(q)}$ means the mean value of the q th long-window frame of the current MP3 audio. The Δsf_{jx} donates the element in the j th row and the x th column in Δsf. $x = 1, 2, \cdots, n$, $n = 21$, $q = 1, 2, \cdots, I$, $j = qC - q, qC - q + 1, \cdots, qC$. The C shows the constant value of the number of the audio's channels, $C = 2$ for the mono audio, and $C = 4$ for the dual-channel audio.

$$\overline{\Delta sf_{frame}(q)} = \frac{1}{nC} \sum_{j=qC-q}^{qC} \sum_{x=1}^{n} \Delta sf_{jx} \tag{8}$$

<4> Then all the mean values of each long-window frame are taken as the primal features. In order to have a suitable balance between high classifying accuracy and computational complexity, the first nonzero value is selected as the first feature and the 250 consecutive values are chosen for the final features [13] in each MP3 audio sample. Finally the row vector of the features is used for training and testing.

IV. Experiment And Discussion

A. Experiment

Experimental sample library contains five different styles audios of raw format, such as blues, classical, pop, country and folk. 4608 single MP3 compressed audios with 56, 64, 80, 96, 112, 128, 160 and 192 kbps are obtained by lame 3.99.5 which is the latest and most popular MP3 encoder nowadays [14]. The duration of each audio is five seconds. Respectively, these single MP3 audios are decompressed and recompressed at the same bitrate using the same encoder and parameters. Afterwards the same operations are implemented again based on the secondary compressed MP3 audios. Finally, 9216 audios are obtained totally from the secondary compression and the third compression. The features described in section 3 are extracted from all these audios. To discriminate the double compressed MP3 audios from the single compressed MP3 audios at the same bitrate, ten experiments are done with the SVM for classification for each bitrate. 70% features are randomly chosen for training and the rest of them are used for testing in our experiments.

Table 1 shows the average accuracy of the double MP3 compression detection, wherein 'Bitrate' denotes the detection between the single MP3 compressed audios and the double MP3 compressed audios at the same 'Bitrate'. The results consist of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). The True Positive Rate (TPR) and the True Negative Rate (TNR) are denoted as TP/(TP+FN) and TN/(TN+FP) respectively. Finally, the accuracy is calculated as (TPR+TNR)/2 in these experiments.

TABLE I.
TESTING ACCURACY ON DOUBLE COMPRESSION DETECTION (%)

Bitrate	TPR	FPR	TNR	FNR	Accuracy
56kbps	96.35	3.65	89.05	10.95	92.70
64kbps	98.54	1.46	91.53	8.47	95.04
80kbps	99.27	0.73	92.70	7.30	95.99
96kbps	99.37	0.63	96.85	3.16	98.11
112kbps	98.61	1.39	90.66	9.38	94.64
128kbps	98.54	1.46	95.62	4.48	97.08
160kbps	96.60	3.40	94.66	5.34	95.64
192kbps	93.32	6.68	90.32	9.68	91.82

Furthermore, the testing results can also be expressed in Figure 6. The vertical coordinate denotes the true positive rate, which means the single compressed MP3 is detected as the single compressed MP3 and the double compressed MP3 is also detected correctly. The horizontal ordinate shows the false negative rate (FPR), meaning the incorrect detection. When FPR is equal to or bigger than 0.1, all of the corresponding TPR are bigger than 0.9, which means high accuracy of the proposed algorithm.

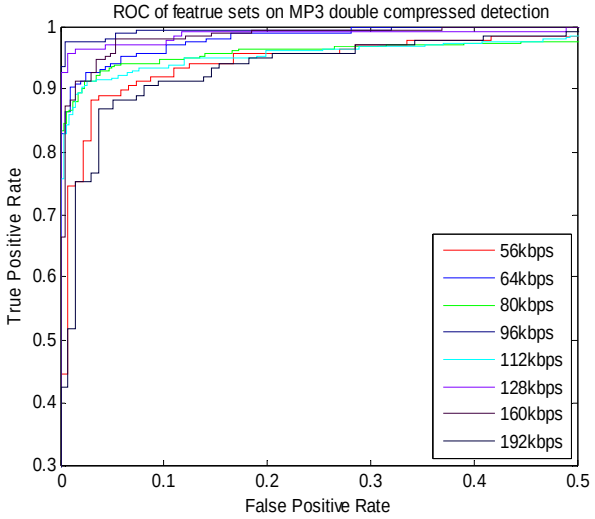


Figure 6. The ROC curve of the testing results.

B. *Discussion*

The experimental results demonstrate that the proposed method of detecting double MP3 compression at the same bitrate performs well, except individual lower bitrate and higher bitrate. For the lower bitrate, the compression rate is so high that many effective features have been lost. Moreover, the span of each frame position in audio is large, and the corresponding relation of the extracted effective features becomes poorer after several compression. These reasons have a bad effect on the detection accuracy. Actually double MP3 compression under low bitrate is rare in the real world, because every time of low bitrate compression will affect the perceived quality seriously, which may arouse suspicion easily. For the higher bitrate, there are enough bits to convey the audio information, including the high frequency information, few features of the audio information change after decompression and recompression, and the extracted corresponding features are approximately similar [15]. So the low detection accuracy will be caused.

V. CONCLUSIONS

In this paper, we have investigated the statistical properties of the scale factors in the frame where the long window is applied, then proposed a method to detect double MP3 compression at the same bitrate based on these features. Experiment results show that our method performs well. Another contribution of this paper is: it is the first time that the scale factor is taken as the feature for the double MP3 compressed detection. However, it is still a challenge for the detection at the lower and the higher bitrates. Also more works would be conducted to forecast multiple compressed MP3 audios to explain the nature of audio compression.

ACKNOWLEDGEMENT

This work is sponsored by K.C. Wong Magna Fund in Ningbo University, National Natural Science Foundation of China (NSFC: 61170137, 61300055, 61301247), Doctoral Fund of Ministry of Education of China(20103305110002), Zhejiang natural science foundation of China (ZJNSF: LY13F020013), Ningbo natural science foundation of China (2013A610057), Scientific Research Fund of Zhejiang Provincial Education Department (Y201119434), Open Fund of Zhejiang Provincial Top key Discipline of Information and Communication Engineering (XKXL1313, XKXL1310).

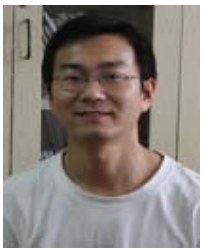
REFERENCES

[1] Swati Gupta, Seongho Cho, C.-C. Jay Kuo. Current developments and future trends in audio authentication. IEEE MultiMedia, 2012, 19(1): 50-59.
[2] Rui Yang, Yunqing Shi, Jiwu Huang. Detecting double compression of audio signal. Media Forensics and Seeurity11, San Jose, California, USA, 2010: 7541:1-10.
[3] Rui Yang, Yunqing Shi, Jiwu Huang. Defeating fake-quality MP3. Proceedings of the 11th ACM Workshop on Multimedia and Security. 2009: 117-124.
[4] Qingzhong Liu, Andrew H. Sung, Mengyu Qiao. Detection of double MP3 compression, Cogn Comput, 2010, 2(4):291-296.
[5] Mengyu Qiao, Andrew H. Sung, Qingzhong Liu. Revealing real quality of double compressed MP3 audio. ACM Multimedia, 2010 : 1011-1014.
[6] Jie Zhu, Rangding Wang, Juan Li, Diqun Yan. The filterbank in MP3 and AAC Encoders: A Comparative Analysis. 2011 International Conference on Electronics, Communications and Control, 2011:1110-1113.
[7] Xuan Fu, Jian Chen, Sheng Xv. The improvements on the quantify Module in MP3 encoding, voice technology, 2004: 52-55.
[8] Zhiheng Gao, Gang Wei. The core technology of the broadband MP3 audio compression. Electroacoustic technology, 2000, 9:9-13.
[9] Renyi Wen, Feng Pan, Junwei Shen. Audio water- marking algorithm against scale factor in MP3 compressed domain, Computer Engineering and Applications, 2012, 48(27): 58-62.
[10] Brian D'Alessandro, Yunqing Shi. MP3 bit rate quality detection through frequency spectrum analysis. Proceedings of the 11th ACM Workshop on Multimedia and Security 2009:57-61.
[11] Shifei Ding, Junzhao Yu, Huajuan Huang, Han Zhao. Twin Support Vector Machines Based on Particle Swarm Optimization. Journal of Computers, 2013, 8(9): 2296-2303.
[12] Jiansi Ren, SVM-based Automatic Annotation of Multiple Sequence Alignments. Journal of Computers, 2014, 9(5): 1109-1116.
[13] Hai Sun, Hongzhi Hu, Weihui Dai, Huajuan Mao, Yan Zhang. Intelligent System for Customer Oriented Design and Supply Chain Management, Journal of Computers, 2012, 7(11):2842-2849.
[14] http://lame.sourceforge.net/. Lame3.99.5, MP3 encoder.

[15] Giancarlo Vercellesi, Andrea Vitali, Martino Zerbini. MP3 audio quality for single and multiple enciding. IEEE International Conference on Multimedia and Expo, ICME 2007, Beijing, China, July 2-5, 2007: 1279-1282.

Pengfei Ma is born in 1987. He is currently the master student in College of Information Science and Engineering, Ningbo University, China. His research interests mainly include Multimedia Information Security, double audio compression detection, information hiding, audio forensics and digital signal processing.

Rangding Wang is born in 1962. Received his M.S. degree in the Department of Computer Science and Engineering from the Northwest Polytechnic University, Xian in 1987, and received his Ph.D. degree in the School of Electronic and Information Engineering from Tongji University, Shanghai, China, in 2004. He is now a professor at Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include multimedia security, digital watermarking for digital rights management, data hiding, and steganography for computer forensics.



Diqun Yan received B.S and M.S. degrees in Circuit and System from Ningbo University, ZhejiangChina, in 2002 and 2008, respectively. He is currently a Ph.D. candidate in College of Information Science and Engineering, Ningbo University. His research interests include digital audio processing and data hiding.

Chao Jin is born in 1990. He is currently the PhD student in College of Information Science and Engineering, Ningbo University, China. His research interests mainly include network and information security, information hiding, audio steganalysis, and audio signal processing