



ICONI



KOREAN SOCIETY FOR INTERNET INFORMATION

The 12th International Conference on Internet (ICONI 2020)

Dec. 13-16, 2020, Landing Convention Center,

Jeju Shinhwa World, Korea

<http://www.iconi.org>

Proceedings of ICONI 2020

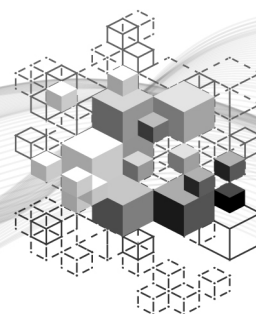
| Organized by |

Korean Society for Internet Information (KSII)

| Sponsored by |



Contents



9-3	Automation Mechanism of Detecting Suspected Piracy Video Streaming Sites Minkyung Lee, Deuk-Hun Kim, Jin Kwak (Ajou Univ., ROK)	392-396
9-4	Prediction Method for New Domain of Suspected Copyright Infringement Site Hae-Seon Jeong, Deuk-Hun Kim, Jin Kwak (Ajou Univ., ROK)	397-402
9-5	Automated Detection Architecture for Piracy Sites Based on Link Collection Sites Eui-Jin Kim, Deuk-Hun Kim, Jin Kwak (Ajou Univ., ROK)	403-406
9-6	Usability and Performance Evaluation of Medical Institution Websites in Malaysia Nor Azman Ismail, Mohd Ainal Farhan Mohamad Johari, Mohammad Ridzwan Syah Irwan, Mohamad Faiz Hakimi Ishak, Su Elya Namira Mohamed (Universiti Teknologi Malaysia, Malaysia)	407-413
9-7	Exploring Levels of Credit Risk in P2P Lending: A Quantitative and Qualitative Emotion Mining Method Min Gyeong Kim, Francis Joseph Costello, Tolchinski Rita, Kun Chang Lee (Sungkyunkwan Univ., ROK)	414-419
9-8	Sentiment Analysis: Study of Social Behaviour of Youth in Different Social Class Nor Azman Ismail, Siti Noorafishah Abd Latif, Su Elya Namira Mohamed (Universiti Teknologi Malaysia, Malaysia)	420-426
9-9	A Novel Approach for Airport Recommendation Using the Comparison of Hierarchical Regression, PLS-SEM, and Bayesian Network-Based PSEM Cheong Kim, Dong Eun Lee, Kyun Sun Eo, Kun Chang Lee (Sungkyunkwan Univ., ROK)	427-432
9-10	C3BR: A Novel Multi-Label Classification Ensemble Mechanism Jong Yoon Won, Chang Min Kang, Dong Eun Lee, Kun Chang Lee (Sungkyunkwan Univ., ROK)	433-438

Automation Mechanism of Detecting Suspected Piracy Video Streaming Sites

Minkyung Lee¹, Deuk-Hun Kim¹ and Jin Kwak^{2*}

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

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

[e-mail: leemkl@ajou.ac.kr, kimdh1206@ajou.ac.kr, security@ajou.ac.kr]

*Corresponding author: Jin Kwak

Abstract

Recently, watching videos such as movies and dramas is increasing by using video streaming services rather than video downloading services worldwide. As a result, the scale of a video streaming market is increasing. However, video streaming sites that infringe copyrights such as movies and dramas are also on the rise. In response to this, it takes action such as closing piracy streaming sites, but the sites are becoming more sophisticated and intelligent to bypass it. Even intelligent piracy sites are almost identical to legal streaming sites. So, it could not be detected by existing piracy streaming site feature-based detection techniques. Therefore, in this paper, we analyze features of legal streaming sites and suspected piracy video streaming sites that cause copyright infringement. Also, we propose a technique of detecting suspected copyright infringement video streaming sites based on a whitelist to respond to increasing and intelligent piracy sites.

Keywords: Copyright, Copyright infringement, Piracy sites, Whitelist

1. Introduction

Digital contents such as films and TV programs can be accessed more easily through web browsers as IT(Information Technology) advances. Recently, however, there have been piracy sites such as unauthorized distribution or reproduction of digital contents without the the copyright holder's permission. Besides, they distribute digital content in various types, such as downloading, torrent, and streaming [1]. Among them, video streaming service is increasing rather than directly downloading videos, and piracy video streaming sites are also increasing [2]. There are also increasing advanced piracy video streaming sites similar to legal video streaming sites and not characteristic of general

piracy video streaming sites. Therefore, studies are needed to detect automatically in order to efficiently detect piracy video streaming sites and are to detect advanced piracy video streaming sites that are similar to legal video streaming site. Accordingly, in Chapter 2, we analyze features of piracy video streaming sites and detecting methods of piracy sites in the existing studies. In Chapter 3, we propose an automation mechanism of detecting suspected piracy video streaming sites using a whitelist to detect advanced piracy video streaming sites. Finally, we present the conclusion in Chapter 4.

2. Related Works

2.1 Feature of piracy video streaming sites

Feature of piracy video streaming sites for detecting piracy video streaming sites that infringe on digital content are as follows [1].

- Content list

The piracy video streaming sites include a content list that contains the title of content so that users can easily access to watch digital content.

- Advertisement

The piracy video streaming sites include advertising banners in various parts for profit-making reasons.

- Streaming server list

In some cases, the piracy video streaming sites also include another streaming server list without embedded streaming video content. Even if one streaming server is blocked, a list of two or more streaming servers can be provided for the user to select the desired quality.

2.2 Methods of detecting piracy sites

The proposed methods for detecting piracy sites in previous studies are as follows.

- Feature analysis and detection technique for piracy sites

The piracy site was classified as a torrent, video streaming and webtoon site, and feature analysis was performed in this technique. Subsequently, the torrent, video streaming, and webtoon sites were detected as piracy sites according to each piracy site [1].

- A detection of webtoon infringement site using page source analysis

This technique can detect webtoon infringement sites through page source analysis by crawling websites. This technique derived a business registration number in legal sites and keyword characteristics used in webtoon infringement sites. As a result, they can detect suspected infringement sites according to the keyword list. In comparing and analyzing the keyword list, the accuracy was increased using the process of separating compound words rather than simply comparing them [3].

- Detection technique of suspected webtoon infringement site using web site advertisement banners

This technique can detect webtoon infringement sites by analyzing advertising banners in piracy websites, a profit-making element of piracy sites. After extracting websites in image form through Full Page Screen Capture, a google extension program, they used gambling-related keyword lists used in piracy sites through OCR(Optical Character Recognition) analysis to distinguish legal sites from piracy sites [4]. Piracy video streaming sites also utilize advertising banners in web sites, which are a profit-making factor, and the advertising banners are the same as the keyword list used by the piracy sites. Therefore, we propose an automation mechanism using the analysis of advertising banners.

2.3 Whitelist

We use a whitelist technique in the proposed mechanism for detecting piracy video streaming sites in this paper. The whitelist is used as a positive security method compared the blacklist, which allows only proven ‘security’ and blocks proven ‘malicious’ [5]. As piracy video streaming sites are advanced and intelligent with similar features to legal streaming sites, advanced piracy video streaming sites have difficulty detecting piracy sites using features of legal video streaming sites and piracy video streaming sites. Therefore, we apply a whitelist technique to detect or monitor malware, phishing sites, etc. to detect advanced piracy video streaming sites.

3. Proposed Scheme

3.1 Analysis of legal and piracy video streaming sites

We organize the legal video streaming site and piracy video streaming site features as shown below in Table 1. Features of sites used in Table 1 are as follows:

- Business Registration Number

The legal video streaming site includes the business registration number, which is the National Tax Service’s management number.

- Contents list

Name of contents for providing digital contents on the video streaming sites

- Banner Advertising

Gambling-related advertising banner

- Streaming Server list

A link list to another external streaming site

- Embedded Video

Video to watch digital content directly from a site without moving to an external site

Table 1. Analysis of comparison legal video streaming sites to piracy video streaming sites

Feature of sites	Legal sites	Piracy sites	Advanced Piracy sites
Business Registration Number	O	X	O (Invalid number)
Content lists	O	O	O
Banner Advertising	X	O	X
Streaming Server list	X	O	X
Embedded video	O	X	O

While piracy video streaming sites contain streaming server lists and advertising banners, there are advanced piracy sites that do not include streaming server lists and advertising banner. They also include a business registration number and an embedded video similar to the feature of legal video streaming sites, such as **Table 1**.

3.2 Feature of piracy video streaming sites

□ Streaming keyword

Table 2. List of streaming keyword using piracy sites

Video Keyword	Piracy Site		
	ogu*** .com	the*** .guru	joo*** .net
Replay link	O	O	-
Only link	O	O	O
Legal issue	O	O	-
Copyright	O	O	O
Provided link	O	O	-
Inquiry of advertising	-	O	-

Streaming keywords, which are mainly used in video streaming sites, are shown in **Table 2**.

□ Link list keyword

Table 3. List of link keyword using piracy sites

Link List Keyword	Piracy Site		
	ogu*** .com	the*** .guru	joo*** .net
Glimited	-	O	O
HPlay	-	O	O
JawCloud	-	O	-
Jload	-	O	O
MixDrop	O	-	-
JetLoad	O	-	-
FlashVid	O	O	-
SuperVid	O	O	O

Piracy site provides streaming services by providing external streaming sites. Therefore, a link list used in piracy sites is derived to extract commonly used external streaming sites. The link list, which is commonly used on the three piracy video streaming sites, is shown in **Table 3**.

3.3 Proposed automation mechanism for detecting piracy video streaming sites

In this paper, we propose a whitelist-based automation mechanism of detecting piracy video streaming sites to detect advanced piracy video streaming sites. Accordingly, the analysis of legal and piracy video streaming sites and the feature of piracy video streaming sites are utilized, and the mechanism proposed in this paper are as follows, **Fig. 1**.

Step 1. Site Crawling and Suspicion

Kim et al. proposed the method of site crawling and checking a business registration number. Unlike the existing proposed method, piracy video streaming sites are advanced and intelligent, similar to legal video streaming sites. Therefore, it is difficult to determine that it is not a piracy video streaming site based on whether business registration numbers are included. If the business registration number is included in the site, it goes to *Step 4*. The pseudo-code for implementing site crawling and checking a business registration number is shown in **Table 4**.

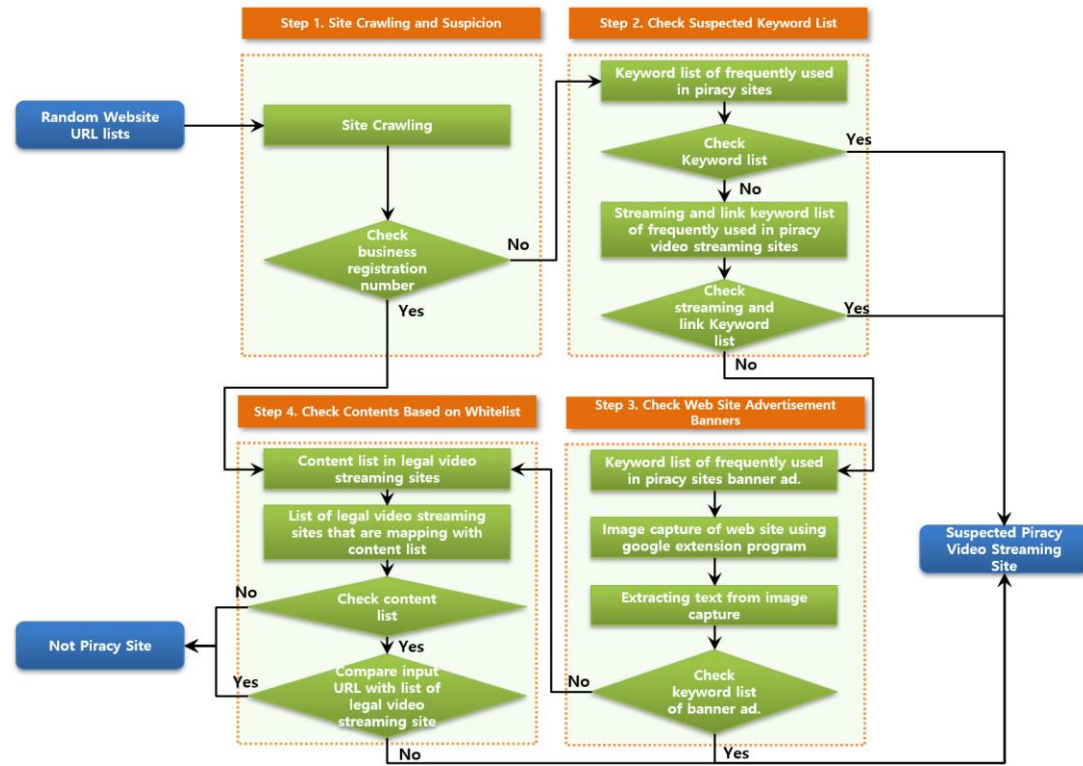


Fig. 1. Proposed mechanism for detecting piracy video streaming sites

Table 4. Pseudo-code for site crawling and checking suspicion

Function for site crawling and checking suspicion:	
1	$str_html \leftarrow$ crawled site
2	$license_signature \leftarrow$ regular expression for determining business registration number
3	IF there is business registration which is matching with $license_signature$ in a site,
4	RETURN business registration number
5	ELSE, RETURN -1

Step 2. Check Suspected Keyword List

This step compares to the keyword list mainly used in the piracy sites used in the proposed method by Kim et al. The following Table 5 is a pseudo-code for Step 2.

Table 5. Pseudo-code for checking keyword list

Function for checking keyword list:	
1	$str_html \leftarrow$ crawled site in Step 1
2	$commonKeyword \leftarrow$ keyword list mainly used in piracy site
3	FOR keyword in $commonKeyword$:
4	IF keyword is mapped by str_html , RETURN 0
5	RETURN -1

Besides, the keyword of the piracy video streaming sites derived from this paper is applied in Step 2. This step compares with the streaming keyword list (streaming keyword in Table 2) and the link keyword list (link list keyword in Table 3) based on features of piracy video streaming sites. The following Table 6 is a pseudo-code for implementing Step 2.

Table 6. Pseudo-code for checking streaming and link keyword list

Function for checking streaming and link keyword list:	
1	$str_html \leftarrow$ crawled site in Step 1
2	$streamingKeyword \leftarrow$ streaming keyword list
3	$linkKeyword \leftarrow$ link keyword list
4	FOR keyword in $streamingKeyword$:
5	IF keyword is mapped by str_html , RETURN 0
6	FOR keyword in $linkKeyword$:
7	IF keyword is mapped by str_html , RETURN 0
8	RETURN -1

Step 3. Check Web Site Advertisement Banners

It can check advertisement banner in piracy video streaming site using the proposed method by Kim et al. The following **Table 7** is a pseudo-code for implementing *Step 3*.

Table 7. Pseudo-code for checking banner advertising

Function for checking banner advertising:	
1	$Ad_keyword \leftarrow$ keyword list of banner advertising
2	$img \leftarrow$ image capture using a google extension program
3	$img_text \leftarrow$ converting img to text
4	FOR keyword in $Ad_keyword$:
5	IF keyword is mapped by str_html ,
	RETURN 0
6	RETURN -1

Step 4. Check Contents Based on Whitelist

It proposes a mechanism based on a whitelist for sites(false negative) that have not been detected as suspected piracy video streaming sites through the results of *Step 1* to *Step 3*. It prepares a list of the content list that was broadcasted and posted on the legal video streaming sites before using a whitelist. It also prepares a URL(Uniform Resource Locator) list of legal video streaming sites, which are mapping with a prepared list of the content. After that, it compares a content list in not suspected piracy video streaming sites to the prepared content list. If there is no content in a site, it detects not suspected video streaming sites. If there is content in a site and the URL of the site is not legal video streaming sites, it detects suspected video streaming site. The following **Table 8** is a pseudo-code for implementing *Step 4*.

Table 8. pseudo-code for detection piracy sites using a whitelist

Function for detection piracy sites using whitelist:	
1	$str_html \leftarrow$ crawled site in <i>Step 1</i>
2	$contents \leftarrow$ content list that is posted in legal video streaming sites
3	$legal_list \leftarrow$ legal video streaming site URL corresponding to $contents$
4	While(repeat by the length of $contents$):
5	IF there is the same content as the $contents$ in str_html ,
6	IF URL of str_html does not match in the $legal_list$, RETURN -1
7	ELSE RETURN 0
8	RETURN -1

In this paper, we proposed the mechanism, which applied features of piracy video streaming site and the technique of previous studies to detecting

piracy sites in *Step 1* to *Step 3*. In addition, we applied whitelist to sites that were not detected as sites suspected of piracy video streaming sites in *Step 1* to *Step 3* to improve the detection rate of suspected piracy video streaming sites.

4. Conclusion

Various piracy sites can detect suspected piracy sites using common features because piracy sites are similar to each other. Besides, the detection rate can be improved by applying a whitelist technique if the feature of the piracy sites is similar to those of legal sites due to the sophistication and intelligence of piracy sites. In the future, we will study ways to improve the detection rate of suspected piracy sites of various types of infringement such as torrents and webtoons through whitelist-based techniques and apply high-speed search algorithms to improve the detection performance of suspected piracy sites.

References

- [1] S.K. Choi and J. Kwak, "Feature Analysis and Detection Technique for Piracy Sites", *KSII Transactions on Internet and Information Systems*, Vol 14, No. 5, pp 2204-2220, May. 2020.
- [2] European Union Intellectual Property Office (EUIPO), "Online Copyright Infringement in The European Union: Music, Film, TV (2017-2018), Trends and Drivers", *European Union Intellectual Property Office*, Nov. 2019.
- [3] E.J. Kim, S.K. Choi and Jin Kwak, "Detection Technique of Suspected Webtoon Infringement Site using Web Site Advertisement Banners", *Proceedings of KSII Spring Conference*, May. 2020.
- [4] E.J. Kim, S.K. Choi and J. Kwak, "A Detection of Webtoon Infringement Site Using Page Source Analysis", *Proceedings of the 15th Asia Pacific International Conference on Internet Science and Technology (APIC-IST 2020)*, Jul. 2020.
- [5] A. Sedgewick, M. Souppaya and K. Scarfone, "Guid to Application Whitelisting", *National Institute of Standards and Technology (NIST)*, Oct. 2015.