Heavy-Uploader Tracking System Design in BitTorrent Environment

Jihah Nah and Jongweon Kim

***Abstract*—Copyright infringement using BitTorrent has become a major problem in the creative industries; however, no reliable technical means are available to prevent piracy. The packet filtering method, which is used to inspect network traffic and to detect problems, puts a heavy strain on network resources. At the same time, there is a big controversy over the privacy in communications and the authority to investigate. This paper proposes a tracking system to trace a heavy uploader, who illegally shares a significant amount of copyrighted contents through BitTorrent. The proposed tracking system analyzes the seed file of an illegal torrent, acquires the swarm tracker's address, and retrieves the IP address or MAC address of a peer from the swarm list, which is a list of peers sharing copyrighted contents from the tracker. Interception and/or inspection of network packets with personal information can be an offense against privacy laws. However, the proposed heavy-uploader tracking method avoids these legal issues by gathering information about content-sharing peers through a client program. A heavy uploader can be determined using an acquired identifying address and a list of copyrighted content.**

***Index Terms*—BitTorrent, copyright infringement, heavy uploader, tracking system.**

I. Introduction

As the manipulation of digital content becomes easier, copyright infringement becomes a serious problem. Mp3 distribution sites, including mp3.com and Napster, were closed or bought by other companies as a result of lawsuits, because peer-to–peer (P2P) file-sharing technology was blamed for enabling copyright infringement [1], [2]. In Korea, Soribada transformed their representative P2P music service into a service platform for copyright management in response to the outcome of a lawsuit [3]. Since then, the use of P2P services decreased; however, copyrighted contents can still be acquired easily using online storage services, such as Webhard.net. In Korea the Electric Communication Law, enacted in May 2012, required service providers to be registered; therefore, interest in P2P services is rising again.

BitTorrent is a P2P file-sharing technology developed by Bram Cohen in 2001. BitTorrent provides file-sharing capability at high speed using 1-to-*n* peer connections, unlike previous P2P technologies, which used 1-to-1 peer connections. BitTorrent is a highly innovative technology for distribution of digital content, and copyright owners have blamed it for enabling infringement, because it allows the sharing of digital content with unspecified individuals. Furthermore, files can be transferred effectively although under flash crowd and free-rider can be prevented efficiently. For this reason, the use of BitTorrent is rapidly growing.

Online storage services, such as Webhard, combine technical copyright protection with various filtering technologies; P2P services, like those using BitTorrent, do not provide the same protection. Therefore, it diminishes the effectiveness of Webhard service provider's efforts to transition to a legal service.

In this paper, we propose a heavy uploader tracking algorithm for copyright protection in a BitTorrent environment because heavy uploaders need to be purged from the distribution system for copyrighted contents. The proposed algorithm can trace a heavy uploader using information about the peers and the contents from the tracker server, which participates in the operations of BitTorrent and manages the lists of peers.

II. BitTorrent

BitTorrent is one of the most common P2P protocols for distributing large files over the Internet. In February 2009, P2P networks accounted for roughly 43% and 70% of all Internet traffic in the US and Eastern Europe, respectively [4]. Although this figure has now fallen below 13% in the US, P2P traffic continues to rise in Europe and Asia.

The first BitTorrent version was released in July 2001 [5], and it is maintained by BitTorrent, Inc. BitTorrent clients are available for several computing platforms.

As of January 2012, BitTorrent had 150 million active users. Based on this, the current number of monthly BitTorrent users could be more than a quarter billion [6]. On average, it has more active users than YouTube and Facebook combined [7], [8].

The BitTorrent protocol reduces the impact of distributing large files on both the server and the network. Instead of downloading a file from a single source server, BitTorrent allows users to join a "swarm" of hosts who simultaneously download and upload from each other. The protocol is an alternative to the older single-source-multiple-mirrors technique of distributing data. In addition, it can work over networks with lower bandwidths, so that many small computers, like smart phones and pads, can efficiently distribute files to many recipients.

P2P network like BitTorrent makes session directly between clients. The file to be distributed is divided into small pieces. As each peer receives a new piece of the file, it becomes a source (of that piece) for other peers, relieving the original seed from having to send that piece to every peer

Manuscript received August 25, 2013; revised December 17, 2013. This work was supported by Ministry of Culture, Sports and Tourism (MCST) and from Korea Copyright Commission in 2013.

Jihah Nah is with the Copyright Protection Research Institute, Sangmyung University, Korea (e-mail: jihah.nah@gmail.com).

Jongweon Kim is with the Dept. of Intellectual Property, Sangmyung University, Korea (e-mail: jwkim@smu.ac.kr).

requiring a copy. With BitTorrent, the task of distributing the file is shared by those who want it; it is possible for the seed to send only a single copy of the file, which is eventually distributed to an unlimited number of peers.

The pieces of the file are typically downloaded randomly and rearranged into the correct order by the BitTorrent client. The client also monitors which pieces it has and which it needs. It can transmit the pieces it has to other peers. Throughout a single download, all file pieces are of the same size. For example, a 10 MB file may be transmitted as ten 1MB pieces or as forty 256KB pieces.

Fig. 1 depicts the download process. Once the user is connected to the swarm, they can download available pieces from several peers (seeder and leechers) simultaneously. This mechanism improves the download speed.

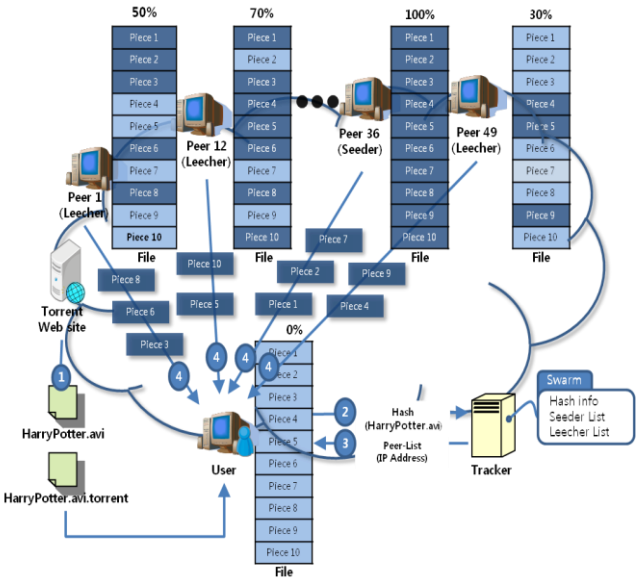

Fig. 1. BitTorrent download process.

III. Entities in BitTorrent [9], [10]

A. Peer

A peer is an entity that executes a BitTorrent client program and transmits copyrighted content to other clients via the internet. A peer generally does not have the complete content but only a small part. A peer joins in swarm.

B. Swarm

A swarm exists for every shared file. It holds its identifier and the information about the list of peers interested in the same shared file. It is managed by a tracker.

C. Seeder

A seeder is a peer with the complete content. When a leech obtains all pieces of the content, that peer becomes a seed. Seeding refers to making a peer's connection available for other peers. Normally, a peer seeds more data than a leech. However, whether to seed or not, or how much to seed, depends on the availability of leeches and the choice of the seeder.

D. Leecher

A leecher is typically a downloader who does not have the complete content. Leeches might be on asymmetric internet connections or might not leave their BitTorrent client open to seed the file after their download has completed. However, some leeches intentionally avoid uploading by using modified clients or by excessively limiting their upload speed.

E. Indexer

An indexer aggregates torrent data from one or more tracker sources into a central database, which is publicly or privately searchable, and provides the metadata contained in the .torrent file, such as the tracker URL.

F. Tracker

A BitTorrent tracker is a server that assists in the communication between peers using the BitTorrent protocol. Clients are required to communicate with the tracker to initiate downloads; they also communicate with the tracker periodically to negotiate with newer peers and to provide statistics; however, after the initial transfer of peer data, peer communication can continue without a tracker.

G. Public Tracker

Public or open trackers can be used by anyone by adding the tracker address to an existing torrent, or they can be used by any newly created torrent. The most popular public trackers are OpenBitTorrent, PublicBitTorrent, and Coppersurfer. The Pirate Bay website formerly ran a tracker, which is now disabled [1], although the website continues to serve as a torrent index.

H. Private Tracker

A private tracker restricts use by requiring users to register with the site. To control registration, many private trackers use an invitation system, in which active and contributing members are given the ability to grant a new user permission to register at the site [2]. Most private trackers monitor how many users upload or download content and, in most situations, enforce a minimum upload-to-download ratio.

Some of the advantages of a private tracker versus a public one are higher speeds, a tighter community, and safer downloads. Almost all private trackers implement a passkey system, where each user is given a personalized announce URL, so that, if a distribution is unauthorized, it can be pinpointed to the responsible user. An example of a private tracker is Oink's Pink Palace, which was forcibly shut down in late 2007 by law enforcement officials.

Some members of a closed community might find it hard to maintain the required ratio. Competitions might be offered with prizes of improved ratios. Some trackers exchange ratio credit for donations, and others use "free leech" systems to improve the user's ratios. When downloading a free leech torrent, only the upload is logged, the download is ignored. Large torrents are usually offered as free leech. For the sake of internet etiquette, seeding to a ratio of at least 1 is still recommended even for free leech torrents.

IV. Information-Gathering Mechanism Design for Tracking a Heavy Uploader

Fig. 2 depicts the proposed system configuration for recognition and tracking of BitTorrent. Torrent server has

torrent meta files and peers share the meta files and contents. Peers seek a seed file (torrent meta file) to find the infringing content in the internet. One of peers is a seeder who makes a seed file and upload it to a torrent server and others are leeachers who wants to download the same content. The information of relevant peers is on a swarm list. The crawler system gathers the seed files from torrent servers in the internet and seed analyzer analysis the seed file for tracking heavy uploaders. After the analyzer analyses the seed file, then it accesses to a tracker, downloads a swarm list and stores the extracted peer information into the peer DB. The peer DB can output peers as heavy uploaders who has contents more than certain amount.

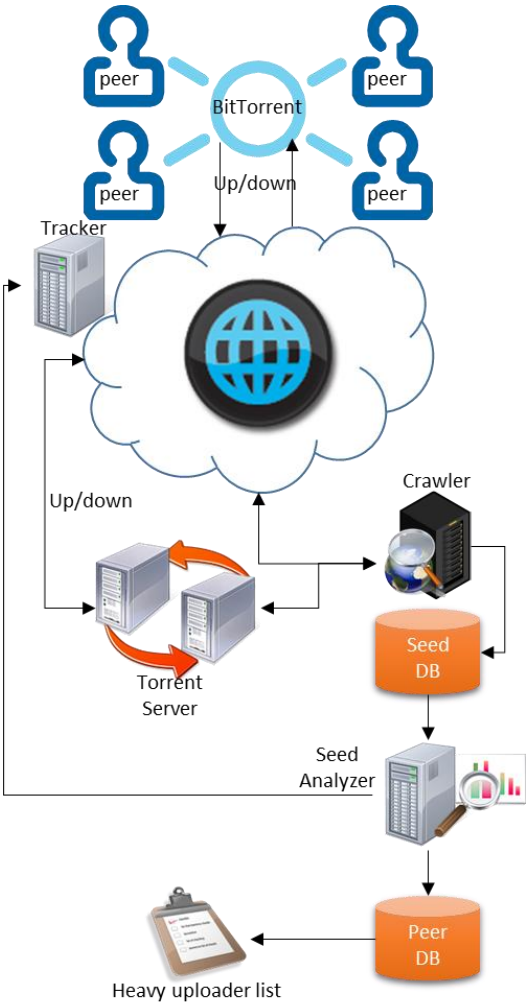

Fig. 2. The proposed system configuration for tracking of BitTorrent.

A tracker performs the most important role in a BitTorrent operation. It manages a swarm list, which is a list of seeders and leechers that share the content specified in the torrent file. Fig. 3 describes messages used in communication between a peer and a tracker. Communication is started to transmit hash information and peer ID as shown in Fig. 3.

Hash values are configured for directory, piece length, private and filename. After confirming hash values, a tracker transfers information about peers and the file location for downloading and uploading. The swarm list can be obtained by connecting to a tracker when a peer runs a client program using the BitTorrent seed file as shown in fig. 4. The tracker gives peers list on a relevant swarm to a client based on hash information given by the BitTorrent client. The BitTorrent specification states that the swarm list can include up to 50 peers; however, a list of 200 peers is actually delivered if a client program is running.

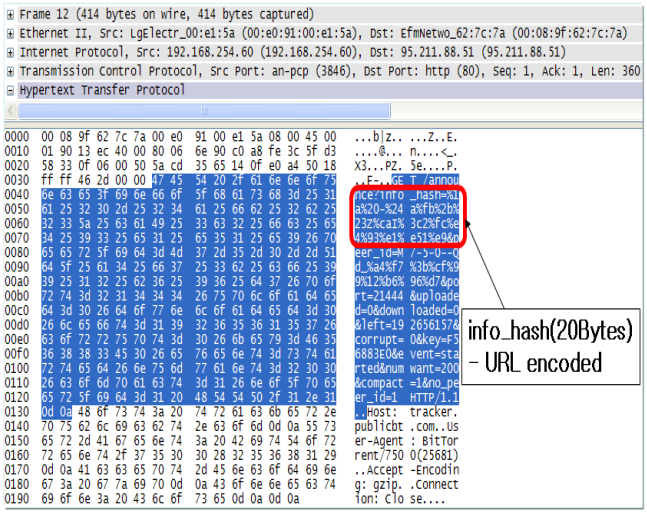

Fig. 3. Hash information in seed analyzer.

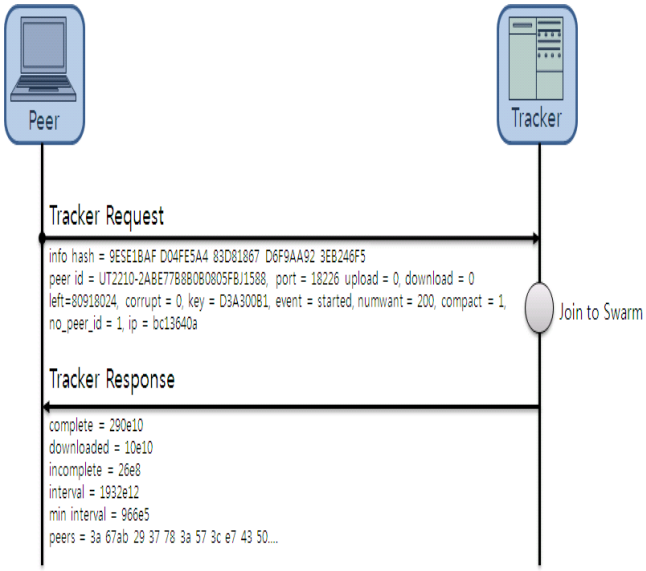

Fig. 4. Communication protocol between peer and tracker.

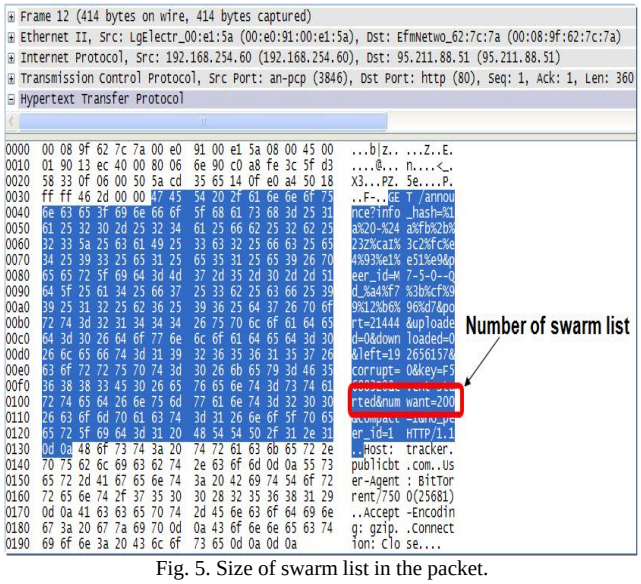

Fig. 5. Size of swarm list in the packet.

Fig. 5 shows the requested number of peers in a swarm list, as detected in a packet.

The swarm list provides each peer's 4-byte IP address and a port number for that IP. Each tracker informs the numbers of involved seeds and peers and current state of a client. This information can be obtained by analyzing a packet containing a swarm list that is exchanged between a peer and a tracker, as shown in Fig. 6.

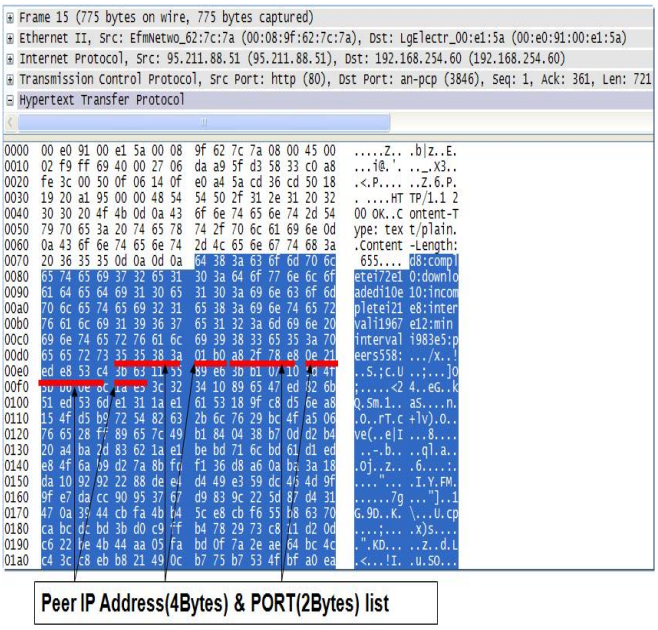

Fig. 6. Peer IP address and port number in the packet.

Fig. 7 shows this mechanism can track MAC address between peers when a tracker communicates with peers by UDP. The method for IP tracking can be complemented by tracing MAC address of the source and the destination.

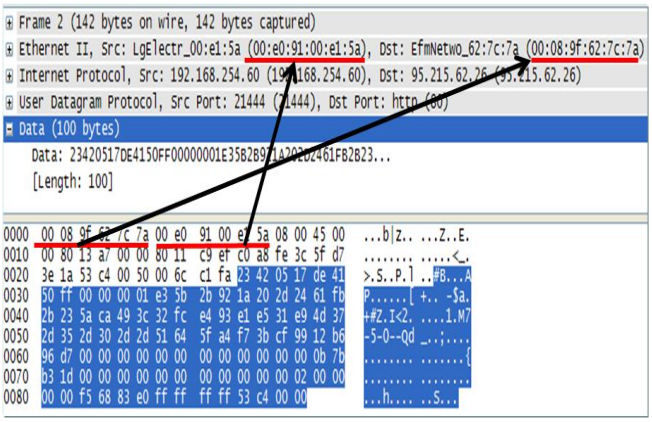

Fig. 7. Peer IP address and port number in the packet.

A swarm list can also be used to track a heavy uploader, and a co-owner of the same content can be identified. If the swarm lists for several contents are collected, we can determine how much content is shared by the same peer based on the IP address. Thus, we can identify and track a heavy uploader. Fig. 8 depicts the flow of a tracking mechanism.

The BitTorrent seed file contains the hash value of the content and a list of tracker servers that can be used to get a swarm list. We acquired swarm lists from these tracker servers. If a peer in an acquired list was in our existing peer database, the appropriate count was increased. If not, the new peer was registered. Additional swarm lists were collected from all tracker servers listed in each new seed file. The database of collected peers was analyzed and processed for statistics after being divided into two groups: a list of peers sorted by content and a list of contents sorted by peer.

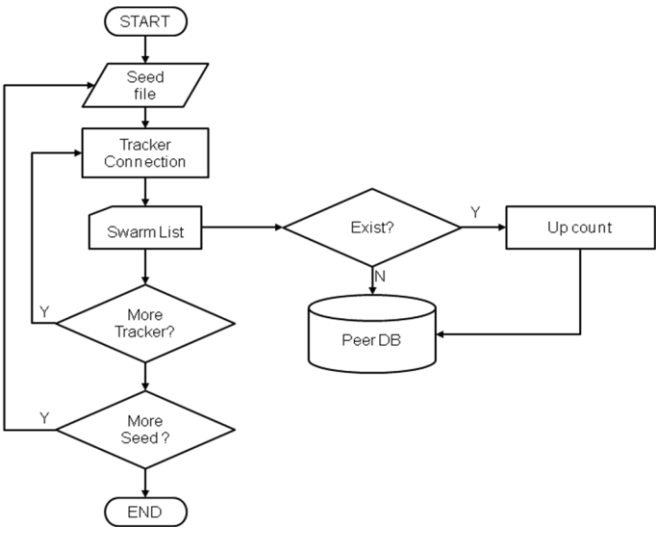

Fig. 8. Tracking mechanism for heavy uploaders.

V. Conclusion

Copyright infringement occurs more frequently in BitTorrent compared to the Webhard service, which is expanding in Korea. A movie with 2GB can be downloaded within 20 minutes through the BitTorrent protocol, which transmits content using a 1-to–N P2P design. After a movie is downloaded, the copyright infringement occurs when it is uploaded. A practical technology for shutting down illegal distribution is difficult to use because of BitTorrent. Although a method that checks and cuts packets on the network can be used, the cooperation of internet service providers (ISP) is necessary, and related laws have to be reviewed.

BitTorrent trackers provide information to receive and send copyrighted works between peers, but, they do not directly save or transfer any copyrighted works themselves. The contents are transferred directly from each peer's computer without intervention of the tracker itself. Nevertheless, many torrent trackers have been the target of lawsuits from copyright owners [11]. Actually, there is more complicated factor that includes a variety of countries and copyright laws.

Interception and/or inspection of network packets with personal information can be an offense against privacy laws. However, the proposed heavy-uploader tracking method avoids these legal issues by gathering information about content-sharing peers through a client program.

Because the proposed method distinguishes content-sharing peers using only the IP address or MAC address, it can discriminate a machine, but not person. The person can be identified with the proper procedure, because some evidence regarding the copyright infringement can be obtained from the specified machine. Copyright infringement through BitTorrent could decrease using the proposed method.

ACKNOWLEDGMENT

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

REFERENCES

[1] The Day the Napster Died. (Aug. 5 2013). [Online]. Available: http://www.wired.com/gadgets/portablemusic/news/2002/05/52540?currentPage=all

[2] MP3.com. [Online]. Available: https://en.wikipedia.org/wiki/MP3.com visited Aug. 5 2013.

[3] Lawsuit against Soribada the prosecution embarassment recoding industry 2000 billion won per pear from illegal copies. (May 21, 2001). The Korea Economic Daily. [Online]. Available: http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=102&oid=015&aid=0000376725

[4] H. Schulze and K. Mochalski, "Peer-to-peer file sharing (P2P) still generates by far the most traffic in all monitored regions–ranging from 43 percent in Northern Africa to 70 percent in Eastern Europe," *Internet Study 2008/2009*, ipoque, Leipzig, Germany 2009.

[5] B. Cohen, *BitTorrent-a new P2P app*, Yahoo eGroups, 2001.

[6] Bittorrent.com, *BitTorrent and µTorrent Software Surpass 150 Million User Milestone*, 2012.

[7] Fastcompany.com. (Aug. 5, 2013). [Online]. Available: http://www.fastcompany.com/

[8] Comscore.com. (2010). [Online]. Available: http://www.Comscore.com/

[9] Glossary of BitTorrent terms. (Aug. 5 2013). [Online]. Available: http://en.wikipedia.org/wiki/Terminology_of_BitTorrent visited

[10] BitTorrent tracker. [Online]. Available: http://en.wikipedia.org/wiki/BitTorrent_tracker visited Aug. 5 2013.

[11] J.-P. Sarah," Copyright Trolls: 200,000 BitTorrent Users Sued Since 2010," *PC World*, 2011.

Jihah Nah received an MS degree in electronics engineering from the University of Seoul. She is currently a graduate student in the Electrical and Electronic Engineering Department at Yonsei University, Korea. From 1993 to 2003, she was a member of the research staff of the ATM Switching Lab at ETRI. She was manager of the Embedded Software Team at KIPA from 2003 to 2008. Her research interests are in the areas of copyright protection, image processing, and embedded software development.

Jongweon Kim received a Ph.D. degree from the University of Seoul with a major in signal processing in 1995. He is currently a professor at the Department of Intellectual Property at Sangmyung University in Korea. He has considerable practical experience in digital signal processing and copyright protection technology in the institutional, industrial, and academic environments. His research interests are in the areas of copyright protection technology, digital rights management, digital watermarking, and digital forensic marking.