A Novel Steganography Method Based on Modifying Quantized Spectrum Values of MPEG/Audio Layer III*

BEIXING DENG[1], JIE TAN[2], BO YANG, XING LI[3]
Department of Electronics Engineering
Tsinghua University
Department of Electronics Engineering，Tsinghua University，Beijing 100084, CHINA P.R

Abstract: This paper proposes a novel information-hiding method to hide much more information into audios, compared with software mp3stego. The music files in MP3 format is so popular on Internet due to its characteristically high compression ratio and tone quality. Our proposed method can simultaneously accomplish the information embedding procedure when performing compression of MP3. In the proposed method, we modify some quantized spectrum values of audio layer III to embed secret information into audios. The capacity of our method is six times as large as that of mp3stego. At the same time, the difference between original audio and audio with secret information is imperceptible. In the experimental results, we hide more characters into audios and extract them correctly. The audios with secret information are indiscernible to human ears.

Key-Words: Information Hiding, Steganography, MP3, MP

1 Introduction[1]

Owing to popularity and low cost of sales, more and more music framers choose Internet as a channel to issue their creations. Music files in MP3 format are popular on Internet. But steganography software using MP3 audios as hosts is rare. The found information-hiding software using MP3 audios as hosts is mp3stego [1], but its capacity is low. Mp3stego doesn't have the information about the length of the secret information, thus, when extracting secret information, the extracted information includes two parts: the former part is the correct secret information and the latter part is unrecognizable characters. In our paper, this drawback is corrected through embedding information about the length of the secret information in host audios.

A successful information hiding technique should include the following characteristics:

[1] This work is Supported by National Science Foundation of China (No.60473087)

1.1 Imperceptibility

People can't distinguish between the original creation and one with secret information; otherwise, they would detect the existence of secret information.

1.2 Undetectibility

Only those who know the specified key can correctly extract the secret information embedded in hosts. One without the correct key can't extract the secret information.

1.3 High capacity

An information hiding system should have high capacity, that is, we should embed as much information into the same hosts as possible. Our method can embed 12 bits into a frame if the host audio signal is single-channel and 24 bits if the audio signal is dual-channel. Thus, the hiding capacity can achieve about 460bit/s for a single-channel audio signal, 44.1khz sampled, and about 920bit/s for a dual-channel audio signal, 44.1khz sampled. Under the same condition, mp3stego only achieves about 76bit/s for a single-channel audio and about 152bit/s for a dual-channel audio. The capacity of our method is six times as large as that of mp3stego.

1.4 Convenient extracting

A system can extract the secret information without the auxiliary of the original audio data.

Currently, information-hiding techniques are divided into two types: time-domain and frequency-domain. The former is achieved through the modification of the original files [2,5,6]; the latter transforms and encodes the music files into frequency domain before the embedding of secret information. Because frequency coefficients represent different features of the original music data, we can embed secret information into an original by modifying coefficients, slightly influencing musical quality. Up to now, most researches focus on techniques of time-domain information hiding, which can be easily detected. The frequency domain of information-hiding systems is more complicated and requires a large number of computations. MP3 compression [7, 10], is too time-consuming [11]. We therefore propose a information-hiding method which can simultaneously achieve MP3 compression and information embedding. The experimental results show that our proposed method has higher hiding capacity, imperceptibility, undetectibility and secret information can be easily extract in the procedure of MP3 decoding.

During MP3 compression, Layer III filterbank [4] is used to translate time domain into frequency domain. After this translation, the original 576 music signals will be divided into 576 MDCT coefficients [11]. Then, quantization is performed on 576 frequency spectrum lines. We modify those quantized values to embed secret information. More specifically, we modify the last count1-quadruple [4] if this granule [4] doesn't contain short blocks [4] and the last 4 quantized values (4 highest frequency lines) if this granule contains short blocks, which has no count1-quadruples [4]. Through modifying 4 quantized values, we can embed 6 bits in each granule. Findings of our experiment: audio data is slightly affected in the time domain when we modify the quantized values in the frequency domain. An audience cannot tell the quality difference between the original and the revised music.

Remaining of this paper is organized as follows. Section 2 describes the study of related information hiding literature. The encoder and decoder of MPEG Audio layer III are discussed in section 3, our proposed method in Section 4, experimental results and comparison in Section 5. Finally, we present conclusions in the last section.

2 Reviews of related works

Previously, many scholars propose several effective information-hiding schemes [5, 6, 9]. Most directly hide information in audio files, that is, embed secret information in the time domain [3]. In frequency-domain information hiding, each frequency coefficient represents different traits of original audios. If we modify insignificant characteristics, that is, insensitive part of original signals, the modification cannot be detected by human ears [11].

Researchers propose time-domain information-hiding schemes [2, 5, 9] that achieve the goal of information embedding by adding delay into the original audios. The delay U is determined by psychoacoustic models [5] so that U is as long as possible, and yet not found by users. Audios are divided into many sections. Each section adopts appropriate U. When embedding, if secret information bit is 1.1U delay is adopted; if the secret information bit is 0.2U delay is adopted.

Suppose H be a collection of audio signals whose value exceeds a predetermined threshold; I and J are adjacent elements of H. This method [6] changes sample numbers between sample I and J to embed secret information. The number of samples between the two Hs is adjusted to even number for bit 1 and odd number for bit 0.

In [5], the scheme segments original audio data by using fixed length and then three segments as S1-S3, which are adjacent, to represent some operating unit. This scheme computes the energy of three segments independently, and uses Emax, Emid, Emin to represent the maximum energy, secondary maximum energy and minimum energy. Then:

A=Emax-Emid,

B=Emid-Emin.

Then we embed 1 bit into three segments trough adjusting the relationship between A and B.

- The bit of secret information is 1
  If (A-B≥(Emax+2*Emid+Emin)*d), changes are not needed. Otherwise, increase Emax or decrease Emid to satisfy the condition. Where d is a parameter to determine the energy of secret information. d is larger, the secret information is more difficult to erase.
- The bit of secret information is 0
  If (B-A≥(Emax+2*Emid+Emin)*d), no changes are needed. Otherwise, increase Emid or decrease Emax to satisfy this condition.

When extracting, if (A≥B), then the secret bit is 1. Otherwise, the secret bit is 0.

While the mentioned methods have high imperceptibility, they require complicated computations and much time spent in embedding procedure. If MP3 compression is considered, it lengthens waiting time for users. Thus, we propose a method that simultaneously achieves MP3 compression and information embedding.

The existing software that simultaneously achieves MP3 compression and information embedding is mp3stego. It embeds 1 bit into each granule (every 576 audio signals) [4] through modifying the part2_3_length [4]. It adjust part2_3_length to be even if the secret bit is 1 and odd if 0. Although it has high imperceptibility, undetectibiltiy and secret information can be conveniently extracted, its hiding capacity is relatively low. Thus, we propose our method that has hiding capacity six times as large as that of mp3stego, at the same time, keeps high imperceptibility, undetectibility and secret information can be conveniently extracted without the auxiliary of original audio signals.

3 MPEG Layer III Encoder and Decoder

MPEG Audio Layer III's [4] high compression ratio and high quality tone make it popular, most people use this format to compress music. But the only existing information hiding system using MP3 audio file as hosts is mp3stego whose capacity is relatively low. Thus, we propose a method that has much higher hiding capacity.

3.1 MP3 Encoder

The encoder of MP3 is briefly described. The original audio signal is decomposed into 32 subbands after it is dealt with Analyser of Filterbank [4]. In order to increase resolution of frequency, MP3 encoder further divides the 32 subbands with MDCT, and then obtains 576 lines of data. Then, MP3 encoder performs the procedure of quantization and Huffman encoding. Finally, the encoded bitstream is formatted and output [11].

MP3 encoder deletes the subbands, unsusceptible to human ears and masked by other signals, to achieve the goal of reducing bit rate. Then it further compress the signals by using Huffman coding to reduce bit rate and carry out the goal of a large number compression[11].

3.2 MP3 decoder

First, the bitstream is unpacked and the relative information of the bitstream is retrieved. Then, Huffman decoding [4] is performed. We can obtain 576 MDCT lines of data, which have been reconstructed. Then 32 subbands are obtained using IMDCT to deal with the lines of data. Finally, audio data of time domain are obtained by using synthesis filterbank to deal with the subbands [11].

4. The proposed audio information-hiding method

Usually, the value of the high frequency lines of 576 lines in a granule is small. After quantization, most high frequency lines are quantized as 0. Therefore, we slightly modify a few high frequency quantized values will not cause distortion detected by human ears. Our method modifies the last count1-quadruple [4] if the granule has no short blocks, and the 4 quantized highest frequency lines if the granule contains short blocks.

4.1 Embedding procedure

In order to keep fixed bit rate, our method embed secret information in the procedure of quantization. A random generator with a secret key is used to choose frames used for embedding. If the audio is ingle-channel, the first two chosen frames are used to embed information about the length of secret information. If the audio is dual-channel, the first chosen frame is used to embed information about the length of secret information. The embedding method is show in Fig.1. The embedding procedure is show in Fig.2.

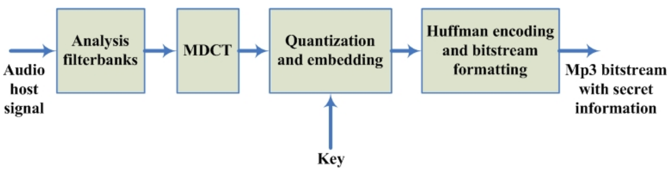


Fig. 1. The embedding method

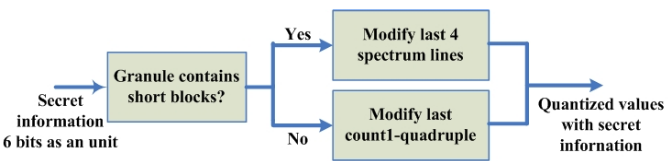
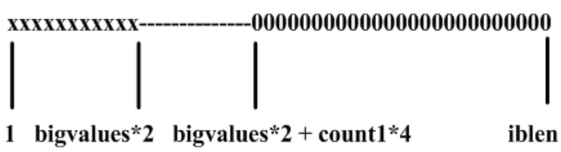


Fig. 2. The embedding procedure

MP3 quantized values are organized as follows [4]:



The values 000 are all zero.

The values --- are -1,0 or +1. Their number is a multiple of 4.

The values xxx are not bound.

Quadruples of quantized values with absolute value not exceeding 1 (-1, 0, 1) are counted as count1. So, quadruple has 81 different combinations. Obviously, the last two values of the last quadruple cannot be 0 at the same time, otherwise, this pair of values will be counted as rzero [4]. Therefore, the number of available combinations is 72. We use 64 of them to represent 6 bits secret information. For example, if we use {-1,-1,-1,-1} to represent 000000, then if the secret bits are 000000, we will modify the last count1-quadruple (granule contain no short blocks) to be {-1,-1,-1,-1} or the last 4 quantized spectrum lines (granule contain short blocks, which has no count1-quadruples) to be {-1,-1,-1,-1}. Generally speaking, our method can be finished in two steps:

- Use a key that is chosen by user as a random generator seed to choose a set of frames for embedding;
- If a frame is chosen to embed secret information, modify special quantized spectrum lines(base on whether a granule contains short blocks or not) of each granule in this frame.

4.2 The extracting procedure

The secret information can be extracted when MP3 decoding is performed. After Huffman decoding, we look over the special quantized spectrum lines to extract secret bits. The extracting method and the extracting procedure are show in Fig.3 and Fig.4 , respectively.

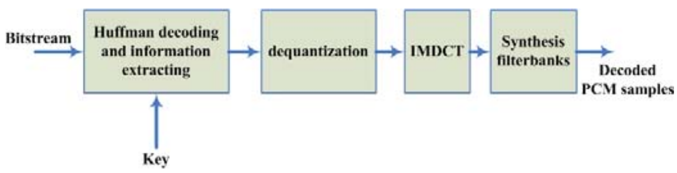


Fig. 3. The extracting method

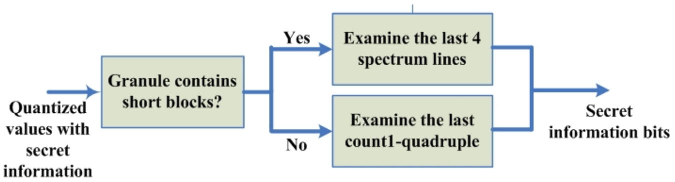


Fig. 4. The extracting procedure

Generally speaking, the extracting procedure can be performed in two steps:

- Use a user-specified key to determine the set of frames that have secret information;
- If a frame has secret information, examine the last 4 spectrum lines or the last count1-quadruple of each granule in this frame to extract information. In the above mentioned example, if the 4 special spectrum lines are {-1,-1,-1,-1}, then the secret bits are 000000. The first 24 extracted bits are information about the length of secret information.

5 Experimental results

We embedded characters into classical, jazz, folk and symphony, and then extract the embedded characters correctly. People cannot distinguish between the original music and those with secret information. Objective tests of audios are unreasonable. High signal-to-noise ratio and high similarity coefficient do not necessarily mean high imperceptibility, so we use subjective tests to test the imperceptibility of our method. We performed two kinds of experiments: one is to present original music and the music with secret information to 20 different listeners (including 5 experts on music) and to ask them to tell the difference between the two; the other is to randomly present an original music and the revised music to an listener and to ask him whether the music is embedded with secret information or not, and then average the discrimination rate of the 20 listeners.

We performed experiments on a large set of music. Experimental results show that most people cannot distinguish between the original music and the music with secret information. In the first experiment, the number of people that cannot tell the difference between the original music and the music with secret information is recorded. In the second experiment, the discrimination rate (the number that people correctly tell whether one music is original or not divided by the total number) is recorded. Some experiment results are show in Table.1.

Table 1. Results of experiment

	Discrimination rate	The number of people that cannot distinguish
Classical	50.1%	20
Jazz	50.01%	20
Folk	49.99%	20
Symphony	50.03%	19

6 Conclusions

We proposed an audio information-hiding system based on modifying special spectrum lines of MPEG layer III. The proposed scheme has the following characteristics. In imperceptibility, most people cannot distinguish between the original music and the one with secret information. In undetectibility, one cannot extract the correct information without the secret key. In hiding capacity, our method has a capacity six times as large as that of mp3stego. In extracting procedure, we can extract secret information without original audio. Our information scheme has characteristics that information-hiding techniques must be responsible for prerequisites.

References

[1]http://www.watermarkingworld.org/WMMLArchive/9808/msg00000.html

[2] P.Bassia and I. Pitas, “Robust audio watermarking in the time domain”, *Europ. Signal Processing Conf*, Rhodes, Greece, 1998.

[3] M. Ikeda, K. Takeda and F. Itakura, “audio data hiding by use of band-limited random sequences”, *IEEE International Conference on*

Acoustic, Speech, and Signal Processing, 4, pp.2315-2318, 1999.

[4] ISO/IEC 11172-3, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s-part 3: Audio", *ISO/IEC, JTC1/SC29*, 1993.

[5] W. N. Lie and L. C. Chang, "Robust and high-quality time-domain audio watermarking subject to psychoacoustic masking", *IEEE International Symposium on Circuits and System*, 2, pp .45-48, 2001.

[6] M. Mansour and A. Tewfik, "Audio watermarking by time-scale modification", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3, pp.1353-1356, 2001.

[7] D. Pan, "A tutorial on MPEG/Audio compression", *IEEE Multimedia*, 2(2), pp. 60-74, 1995.

[8] J. W. Seok and J. W. Hong, "Audio watermarking for copyright protection of digital audio data", *Electronics Letters*, 37(1), pp.60-61, 2001.

[9] C.S. Xu, J. K. Wu and Q. Sun, "A robust digital audio watermarking technique", *in Proceedings of the Fifth International Symposium on Signal Processing and Its Applications*, 1, pp.95-98, 1999.

[10] E. Ambikairajah, AG Davis and W. Wong, "Auditory masking and MPEG-1 audio compression", *IEE Electronics & Communication Engineering Journal*, 9(4), pp.165-175, 1997.

[11] Cheng-Te Wang, Tung-Shou Chen and Wen-Hung Chao, "A new audio watermarking based on modified discrete cosine transform of MPEG/Audio Layer III", *Proceedings of the 2004 IEEE International Conference on Networking, Sensing & Control*, pp.984-989, 2004.