

# Partial Identification Analysis for MP3 Music

Xun Jin and Jong Weon Kim

**Abstract**—With the advent of BitTorrent protocol, internet users can easily share their movie and music files. Although the BitTorrent provides a highly efficient mechanism, the illegal use of BitTorrent on unauthorized distributing of copyright-protected digital content is increasing rapidly. To prevent illegal distribution of copyrighted files from spreading over the internet, the file pieces split by torrent file must be detected as soon as possible. In this paper, we propose a method of audio recognition on a few pieces of audio files downloaded by BitTorrent client program. The audio identification algorithm is made up of beat tracking and chroma features. To identify audio in the piece, the beat-by-chroma representations obtained from the piece and original audio are cross-correlated and look for sharp peaks. The experimental results show probabilities of successfully identifying the pieces with different piece sizes and different number of pieces.

**Index Terms**—BitTorrent protocol, MP3 file analysis, piece verification, audio identification.

## I. INTRODUCTION

BitTorrent protocol designed by Bram Cohen in 2001 is currently the most popular peer-to-peer (P2P) network for multimedia file sharing [1]. As of February 2013, BitTorrent was responsible for 3.35% of all worldwide bandwidth [2]. However, its illegitimate use on unauthorized sharing of copyright protected files is increasingly rampant. To prevent the illegal distribution of copyright-protected materials through the internet, we had better extract the feature of the content and identify whether it is a copyright-protected content. But first of all, we have to analyze the data downloaded by the BitTorrent client program whether it is enough to extract the correct feature. For the purposes of transfer, files are split into fixed-size pieces which are all the same length except for possibly the last one which may be truncated. In this paper, we proposed a method of audio recognition on a few pieces of MP3 files.

Copyright laws protect musical works from being reproduced and distributed without the copyright owner's permission. MPEG-1 more commonly referred to as MP3, is the most common audio file format used today. Most of audio files transferred through the BitTorrent are MP3 files. The structure of MP3 file consists of a number of independent frames and each frame has a header and side (audio) information which is used to decode the compressed audio data. If there is no header or side information in a piece, it is not only hard to make sure the format of audio file, but also

difficult to find out where to start decoding. In this case, we have to analyze another piece. Thus, header and side information play an important role in process of file analysis. After the compressed data is decoded, then we extract the feature from the audio data.

There are several kinds of musical features can be extracted from audio files, such as timbre, tonality, rhythm and chroma [3]. In this paper, the audio identification algorithm consists of beat tracking and chroma features [4]. Beat tracking is important as a step in understanding how people process temporal information in the editing of audio data. Chroma feature is a powerful and interesting representation for audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones of the musical octave. A beat tracker is used to generate a beat-synchronous representation with one feature vector per beat and the representation of each beat is a normalized chroma vector which sums up spectral energy into 12 bins. To identify the downloaded piece, two beat tracks obtained from original audio file and downloaded piece are matched by using cross-correlation.

The rest of this paper is organized as follows. In Section II, the working principle of BitTorrent technology is presented. In Section III, we give an overview of the structure of MP3 file and point out important information used to decode the piece of audio. In Section IV, we present the audio identification algorithm. Finally, we discuss the evaluation of simulation results in Section V.

## II. BITTORRENT PROTOCOL

BitTorrent is a P2P file distribution protocol which is designed to allow efficient distribution of large files, such as video and audio files. To share a file, BitTorrent splits the file into several fixed size pieces. The size of a piece is usually with byte sizes of a power of 2 and typically between 32 KB and 16 MB. BitTorrent users called peers connect to each other directly to send and receive these pieces. However, there is a central server called tracker coordinates the action of the peers [5], [6]. It acts as an information exchange center and sends a randomly chosen subset of peers who have pieces of the file. The subset of peers is called swarm list. Firstly, there should be at least one file owner with the whole pieces of a file generates a torrent file which contains metadata about the file to be shared and the address of the tracker, and this user is called initial seeder. The torrent file also contains SHA1 hash values of each piece to verify integrity of the pieces. After the torrent file is created, it is registered into a tracker and the file owner places it on websites to make the torrent file available to other BitTorrent users. After downloading the torrent file from the BitTorrent website, the downloader opens it with a BitTorrent client program which

Manuscript received November 29, 2013; revised January 8, 2014.

Xun Jin is with Dept. of Copyright Protection, Sangmyung University, Seoul, 110-743, Korea (e-mail: jinxun@smu.ac.kr).

JongWeon Kim is with Dept. of Intellectual Property, Sangmyung University, Seoul, 110-743, Korea (corresponding author, e-mail: jwkim@smu.ac.kr).

connects to the tracker and manages the transfer of the pieces.

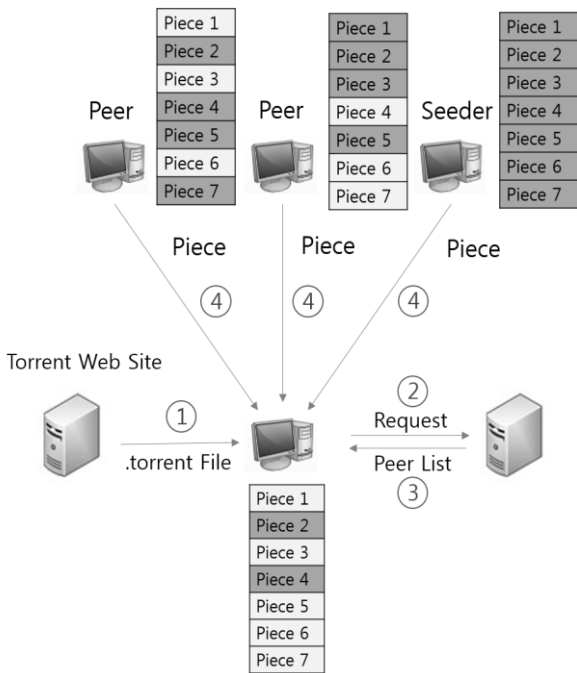


Fig. 1. The working protocol of a BitTorrent network.

The working protocol of a BitTorrent network is shown in Fig. 1.

### III. MP3 FORMAT ANALYSIS

MPEG Audio Compression is one of many methods to compress audio in digital form trying to consume as little space as possible but keep audio quality as good as possible. It is a lossy compression, which means we lose some audio information. But this loss can hardly be perceived because the compression method tries to control it. To decode these compressed data, we must need several parameters to represent compression strength, compression method, compressed data size and location of compressed data. These parameters are allocated in header and side information.

#### A. Frame Header

An MP3 file is built up from a number of small parts called frames. These frames are independent items and each frame has its own header and side information. There is no file header, thus, we can cut any part of MP3 file and play it correctly. But this is not always correct especially for MPEG 1 Layer III files, because frames are sometimes dependent on each other. The parameters of header and side information are not always the same. Variable bitrate MPEG files may use bitrate switching which means the bitrate changes according to the content of each frame to make better compression while keeping high quality of audio. This way lower bitrates can be used in frames and it will not reduce audio quality. The header is constituted by the very first 4 bytes in a frame. The first 12 bits of a header are always set and they are called frame sync. Therefore, we can search through the file for the first occurrence of frame sync. After the frame sync is found, we verify next parameters and check if the values are correct. Then we read the whole header if there is no invalid value. Fig. 2 shows the structure of the header.

Sync		
ID	Layer	Prot_bit
Bitrate		
Frequency	Pad_bit	Priv_bit
Mode		Mode extension
Copy	Original	Emphasis
Side Information and Audio Data		

Fig. 2. Frame header.

#### B. Important Parameters

There are 6 parameters in the header are very important, and they are Version ID, Layer, Bitrate, Frequency, Padding bit and Channel Mode. The practical bitrate is obtained from a bitrate index table with the Version ID, Layer and Bitrate. The values of Version ID and Layer together represent the column and Bitrate represent the row in the table. The frame size is calculated with the bitrate, Frequency and Padding bit. The size of side information depends on Channel Mode. After a piece of a file is downloaded, we read these parameters to verify if the piece is valid. Then side information parameters are used to decode the audio data.

#### C. Piece Verification

The piece length specifies the nominal piece size, and is usually a power of 2. The piece size is typically chosen based on the total amount of file data. If the piece size is too small, resulting in a large torrent metadata file, and piece sizes too large cause inefficiency. The most common sizes are 256 KB, 512 KB and 1 MB. BitTorrent clients download pieces in a random order to increase the opportunity to exchange data. When a peer finishes downloading a piece, the client checks that the hash matches, then passes the piece to the piece verification module. It is almost impossible to split MP3 file from the beginning of the header, which means the beginning of the piece may be frame data, side information or header. Thus, when a piece is downloaded, we begin to search the header immediately. If there is an available header, we decode the piece and begin the process of audio identification, otherwise we have to analyze next piece. If the next piece is just next to the pieces already downloaded, the probability of successful identification will be increased.

### IV. AUDIO IDENTIFICATION

#### A. Beat Tracking

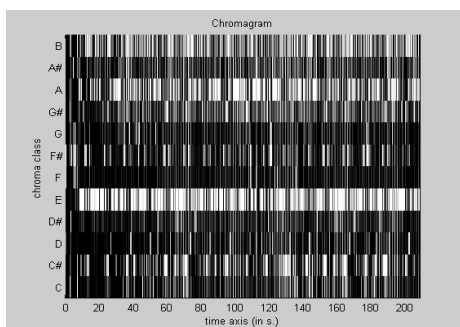
Beat tracking is an important initial step in computer emulation of human music understanding, since beats are fundamental to the perception of Western music. Even though people cannot completely identify every audio component, we can track musical beats and keep time to music by foot-tapping. Therefore, we can build a computational model of beats and track the beats [7], [8].

The first stage of beat tracking converts the audio into a one-dimensional function of time at a lower sampling rate

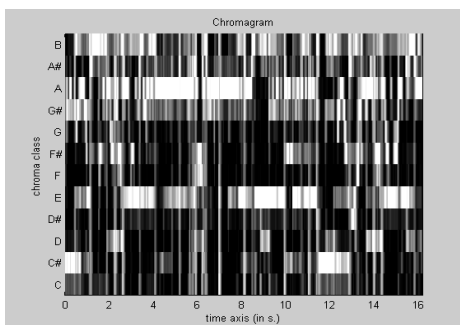
which reflects the strength of onsets by taking the first-order difference along time in a log-magnitude 40-channel Mel-frequency spectrogram. Discard negative values and sum across frequency. Then slowly varying DC offsets are removed by a high-pass filter. The onset strength for the entire signal is auto-correlated to estimate an approximate global tempo. Tempo is a pace reference that typically ranges from 40 to 260 BPM (beats per minute) with a mode roughly around 120 BPM. The best BPM is passed to the beat tracking module which is implemented with dynamic programming. It attempts to find a sequence of beat times that optimize both the onset strength at each beat and the spacing between beats.

**B. Chroma Features**

Useful musical information can be obtained from the distribution of chroma even without the absolute frequency. Chroma features consist of a 12 element vector with each dimension representing the intensity which is associated with a particular semitone. Record a single feature vector per beat and 12 element chroma features are used to capture both dominant note and the broad harmonic accompaniment. To identify strong tonal components in spectrum and get a higher resolution estimate of underlying frequency, use phase derivative within each FFT bin. Each recording is represented by a matrix of 12 chroma dimensions. Fig. 3 shows chromagrams of an audio recording and one piece of it. Finally, the identification is implemented by cross-correlating the two feature matrices of piece and original.



(a) Chromagram of an audio recording.



(b) Chromagram of one piece.

Fig. 3. Chromagrams of an audio recording and one piece of it.

**V. EVALUATION**

The most common types of copyright infringement of audio on BitTorrent are pop songs being used illegally. Thus, we evaluate the proposed method with 30 pop song samples and sample file size is between 2 MB and 5 MB. Suppose the piece size is 256 KB, the playback time of the piece will be 16

seconds with the bitrate of 128 Kbps. But in fact, the bitrate of audio file on BitTorrent is unpredictable, which means it depends on how the MP3 file owner compress the audio, thus, the bitrates of samples were not fixed. Generally, the larger piece size is the better result we get. Fig. 4 shows the probabilities of successfully identifying pieces of 30 samples against different piece sizes. In this test, each sample was identified with only one piece of it. The results show if the piece size is less than 256 KB, it is difficult to identify the piece correctly. It also means if the playback time of an audio is less than or equal to 8 seconds, the audio will be unrecognizable. There are several samples were encoded with bitrate of 256 Kbps, therefore, the playback time of pieces with size of 256 KB from those samples is equal to 8 seconds. This kind of phenomenon will have a negative impact to the probability and needs further research.

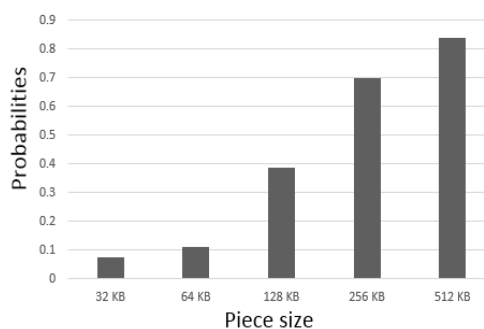
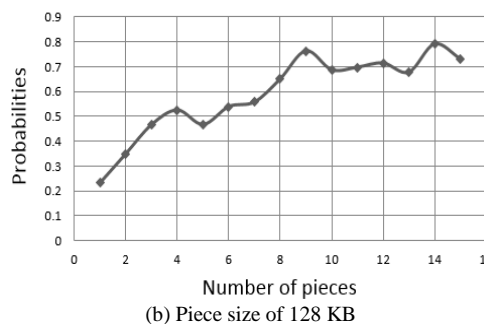


Fig. 4. Probabilities against different piece sizes.

To analyze the relationship between number of pieces and probability, we also evaluate probabilities of successful identification with different numbers of pieces randomly ordered. Fig. 5 shows the probabilities of successful identification against different number of pieces with fixed piece sizes. In the case of piece size being equal to 64 KB and 128 KB, all the samples are evaluated 15 times, whereas 8 times for 256 KB, and each time the number of pieces is increased by 1. Obviously, the results show the more pieces we have, the better identification results we can obtain.



(a) Piece size of 64 KB



(b) Piece size of 128 KB

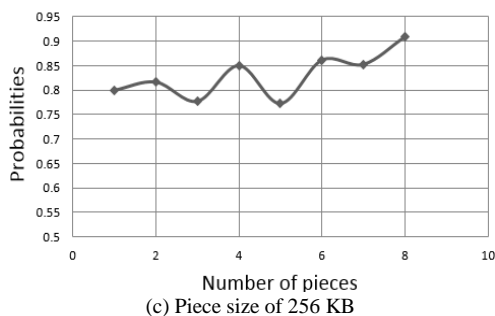


Fig. 5. Probabilities against different number of pieces with fixed piece sizes.

As shown in Fig. 5 (a), it is difficult to correctly identify the pieces, especially when the pieces are less than 12, the probability drops below 60%, and most of probabilities drop below 50%. Fig. 5 (b) shows better results than (a), but there are still half of probabilities drop below 60%. Fig. 5 (c) shows the evaluation results of piece size of 256 KB, and all of them are greater than 77% as well as half of them are greater than 85%. When the piece size is equal to 512 KB, the probability will be greater than 90% if we have more than 1 piece. As mentioned above, the most common piece sizes used on BitTorrent are 256 KB, 512 KB and 1 MB, and our method shows good performances on these sizes.

## VI. CONCLUSION

In this paper, we proposed a method of audio identification on a few pieces of MP3 files downloaded by BitTorrent client program. The pieces split by torrent file are randomly ordered, and passed to the piece verification module to verify if the piece is valid by checking the important parameters of frame header. Then we decode the piece and extract the feature to identify the piece. The experimental results show the probabilities of successfully identifying the pieces with different piece sizes and different number of pieces. When the piece size is greater than or equal to 512 KB, the probability of successful identification will be greater than 90%. Because of the unpredictable bitrate of MP3 file, the playback time of audio decoded from pieces of same size is variable. It increases difficulty of audio identification and needs to be further discussed in the future research.

## ACKNOWLEDGMENT

This research project was supported by Ministry of Culture, Sports and Tourism (MCST) and from Korea Copyright Commission in 2013.

## REFERENCES

- [1] P. Sharma, A. Bhakuni, and R. Kaushal, "Performance analysis of BitTorrent protocol," in *Proc. 2013 IEEE National Conference on Communications*, 2013, pp. 1-5.
- [2] An torrent app: BitTorrent. [Online]. Available: <http://www.warer.com/windows/apps/bittorrent/>
- [3] O. Lartillot and P. Toivainen, "A Matlab toolbox for musical feature extraction from audio," in *Proc. the 10<sup>th</sup> Int. Conference on Digital Audio Effects*, September 2007, pp. 10-15.
- [4] D. P. W. Ellis and G. E. Poliner, "Identifying 'cover songs' with chroma features and dynamic programming beat tracking," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007, vol. 4, pp. IV-1429-IV-1432.
- [5] K. Bauer, D. McCoy, D. Grunwald and D. Sicker, "BitStalker: accurately and efficiently monitoring bittorrent traffic," in *Proc. First IEEE International Workshop on Information Forensics and Security*, 2009, pp. 181-185.
- [6] K. P. Chow, K. Y. Cheng, L. Y. Man *et al.*, "BTM - an automated rule-based bt monitoring system for piracy detection," in *Proc. Second IEEE International Conference on Internet Monitoring and Protection*, 2007, pp. 2.
- [7] M. Goto and Y. Muraoka, "A real-time beat tracking system for audio signals," in *Proc. the International Computer Music Conference*, 1995, pp. 171-174.
- [8] T. Jehan, "Creating music by listening," PhD thesis, MIT Media Lab, Cambridge, MA, 2005.



**Xun Jin** received his B.S. degree in computer science and technology from Fujian Agriculture and Forestry University, China, in 2011. He is currently pursuing the Ph.D. degree in copyright protection, Sangmyung University, Korea. His research interests are digital image/video watermarking, multimedia forensics, digital signal processing, and information security.



**Jong Weon Kim** received the Ph.D. degree from University of Seoul. He majored in signal processing in 1995. He is currently a professor of Dept. of Intellectual Property at Sangmyung University in Korea. He has a lot of practical experiences in the digital signal processing and copyright protection technology at the institute, the industry, and college. His research interests are in the areas of copyright protection technology, digital rights management, digital watermarking, and digital forensic marking.