



ICONI



KOREAN SOCIETY FOR INTERNET INFORMATION

The 11th International Conference on Internet

(ICONI 2019)

Dec. 15-18, 2019, Lotte Hotels & Resorts Hanoi, Vietnam

<http://www.iconi.org>

Proceedings of ICONI 2019

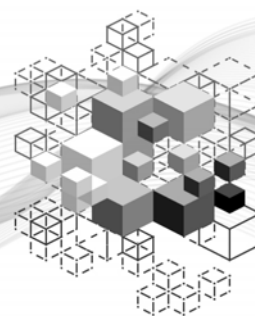
| Organized by |

Korean Society for Internet Information (KSII)

| Sponsored by |

Korea Electronics Technology Institute (KETI)

Contents



9-2	EDISON Simulation Linked Competition Platform Construction and Service Yejin Kwon, Noori On, Inho Jeon, Jerry H. Seo, Jongsuk R. Lee (KISTI, ROK)	253-256
9-3	Knowledge-based Project Management System (KPMS) for Industrial Construction Projects Minhui Ha, Donghee Kim, Jaehyun Choi (Korea Univ. of Technology & Education, ROK)	257-261
9-4	Lessons from HPC course by Government sponsored Research Institutes in Korea Eunha Lee, Jung Seung Wha, Hae Jin Kang, Mijoo Yang, Sunggeun Han (KISTI, ROK)	262-266
9-5	History and trends of Data Education in Korea - KISTI Data Education Based on Statistics in 2001 ~ 2019 Jaehong Min, Jaewoo Do, Jaegwang Lee, Jaesung Kim, Buyoung Ahn (KISTI, ROK)	267-271
9-6	A Knowledge Engineering Approach for Gaming Big Data Analytics using Orient Database Framework Pir Hamid Ali Qureshi, Isma Farah Siddiqui (Mehran Univ. of Engineering and Technology, Pakistan)	272-276
9-7	A Web-based 3D Molecular Orbital Visualizer based on Computational Science Platform Inho Jeon, Noo Ri On, Yejin Kwon, Huiseung Chae, Jerry H. Seo, Jong Suk Ruth Lee (KISTI, ROK)	277-279
9-8	An Effects on Users Perceived usefulness and Ease of Computational Science Platform on User Satisfaction Noo Ri On, Nam-Gyu Kim, Hanbichnale Jang, Jong Suk Ruth Lee (KISTI, ROK)	280-284
10-1	Study of Prediction Research Regarding Smart Office Using Scenario Analysis Ho Seok Lee, Bong Gyou Lee (Yonsei Univ., ROK)	285-288
10-2	System for Analyzing Pirated Content Distribution Channel Deuk-Hun Kim, Jin Kwak (Ajou Univ., ROK)	289-293

System for Analyzing Pirated Content Distribution Channel

Deuk-Hun Kim¹ and Jin Kwak²

¹ISAA Lab., Department of Computer Engineering, Ajou University
Suwon, Republic of Korea

²Department of Cyber Security, Ajou University
Suwon, Republic of Korea

[e-mail: dhkim.isaa@gmail.com, security@ajou.ac.kr]

*Corresponding author: Jin Kwak

Abstract

Technology and culture are creating content added value by integrating with various industries. In addition, countries are creating an infrastructure environment interoperability for integrating new technologies and cultural contents. However, copyright infringement caused by pirated contents continues to occur for copyrighted cultural contents. Various researches are progressed to prevent copyright infringement, but the current prevention technology is mainly post response, which raises the issue of effectiveness. Accordingly, there is a need for a technology for preventing illegal distribution in advance by analyzing a distribution channel of pirated contents and blocking a trigger. In this paper, we propose a system for analyzing distribution channel and blocking trigger through a centralized cloud monitoring server after crawling data from trusted sites and torrent sites.

Keywords: Copyright, Pirated Content, Torrent Site, Crawling

1. Introduction

Technology and culture are creating added value by integrating with various industries[1]. However, copyright infringement caused by pirated contents continues to occur for copyrighted cultural contents. Various researches are progressed to prevent copyright infringement, but the current prevention technology is mainly post response, which raises the issue of effectiveness. Accordingly, there is a need for a technology for preventing illegal distribution in advance by analyzing a distribution channel of pirated contents and blocking a trigger.

In this paper, we analyze the definition of copyright, the current state of the pirated market, and the torrent site in Chapter 2. Afterwards, we

propose a trigger blocking model through the analysis of illegal distribution channels in Chapter 3. Finally, conclusions are made in Chapter 4.

2. Related Works

2.1 Copyright

Copyright is the legal right of the content creator to the content. In other words, it aims to develop culture by protecting the rights of creator[2]. However, as copyright infringement caused by pirated contents has spread, difficulties continue in protecting content copyright.

2.2 Pirated Content Market

As of 2017, about 16 million people had experience using pirated contents. Pirated contents are distributed through web hard, P2P(Peer-to-Peer), torrents, etc. Recently, copyright infringement has also occurred through streaming services.

In particular, the torrent is a one-to-N method of distributing content files in pieces and merging them into pieces. So, the boundary between the uploader and the downloader is unclear, which makes it difficult to crack down and highest share of pirated contents distribution[1].

2.3 Torrent Site

Torrent site refers to a site that distributes a torrent file(.torrent) that can upload/download files through BitTorrent, a P2P file transfer protocol. Fig. 1 shows the structure of sharing P2P file using BitTorrent protocol.

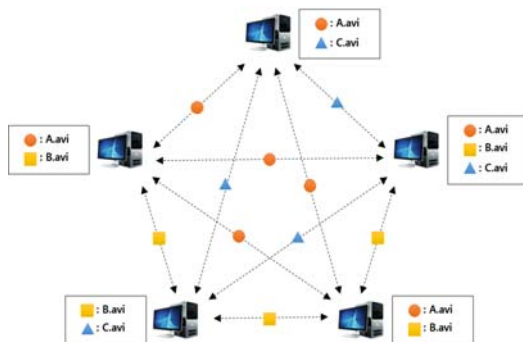


Fig. 1. P2P file sharing using BitTorrent protocol

As this time, there are also legal torrent sites that distribute torrent files of legally redistributable content, but there are many torrent sites that distribute torrent files for pirated contents that infringe copyright law.

Recently, in order to circumvent torrent site crackdown monitoring, there have been an

increasing number of cases in which pirated contents are distributed after opening a site with a server overseas[1]. Accordingly, using only a method of detecting and blocking a torrent site has a problem of effectiveness in preventing content infringement. Therefore, a study is needed to analyze the distribution channels of pirated contents and to use the method to block the trigger.

3. Proposed System Model

This chapter proposes a system model that selects a torrent site and a category of infringement content to analyze the distribution channels of pirated contents in the torrent site, and then blocks the trigger of the torrent files distribution.

3.1 Torrent Site Selection

In order to analyze the distribution channels, the torrent sites were selected according to Alexa daily access time[3] among the torrent sites searched through Google site as of September 2019, as shown in Table 1.

Table 1. Torrent site selection

Torrent Name	Site	Daily Time on Site(MM:SS)
Torrentwal	torrentwal.com	03:37
Torrentbe	torrentbe.com	03:17
Torrentmap	torrentmap.com	02:56
Torrentqq	torrentqq.com	01:53
Freefilewal	filewal.com	01:34

3.2 Infringement Content Category

The infringement content categories common to the selected torrent sites are classified into three categories as analyzed in Fig. 2.



Fig. 2. Analysis of infringement content category

The features of the torrent file for each infringement content distributed in the torrent site are as follows.

@ Movie content

: Torrent file distribution based on the start date of online VOD service

: Example) 토렌트영화, 영화

@ TV media content

: Torrent file distribution progress within one day after the end of main broadcasting

: Example) TV예능, 한국TV, 방송, 티비

@ Music content

: Torrent file distribution progress within one day based on the top 100 real-time sound charts

: Example) 음악, 해외음원

3.3 System Model

We analyze the distribution channels of torrent files and propose a system model as shown in Fig. 3 to block the trigger of the torrent file distribution.

Step 1. Crawl content features on sites that share reliable information for each infringement content category. The torrent sites also crawl content features. The crawling methods for each infringement content are as follows.

@ Movie content

: The trusted site was chosen by the Korea Film Council(KOFIC), which manages online VOD services for movie content[4]. Fig. 4 shows the torrent file feature of movie contents and analyzes the range of data to be crawled. The KOFIC crawls online VOD service start date, which is a feature of movie contents. In addition, each torrent site crawls a post name, a posting date, and a hash value of a torrent file in the movie content category.

@ TV media content

: The trusted site was chosen by the Electronic Program Guide(EPG) that manages the main broadcast schedule of TV media contents[5]. The EPG crawls the main broadcast schedule, which is feature of TV media contents. In addition, the posting type of the TV media content in each torrent site is posted in the form of 「Title.E00.YYMMDD.Quality」, crawls a

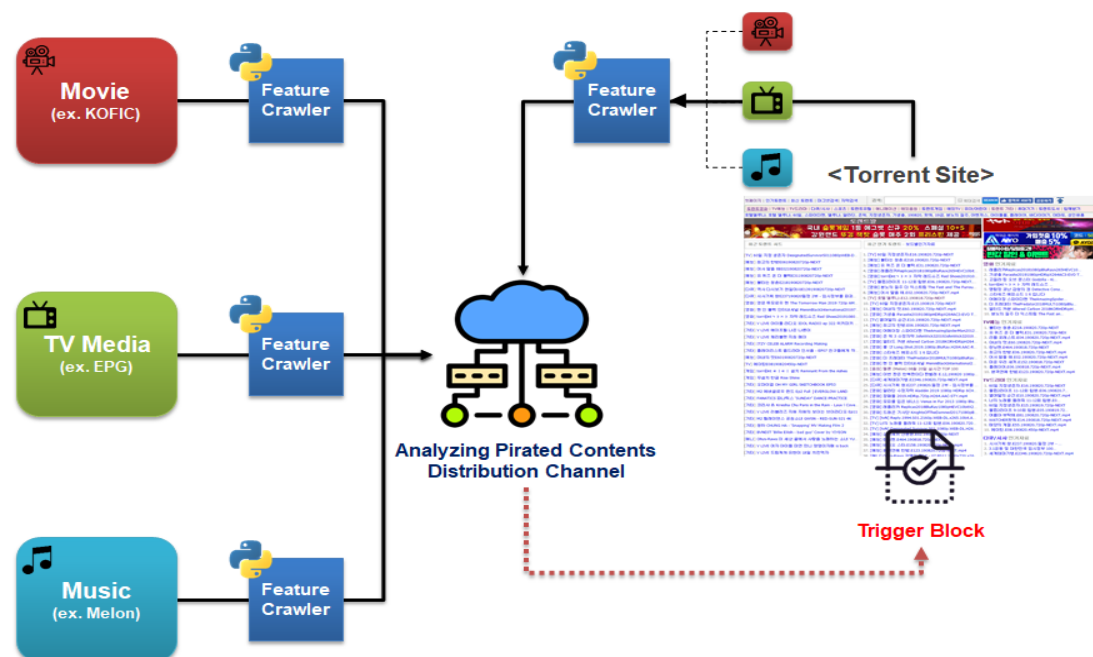


Fig. 3. Proposed system model

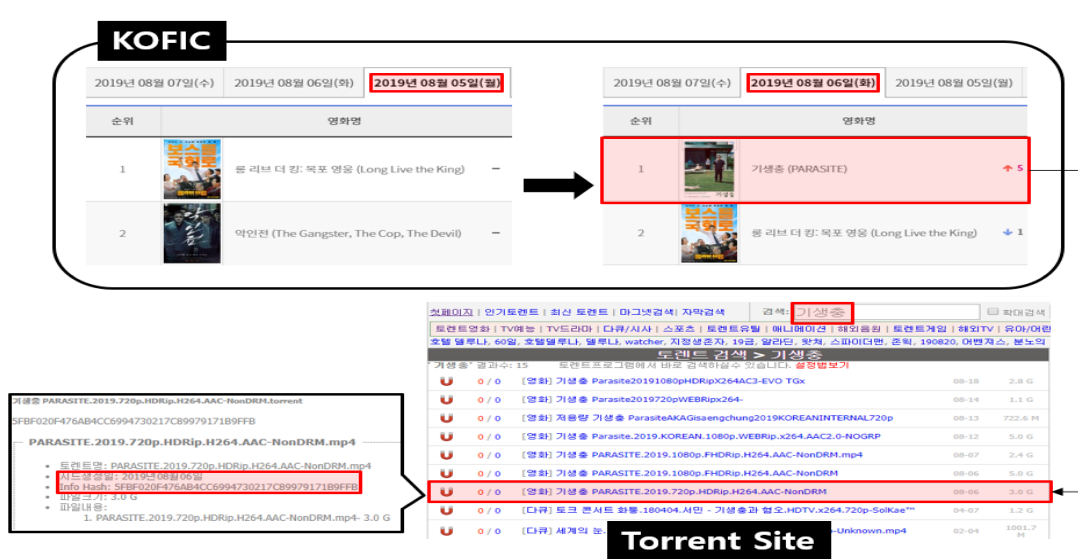


Fig. 4. Example of movie content feature crawling

post name, a posting date, and a hash value of a torrent file in the TV media content category.

@ Music content

: The trusted site was chosen by the Melon, which manages the real-time sound chart of music contents[6]. The Melon crawls top 100 real-time sound charts, which is a feature of music contents. In addition, the posting type of the music content in each torrent site is posted in the form of

「Music Site Name(Melon).MM.DD.Time.

TOP 100」, crawls a post name, a posting date, and a hash value of a torrent file in the music content category.

Step 2. Send the crawled data to a centralized cloud monitoring server. The server blocks the trigger by storing the sent crawled data and analyzing the distribution channels by continuously comparing and analyzing the data.

For example, by comparing the crawled data shown in the Table 2 through the proposed

Table. 2. Crwaled data from movie content

Movie title	Site	Type	Crwaled data				Trigger (Influential)
			Posting title	Upload day	Upload time	Hash value	
기생충 (Parasite)	KOFIC	Trusted site	PARASITE	2019-08-06	-	-	×
	Torrentwal	Torrent site	PARASITE .2019.720p	08-06	-	5FBF020F476AB4CC6994730217C89979171B9FFB	△
	Torrentbe			08-08	20:00	-	×
	Torrentmap			08-13	18:00	-	×
	Torrentqq			08-06	00:11	5FBF020F476AB4CC6994730217C89979171B9FFB	○
	Freefilewal			08-06	13:25	5FBF020F476AB4CC6994730217C89979171B9FFB	×

model in the movie content and analyzing the distribution channels, the trigger of the torrent site is blocked. According to the **Table 2**, the online VOD service start date of 「*Parasite*」 by the trusted site KOFIC is “2019-08-06”. At this time, the site to which the torrent file is uploaded on the same date as the start date is compressed to 「*Torrentwal*, *Torrentqq*, *Freefilewal*」.

After that, the torrent file is compressed into 「*Torrentwal*, *Torrentqq*」 by comparing the uploaded time and hash value, and since the uploaded time from 「*Torrentwal*」 could not be crawled, 「*Torrentqq*」 is analyzed as a powerful trigger and blocked.

4. Conclusions

The distribution of pirated contents in which content copyright is infringed continues to occur through various platforms. In particular, torrents have difficulty finding pirated contents distributors due to the one-to-N file transfer feature of the BitTorrent protocol.

Therefore, in this paper, we analyzed the features of infringement contents in torrent sites, and compared and analyzed the information of trusted sites through data crawling, and proposed a model for distribution channels analysis and trigger blocking. This will help protect content and improve the copyright infringement environment.

References

- [1] “2018 Annual Report on Copyright Protection”, Korea Copyright Protection Agency, pp. 1-393, 2018.
- [2] Wikipedia, “Copyright”, <https://en.wikipedia.org/wiki/Copyright>
- [3] Alexa, “Drive More Website Traffic with Competitive Analysis”, <https://www.alexa.com/siteinfo/>
- [4] KOFIC, <https://kofic.or.kr/>
- [5] EPG, <http://www.epg.co.kr/>
- [6] Melon, <https://www.melon.com/>