International Journal of Security and Its Applications
Vol.7, No.6 (2013), pp.191-198
http://dx.doi.org/10.14257/ijsia.2013.7.6.20

# Piracy Tracking System of the BitTorrent

Jungjae Lee[1] and Jongweon Kim[2]

[1]*Dept. of Entertainment Business, Soongsil Cyber University, Korea*
[2]*Dept. of Intellectual Property, Sangmyung University, Korea*
[1] *jjlee@mail.kcu.ac.kr,*[2]*jwkim@smu.ac.kr (Corresponding Author)*

### *Abstract*

*In this paper, we investigated the problem of copyright infringement in the peer-to-peer (P2P) environment, and proposed a system for piracy tracking. Even though the BitTorrent P2P software is a very efficient way to transfer files, it has become the main tool used to pirate copyrighted materials. Thus, we studied the bit torrent packet itself in order to detect the destination and source of the piracy. The torrent meta-file which has hash values of the pieces of the shared content is also analyzed to identify copyrighted material. As a result, we proposed a tracking system for copyright protection against BitTorrent piracy. Our system consists of hash database, feature database, protocol analyzer, torrent meta-file crawler, and copyright identifier. In this tracking system, the content is identified by three ways. The first way is to use the info hash in the hash DB. The second way is to use the piece hash value in the hash DB. Last way is to use the feature of the content that is able to identify if the content is slightly changed from original. If the proposed system is installed on the ISP network, it is able to protect the content from BitTorrent distribution.*

***Keywords:*** *Piracy Tracking, P2P, BitTorrent, Copyright Infringement, Torrent Meta File*

## 1. Introduction

As computer technology improves and the Internet rapidly becomes ubiquitous, new online services such as web storage and peer-to-peer (P2P) protocols have enabled content distribution. Although these services are widespread because they are useful and convenient for users, they are also blamed for hampering the copyright industry, creating illegal revenue that excludes the copyright owner's share. This is because digital content is easy to copy and distribute via the Internet.

Early in August of 2011, the Seoul District Prosecutors Office announced that it had been investigating whether some web storage service providers were distributing illegal content. According to the announcement, two companies with annual sales of US$25 million and US$15 million had the same owner. It was claimed that they operated a specialized company for uploading illegal content independently from the service companies and swindled copyright owners out of license fees (about US$15 million)[1].

In July of 2013, Contra Piracy, a claimed non-profit group, said they had monitored 2,919 individuals infringing the movie on more than 280,000 occasions. In order to stop these infringements they need the identities of the file-sharers from ISPs. As usual, Swiss-based Contra Piracy isn't the creators of the movie. Instead the outfit obtained "enforcement rights" from Los Angeles-based Hannibal Pictures to pursue the action. With around US$8 million in settlements potentially on the table, it was certainly a deal worth doing [2].

While piracy techniques are becoming more intelligent and advanced, copyright protection technologies and tools for acquiring digital evidence are lagging behind. Therefore, it is

ISSN: 1738-9976 IJSIA

important to study the prevention of copyright infringement on P2P software such as BitTorrent. In particular, to prevent large-scale damage from continued illegal distribution, we must develop a system that can quickly block unauthorized downloads.

Section 2 introduces BitTorrent service, and Section 3 presents an analysis of the BitTorrent protocol. Some counterstrategies against BitTorrent piracy and our proposed tracking system for copyright protection are described in Section 4. Finally, Section 5 contains our conclusions.

## 2. BitTorrent Service

P2P systems allow users to download media files, such as music, movies, and games, using a P2P software client that searches for content on other computers [3]. The “peers” are computer systems connected to each other through the Internet. Thus, the only requirements for a computer to join a P2P network are an Internet connection and P2P software [4]. The first generation of P2P software was the Napster system, a central server-based model that was eventually shut down. The second generation of P2P software included Gnutella and Kazaa, which are user-based models, and BitTorrent is part of the third generation of P2P networks. The difference from previous systems is that BitTorrent creates a new network for every set of files, rather than trying to create one big network of files using SuperNodes, web caches, or servers [5].

BitTorrent is one of the most common P2P protocols for distributing large files over the Internet. It has been estimated that P2P networks accounted for roughly 43% and 70% of all Internet traffic in the US and east Europe, respectively, in February 2009 [6]. Although this figure has now fallen below 13% in the US, traffic continues to rise in Europe and Asia.

Programmer Bram Cohen created the BitTorrent protocol in April 2001, and released the first available version in July of that year [7]. It is now maintained by Cohen’s company, BitTorrent, Inc., which offers numerous BitTorrent clients for a variety of computing platforms.

According to BitTorrent, Inc., BitTorrent had 150 million active users as of January 2012. Based on this, the total number of monthly BitTorrent users can be estimated at more than a quarter billion [8]. Indeed, at any given instant, BitTorrent has, on average, more active users than YouTube and Facebook combined [9, 10]. BitTorrent is widely used to transfer files [11].

The BitTorrent protocol can be used to reduce the impact of distributing large files on both the server and the network. Rather than 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 mirror-sources technique of distributing data. In addition, it can work over networks with lower bandwidth, so many small computers, like smartphones and pads, are able to efficiently distribute files to many recipients.

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

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

 

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

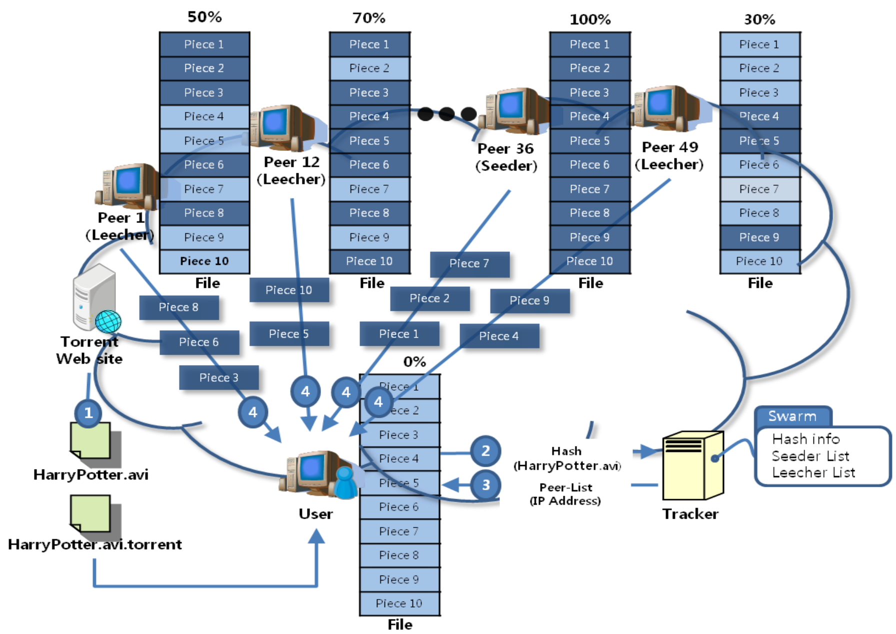


Figure 1. BitTorrent Download Process

## 3. Analysis of BitTorrent Protocol

To download content, users must first download the torrent meta-file. The meta-file extension is ".torrent," and the hash value of the file is formed from the directory and piece length, as well as the private status and filename, in the metadata.

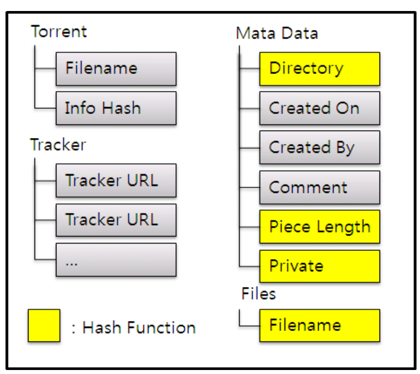


Figure 2. Structure of the Torrent File

In order to analyze the BitTorrent packet, we investigated transport packets using various analysis methods after the torrent client downloaded some content. First, we configured the BitTorrent client to use port 21444 as the default download port.

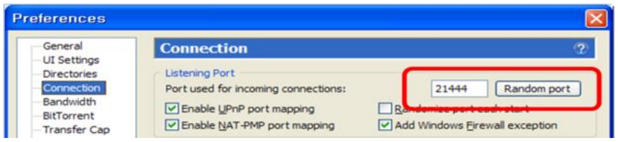


**Figure 3. Port Configuration**

Using the torrent editor, we confirmed information about the torrent file. In this case, the hash info was made up of four hexadecimal parameters beginning with “1A202D24” and ending with “E1E531E9.” Torrent tracker server manages the peers and classifies the content with the info hash.


**Figure 4. Info Hash**

The torrent meta-file also has a lot of hash values which can identify the piece of the content. Figure 5 shows hexadecimal value of the meta-file. As shown in the file, the content has 10780 pieces and the following values are has values of the pieces. If the network protocol analyzer catches a packet of one piece, we can identify what the piece means exactly what the content is. This reduces the identification time and it is helpful blocking the piracy content as fast as possible.

```
00000310  4D 61 6E 20 33 20 32 30 31 33 20 37 32 30 70 20  Man 3 2013 720p
00000320  52 36 20 78 32 36 34 20 41 43 33 2D 4A 59 4B 31  R6 x264 AC3-JYK1
00000330  32 3A 70 69 65 63 65 20 6C 65 6E 67 74 68 69 34  2:piece lengthi4
00000340  31 39 34 33 30 34 65 36 3A 70 69 65 63 65 73 31  194304e6:pieces1
00000350  30 37 38 30 3A E9 49 DA 06 C0 AB 9E 77 DF 90 55  0780:éIÚ.À«žwß.U
00000360  7D 77 31 C7 0A EB 37 1E 95 49 6D CE D9 FB E5 FE  }w1Ç.ë7.•ImÎÙûåþ
00000370  D4 D4 E7 0D EB 78 14 C1 67 53 3E CB 72 52 D6 13  ÔÔç.ëx.ÁgS>ËrRÖ.
00000380  46 7D 6C 19 AF F6 2E AF 3A 75 D3 C3 F8 5C A5 DF  F}l.¯ö.¯:uÓÃø\¥ß
00000390  BD 62 65 9F 77 A7 BB 2B A0 96 85 FF 2D C5 19 51  ½beŸw§»+ –…ÿ-Å.Q
000003A0  FB 96 39 A9 3A 59 FD 4C 6F 03 5E 20 20 7E 6F 34  û–9©:YýLo.^  ~o4
000003B0  EC 3F DD DC 18 7D 82 E5 3A 0E 71 AA 44 F7 4F 82  ì?ÝÜ.},å:.qªD÷O,
000003C0  97 50 C1 14 84 97 F8 CF 52 CD C3 F6 76 23 AA C2  —PÁ.„—øÏRÍÃöv#ªÂ
000003D0  CE 57 B2 46 A1 27 3E 04 7C 48 2F 04 B8 5C E5 E7  ÎW²F¡'>.|H/.¸\åç
000003E0  F4 5D E5 CB C8 26 16 80 C5 05 83 C9 EB 69 AC 1B  ô]åËÈ&.€Å.ƒÉëi¬.
000003F0  BE 0D 7F 53 97 4C D8 E6 72 42 7C 83 B3 79 81 F3  ¾..S—LØærB|ƒ³y.ó
00000400  44 DD E5 02 38 2A 67 05 2F 14 7A 87 0C 72 3B 44  DÝå.8*g./.z‡.r;D
00000410  96 74 15 89 90 37 0D 82 34 67 2B A4 00 46 AA 37  –t.‰.7.‚4g+¤.Fª7
00000420  C0 BA 13 FB F5 56 EC E0 07 75 3D B1 14 5B 41 CF  Àº.ûõVìà.u=±.[AÏ
00000430  6F EF A0 45 A1 BB B4 B7 44 90 92 F9 5C C1 CB 76  oï E¡»´·D.’ù\ÁËv
00000440  50 B7 F6 A1 F2 76 DE A2 5E 68 66 BE 48 AC 7E 74  P·ö¡òvÞ¢^hf¾H¬~t
```

**Figure 5. Hexadecimal Expression of the Meta-file**

After confirming the hash value, the BitTorrent client connected to the tracker and received the information about shared content. The packet sent between a peer and the tracker uses the http protocol. The WireShark open-source software allows us to watch and analyze internet packets. In this packet, we can get the info_hash value and then we can identify the peer tries to share which content.

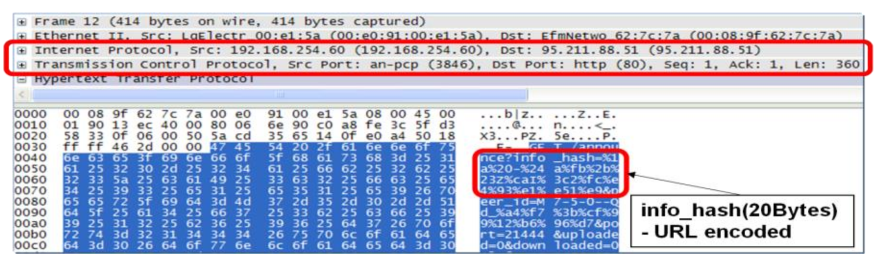


**Figure 6. Hash Info in the Packet**

 

In the next procedure, the tracker sends the peer a list of IP(Internet Protocol) addresses that have content available. This list is known as swarm list of peers who have complete content or incomplete content.

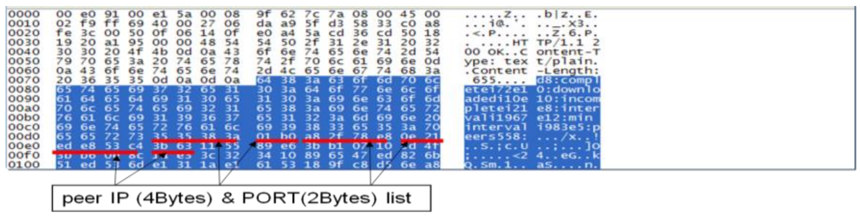


**Figure 7. Peer IP Address and Port List in the Packet**

Figure 8 shows the MAC(Media Access Control) address when the tracker and peer communicate using UDP(User Datagram Protocol).

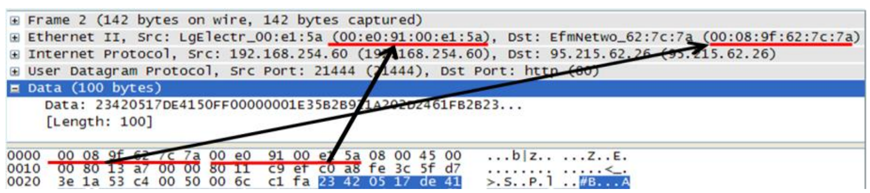


**Figure 8. MAC Addresses in the Packet**

After communication with the tracker is completed, P2P communication commences. This is the handshake procedure, in which the hash value is exchanged. At the moment of the handshake, WireShark captures the hash value, peer ID, source IP, and destination IP.

Once a channel between the peer and a seeder or leecher is opened, the peer asks the seeder or leecher for a piece of the file that the peer requires. If someone has the piece, the peer receives it from the seeder or the leecher. This operation occurs simultaneously for all required pieces, which is why BitTorrent is an extremely fast method of downloading a file. Figure 9 shows the exchange protocol between peers.

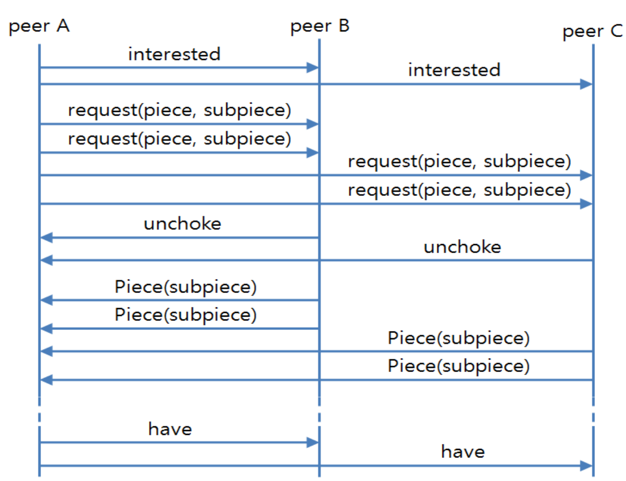


**Figure 9. Exchange Protocol of Bittorrent**

 

## 4. Counterstrategies and the Tracking System for Copyright Protection

Five anti-piracy strategies were announced on torrentfreak.com in 2011 [10]. These are listed in Table 1.

**Table 1. Five Anti-Piracy Strategies**

| Strategy | Status |
|---|---|
| Domain seizures | MPAA (Motion Picture Association of America) and RIAA (Recording Industry Association of America) were prepared to act against sites by seizing their domains. |
| Pressure on webhosts | The trend will continue during 2012, but while it has proven to be an annoying inconvenience, it will probably only be effective in the short-term. |
| Forcing ISPs to block sites | While ISP blocks can be circumvented with ease, they still have the potential to hinder the growth of sites, as inexperienced users will quickly give up trying to access a non-responsive site. |
| Cutting off donations and payment processing | What.cd confirmed that, due to music industry pressure, their donations processing services have been withdrawn several times. Expect this approach to become more widespread. |
| Assaults on advertisers | Blocking the financial incomes of the pirate sites to take further legal action against advertisers. |

Figure 10 illustrates the proposed copyright protection system against BitTorrent piracy. The anti-BitTorrent filter (ABF) system is composed of a network analysis part, content feature extraction module and a torrent tracing part. The network analyzer uses a PCAP(Packet Capture) library to acquire information and extract BitTorrent traffic from the network. The content recognition works by obtaining the unique hash code that every BitTorrent transaction contains, and comparing this with a database of registered hash codes to identify the actual content being transmitted. If the system cannot identify by hash, it uses the feature identification method. When these files are transmitted, the hash value can be changed by creating a new BitTorrent transaction. Once the hash value is changed, the system can no longer recognize the content. This is the important issue in our system, and thus it is essential that the hash DB should be updated.

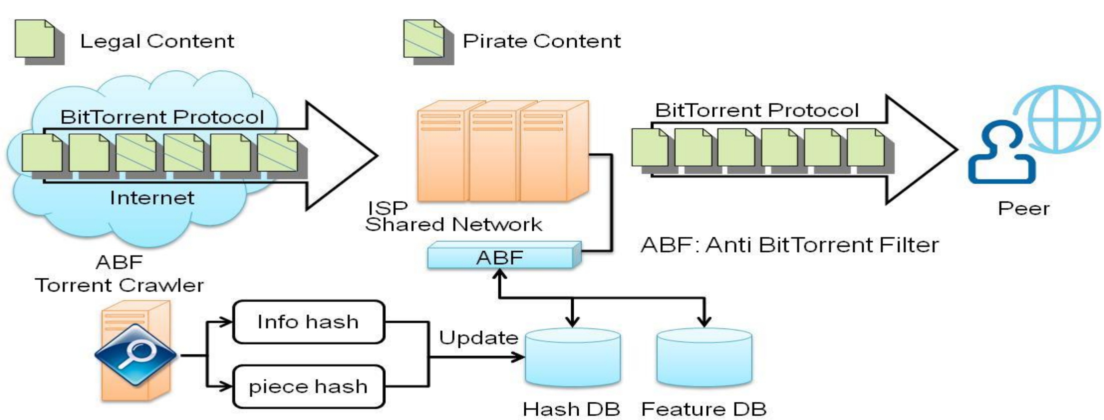


**Figure 10. Proposed Tracking System of Bittorrent for Copyright Protection**

In this tracking system, the content is identified by three ways. The first way is to use the info hash in the hash DB. Before transferring the content, peers exchange the info hash and we can identify the peer is going to transfer content, then the system blocks file transfer if the content distribution is not allowed. The second way is to use the piece hash value in the hash DB. Last way is to use the feature of the content that is able to identify if the content is slightly changed from original. Even if the peer changes the info hash of the content by creating a new transaction, the hash values of the pieces of the content is not changed. It means the proposed system is effective although new transaction is established for avoiding the system.

If there is a new file and a new BitTorrent transaction, it typically takes several hours (with very popular content) before the file has spread to enough computers to allow for high-speed downloads. Thus, we have several hours to update the hash DB. If we missed the packet between peer and tracker server, we cannot get the info hash and there is no way to identify the content distribution is legal or illegal. However, as we just get the packet during the file transferring between peers, we can calculate the hash value of the piece and check what the content is.

The torrent crawler gathers torrent meta-files and information about tracker sites and torrent sites. The crawler also extracts the hash info and piece hash from the torrent meta-file and updates the hash DB. Once the hash DB is updated, ABF uses the new hash to block illegal distribution of new content. Our system can also generate a warning message directing pirate users to the legal purchasing site.

## 5. Conclusions

While piracy techniques are becoming more intelligent and advanced, copyright protection technologies and tools for acquiring digital evidence are lagging behind. Therefore, it is important to study the prevention of copyright infringement on P2P software such as BitTorrent. In particular, to prevent large-scale damage from continued illegal distribution, we must develop a system that can quickly block unauthorized downloads.

In this paper, we proposed a tracking system for copyright protection against BitTorrent piracy. As BitTorrent uses trackers instead of a server, we can obtain the tracker URL from the torrent meta-file. As tracker sites have information about the seeders (uploaders) of content, they aid and abet content piracy. If the proposed system is installed on the ISP (Internet Service Provider) network, it is able to protect the content from BitTorrent distribution.

We have not considered the controversial issue of whether the deep inspection of network traffic by our system is legal or not. Despite the surrounding controversy, BitTorrent is still the main body of copyright infringement, which is why copyright owners would like to see it restricted. Technically, our system can block the illegal distribution of content and trace pirate devices because it has the content hash and can obtain IP and MAC addresses. Our system can also generate a warning message directing pirate users to the legal purchasing site.

## Acknowledgements

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

## References

[1] J. Kim, “Final Report on the Special Type OSP”, Korea Copyright Commission, **(2012)**.
[2] TorrentFreak, “We Need to Kill Piracy to Make BitTorrent Distribution Viable, Trolls Claim”, http://torrentfreak.com/we-need-to-kill-piracy-to-make-bittorrent-distribution-viable-trolls-claim-130712/ , visited Aug. 14 2013
[3] C. Carman, “How BitTorrent works”, http://computer.howstuffworks.com/bittorrent1.htm, visited **(2013)** August 14.
[4] Techterms: http://www.techterms.com/definition/p2p, visited **(2013)** August 14.
[5] Filesharing History: http://en.wikipedia.org/wiki/File_sharing, visited **(2013)** August 14.
[6] H. Schulze and K. Mochalski, “Internet Study 2008/2009: 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. ipoque”, Leipzig, Germany (2009), retrieved **(2013)** August 20.
[7] B. Cohen, “BitTorrent - a new P2P app”, Yahoo eGroups (2001), retrieved **(2013)** August 20.
[8] Bittorrent.com: BitTorrent and µTorrent Software Surpass 150 Million User Milestone. January 2012, retrieved **(2013)** August 20.
[9] fastcompany.com: retrieved **(2013)** August 14.
[10] comscore.com: (2010), retrieved **(2013)** August 14.
[11] S. Jacobsson-Purewal, “Copyright Trolls: 200,000 BitTorrent Users Sued Since 2010”, PC World (2011), retrieved **(2013)** August 14.
[12] Five Anti-Piracy Strategies Designed to Hurt Torrent Sites in 2011: http://torrentfreak.com/5-anti-piracy-strategies-designed-to-hurt-torrent-sites-in-2011-110102/ , visited **(2013)** August 14.

## Authors

**JungJae Lee** is a professor of Dept. of Entertainment Business at Soongsil Cyber University. He received the Ph.D. degree from Soongsil University major in management in 2012. He worked as team manager of Korea Copyright Commission from 2000 to 2012. His research interest is copyright business, entertainment business, copyright management, and digital content.

**Jongweon Kim** received the Ph.D. degree from University of Seoul, major 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.