논문 2023-4-12 http://dx.doi.org/10.29056/jsav.2023.12.12

IP 포트 스캐닝을 이용한 불법 유통 플랫폼 추적 방법

유인재**, 이재청*, 장세영**, 박병찬**, 김석윤**, 김영모**†

An Illegal Distribution Platform Tracking Method Using IP Port Scanning

In-Jae Yoo**, Jae-Chung Lee*, Seyoung Jang**, Byuong-Chan Park**,
Seok-Yoon Kim**, Youngmo Kim**†

요 약

최근 미디어 시장의 규모가 증가하면서 저작권을 침해하는 불법 사이트의 수가 지속적으로 증가하고 있다. 정부에서는 불법 사이트에 대한 대응으로 해당 사이트를 차단하는 방법을 사용하고 있으며, 불법사이트를 차단하기 위해서는 서버의 물리적 위치를 특정할 수 있어야 한다. 하지만 대부분의 불법 사이트는 DNS 우회와 리버스프록시(Reverse proxy) 등과 같은 보안기술을 통해 IP주소를 감추고 있다. 본 논문에서는 이러한 문제점을 해결하기 위하여 은닉된 상태에서 서비스 중인 불법 사이트의 IP를 특정하기 위해 IP 포트 스캐닝을 이용하는 불법 유통 플랫폼 추적 방법을 제안하고자 한다.

Abstract

As the size of the media market has increased recently, the number of illegal sites that infringe media copyrights is continuously increasing. The government uses site blocking methods in response to illegal sites, and in order to block illegal sites, the physical location of the server must be identified. However, most illegal sites hide their IP addresses through security technologies such as DNS bypass and reverse proxy. In order to solve this problem, this paper proposes a method for tracking illegal distribution platforms using IP port scanning to identify the IPs of illegal sites that are in service in a hidden state.

한글키워드 : IP주소, 콘텐츠 전송 네트워크(CDN), 클라우드 서비스, 웹 크롤링, 불법사이트

keywords : Internet Protocol, CDN(Content Delivery Network), Cloud Service, Web Crawling, Illegal site

1. 서 론

최근 미디어 시장의 규모가 증가하면서 미디어의 저작권을 침해하는 불법 유통 플랫폼의 수가 지속적으로 증가 하고 있다[1][2]. 불법 유통 플랫폼은 URL을 지속적으로 변경하거나 리다이렉션 하는 등의 여러 방식으로 단속을 회피하고 있다[3]. 불법 유통 플랫폼을 근본적으로 차단시키기 위한 정책을 진행하기 위해서는 서버의 물리적 위치를 발견해야한다[4]. 대부분의 불법 유통 플랫폼

* (주)비욘드테크
** 숭실대학교 컴퓨터학과
† 교신저자: 김영모(email:ymkim828@ssu.ac.kr)
접수일자: 2023.12.11. 심사완료: 2023.12.16.
게재확정: 2023.12.20.

은 DNS, 리버스프록시 등과 같은 보안기술을 통해서 IP주소를 감추고 있다. 이렇게 은닉된 상태에서 서비스 중인 불법 유통 플랫폼의 실 IP주소를 특정함으로써 불법 유통 플랫폼의 운영자를 추적하고 불법 유통 플랫폼 운영을 근본적으로 차단하는 방안이 요구된다[8][9]. 따라서 본 논문에서는 이러한 문제점을 해결하고자 IP 포트 스캐닝을 이용한 불법 유통 플랫폼 추적 방법을 제안한다.

본 논문의 구성으로 2절에서는 관련 연구로 불법 유통 플랫폼의 특징을 설명하고, 3절에서는 본 논문에서 제안하는 원격지 서버의 IP주소 특정 방법에 대해서 설명한다. 4절에서는 제안하는 방법에 대한 실험 및 결과를 제시하고, 5절에서 결론으로 마무리한다.

2. 관련연구

2.1 불법 유통 플랫폼 특징

불법 유통 플랫폼은 온라인상에 유통되는 불법 정보 및 유해 정보를 포함하는 플랫폼으로서, 유형으로는 저작권 위반 스트리밍, 웹툰, 음란물, 도박, 마약 거래 등이 있다. 국가에서는 불법 또는 유해정보를 제공하는 불법 유통 플랫폼의 접속을 차단하기 위해 DNS(Domain Name Service) 차단 방식, SNI(Server Name Indication) 필드 차단 방식 등 여러 가지 차단 방식을 운영하고 있다[5]. 하지만 불법 유통 플랫폼은 URL을 불규칙하게 변경하거나, 다양한 URL을 갖고 있어 기존의 DNS 차단 방식을 우회하여 계속해서 운영되고 있다[6]. 또한 대부분의 불법 유통 플랫폼은 웹사이트에 사용되는 콘텐츠의 효율적인 전달을 위해 사용하는 CDN 서비스를 사용하고 있다. 특히 CDN 서비스는 불법 유통 플랫폼이 운영하는 실제 IP 주소를 반환하는 것이 아닌 CDN 서비스의 IP 주소를 반환하고 있다.

2.2 Port Scan

포트스캔은 특정 IP에 대하여 열려 있는 포트, TCP/UDP 등 프로토콜의 유형, SSH 등 배너의 종류, 운영체제 버전 등을 확인하는 수단으로 활용된다[7]. 특정 IP에 TCP와 UDP 패킷을 보내고, 응답 패킷을 분석하여 운영체제 핑거프린트를 획득한 후, 기존에 확보해 둔 데이터베이스와 비교하여 운영체제(OS)를 식별하기도 한다.

3. 원격지 IP주소 추적 방법

3.1 원격지 IP주소 추적 개요

그림 1은 본 논문에서 제안한 보안기술을 통해 은닉된 원격지 서버의 IP 특정 방법 구조도이다.

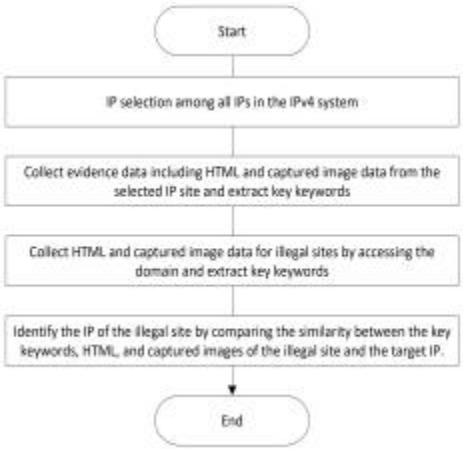

그림 1. IP 특정 방법 구조도
Fig. 1. Structural diagram of IP specification method

검색 단계, 사이트 정보 수집 단계, 표적 사이트 정보 수집 단계 그리고 IP 특정 단계로 이루어져 있다.

3.2 검색단계, 사이트 정보 수집 단계

검색 단계에서는 IPv4 체계의 전체 IP에 대하

여 서비스 여부 및 특정 포트의 개방 여부를 확인하여 분석 대상이 되는 복수의 대상 IP를 선별한다. 즉 42억 개의 IP에 대해 현재 서비스 여부를 확인하고 웹 서비스 포트인 80(HTTP PORT)와 443(HTTPS PORT)의 개방 여부를 확인한다. 이렇게 선별된 IP는 사이트 정보 수집 단계로 전달한다.

사이트 정보 수집 단계에서는 선별된 대상의 IP별로 해당 IP에서 추출되는 HTML 문서 및 메인 페이지 캡쳐 이미지를 증거 데이터로 수집하고, HTML 문서에서 복수의 주요 키워드를 추출하여 색인한다. 정보 수집 단계에서는 전체 IP 중에 현재 서비스 중이면서 80 포트와 443 포트 중 하나가 오픈 되어있는 대상으로, 해당 IP에서 서비스하는 HTML 문서와 사이트 메인 페이지의 이미지 파일을 추출 및 다운로드 하여 이를 증거 데이터로 수집해 둘 수 있다.

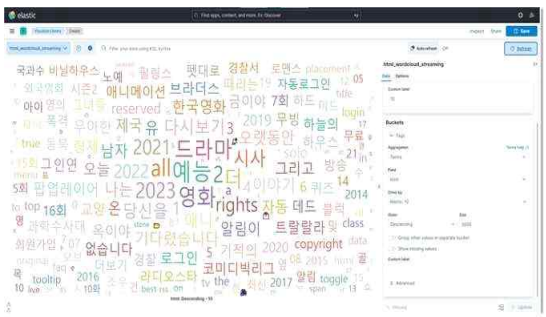

그림 2. 워드클라우드를 이용한 키워드 추출
Fig. 2. Keyword extraction using word cloud

또한 대상 IP에서 추출한 HTML 문서로부터 해당 내에 높은 빈도로 출연한 복수의 주요 키워드를 추출하고 이를 해당 IP에 대한 주요 키워드 값으로 검색엔진에 색인할 수 있다. 그림 2는 HTML 문서로부터 워드클라우드를 사용하여 키워드를 생성한 결과이다.

3.3 표적 사이트 정보 수집 단계

그림 3은 표적 사이트 정보 수집 단계 순서도이다. 표적 사이트 정보 수집 단계는 실제 타겟이 되는 불법사이트에 해당하는 도메인을 접속하여 불법 사이트에 대한 HTML 문서와 메인 페이지 캡쳐 이미지를 수집하고, HTML 문서에 복수의 주요 키워드를 추출하여 색인한다. 표적 사이트 정보 수집 단계에서는 불법 유통 플랫폼에서 추출한 HTML 문서로부터 해당 문서 내에 높은 빈도로 출현한 복수의 주요 키워드를 추출하고, 해당 불법 유통 플랫폼에 대한 주요 키워드값으로 검색엔진에 색인할 수 있다. '즉 불법 유통 플랫폼에서 추출한 HTML 문서를 전처리하여 텍스트들만 추출한 후에 텍스트의 출현 빈도를 반영하여 생성한 워드클라우드 내에서 상위 순위에 해당 키워드를 주요 키워드로 추출할 수 있다.

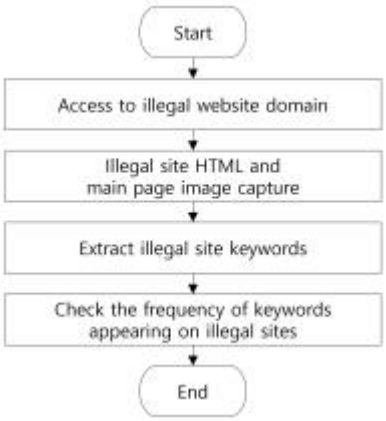

그림 3. 표적 사이트 정보 수집 단계 순서도
Fig. 3. Flowchart of target site information collection steps

3.4 IP 특정 단계

그림 4는 IP 주소 특정 단계 순서도를 표현하였다. IP 특정 단계에서는 불법 유통 플랫폼과 대상 IP 사이에, 주요 키워드, HTML 문서, 메인페이지 캡쳐 이미지의 유사도를 각각 비교 분석한다. 최

종적으로 동일한 사이트로 판단되면, 해당 IP를 불법 유통 플랫폼의 IP로 특정할 수 있다. IP 특정 단계에서 불법 사이트에 매칭된 주요 키워드를 검색엔진에 질의 키워드로 입력하여, 불법 유통 플랫폼과 적어도 하나의 주요 키워드가 일치하는 대상 IP를 탐색하여 검색 결과 리스트에 추가한다. 이와 같이 최소 하나의 키워드가 일치하는 IP를 실제 불법 유통 플랫폼과 관련성이 있는 유사 사이트 범주로 1차적으로 필터링할 수 있다.

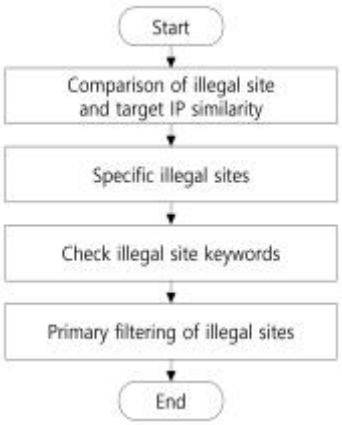

그림 4. IP주소 특정 단계 순서도
Fig. 4. Flowchart to specify IP address

4. 실험 및 결과

본 연구에 대한 실험 및 검증을 위해 다음 표 1과 같은 실험 환경을 구축하였다.

표 1. 실험 환경
Table 1. Experiment Environment

CPU	Intel(R) Xeon(R) Platinum 8259CL
GPU	NVIDIA Geforce RTX 4090
RAM	32GB
OS	Winsodws 10
SSD	100 GB

그림 5는 본 논문에서 제안한 원격지 서버의 IP 주소 특정을 위한 실험방법이고 그림 6은 실험 순서도이다.

한국인터넷진흥원에 등록된 약 1억 개의 활성화된 IP 중 80 Port와 443 Port를 대상으로 Archiving을 진행하였다. 표 2는 한국인터넷진흥원에 등록된 IP 현황이다.

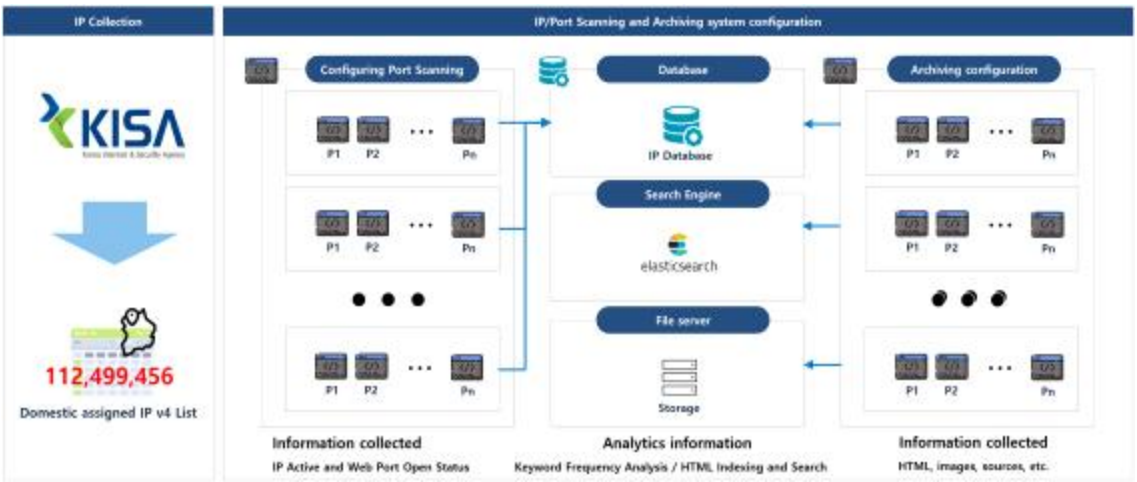

그림 5. IP주소 특정 실험방법
Fig. 5. Experiment method to specify IP address

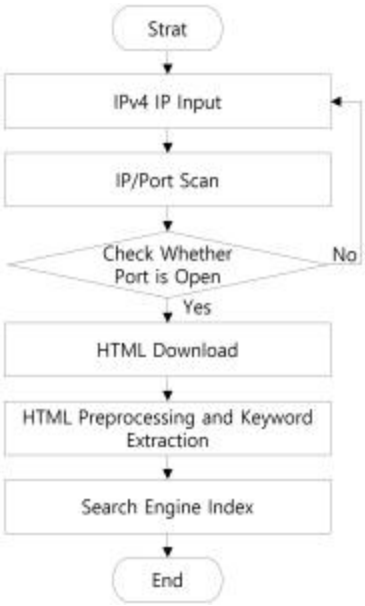

그림 6. IP주소 특정 방법 순서도
Fig. 6. Flowchart of IP address specification method

표 2. 사용중인 IP
Table 2. IP status in use

Type	Active Status
Total IP	112,499,456
Active IP	38,094,248
80 Port	1,373,342
443 Port	597,279

추출한 HTML 문서에서 영화 키워드를 입력하여 IP를 추출하였다. 추출한 IP를 워드클라우드를 사용하여 메인페이지를 확인하였다. 그림 7은 Archiving 데이터를 분석한 내용이다.

그림 8은 실제 DNS 접근을 위하여 HTML 내 사이트 주소로 의심되는 키워드를 추출하였다.

이렇게 추출한 키워드로 직접 접속하여 불법 유통 플랫폼 여부를 확인하였으며, 그림 9는 불법 유통 플랫폼을 직접 확인한 내역이다.

그림 7. Archiving 데이터 분석
Fig. 7. Archiving data analysis

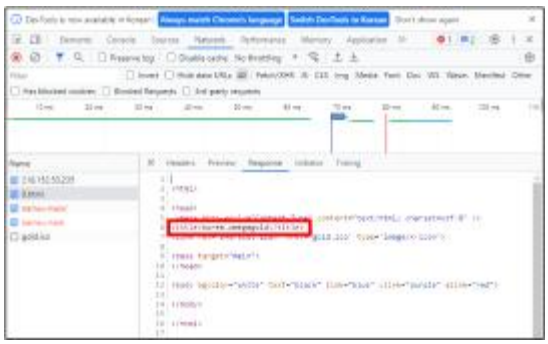

그림 8. 키워드 추출
Fig. 8. Keyword extraction

그림 9. 불법 사이트 여부 확인
Fig. 9. Check for illegal sites

그림 10은 추출한 IP와 Domain으로 접속하여 동일한 사이트인지 비교한 것으로 같은 사이트로 확인 할 수 있다.

그림 10. IP 접속과 Domain 접속 비교
Fig. 10. Comparison of IP connection and Domain connection

그림 11은 서비스 중인 불법 유통 플랫폼의 물리적인 서버 위치를 추정한 결과이다.

그림 11. ISP를 이용한 위치 추적
Fig. 11. Location tracking using ISP

5. 결 론

URL을 주기적으로 변경하거나 리다이렉션하는 등의 여러 방식으로 단속을 회피하고 있는 불법 유통 플랫폼의 물리적인 주소를 찾기 어려운 문제점을 해결하기 위해 본 논문에서는 IP 포트 스캐닝을 이용한 불법 유통 플랫폼 추적 방법을 제안하였다. 리버스프록시 등의 보안기술을 이용하여 IP를 은닉한 상태에서 서비스 중인 불법 어 유통 플랫폼 서버의 원 IP를 추적하고 특정하여 위치까지 추정하였다. 향후 불법 유통 사이트의 적극적인 차단과 수사와 추적에 용이할 것으로 보인다. 향후 불법 유통 플랫폼의 운영자 처벌 강화 및 불법 유통 플랫폼 생성 억제에 대한 연구가 필요할 것으로 보인다.

This research project supported by Ministry of Culture, Sport and Tourism(MCST) and Korea Creative Content agency(KOCCA) in 2022(R2022020109)

참 고 문 헌

[1] Y. S. Hwang, J. H. Han & S. J. Lee. (2022). Real IP address tracking techniques for illegal sites using Cyber Threat Intelligence search services. Journal of digital forensics, 16(2), 116-125. DOI : 10.22798/KDFS.2022.16.2.116

[2] H. G. Kang, H. H. Kim, H. S. Lee & S. J. Lee. (2017). Study on Collecting Server Information through Banner Grabbing. Journal of The Korea Institute of Information Security and Cryptology, 27(6), 1317-1330. DOI : 10.13089/JKIISC.2017.27.6.1317

[3] J. W. Choi, G. Y. Choi & S. J. Lee. (2023). Tracing Copyright Infringement Activities through Illegal Streaming Device Protocol Analysis. Journal of digital forensics, 17(2), 62-72. DOI : 10.22798/kdfs.2023.17.2.62

[4] C. H. Kim, H. J. Yu, S. Y. Kim & S. H Oh. (2022). Effecient Techniques to Block Copyright Infringement Illegal Streaming

Sites. Journal of The Korea Institute of Information Security and Cryptology, 32(5), 837-844. DOI : 10.13089/JKIISC.2022.32.5.837
[5] I. J. Yoo, J. C. Lee, B. C. Park, S. Y. Kim & Y. M. Kim. (2022). A Method for Generating Signature Information to Determine Illegal Distribution of Cloud-based Streaming Video. Journal of Software Assessment and Valuation, 18(2), 239-246. DOI : 10.29056/jsav.2022.12.24
[6] E. S. Choi, Y. M. Kim & M. C Park. (2023). Research on Methods of Feature Information Gathering for Identifying Illegal Copyright Infringement Sites. Journal of Software Assessment and Valuation, 19(3), 1-10. http://www.riss.kr/link?id=A108761284
[7] K. H Lee & G. B Kim. (2020). A study on the possibility of punishment of Portscan in criminal law. The Journal of Police Science, 20(1), 201-224. DOI : 10.22816/polsci.2020.20.1.007
[8] C. Wan & Y. D Kim. (2021). A Study on the Search and Seizure of User Information in Cloud Computing Service. Law, 70(3), 155-189. DOI: 10.17007/klaj.2021.70.3.005
[9] Y. C. Choi & S. J Lee. (2022). Analysis of site relevance through illegal webtoon site information. Journal of Digital Forensics Society, 16(1), 76-87. DOI : 10.22798/KDFS.2022.16.1.76

저 자 소 개

유인재(In-Jae Yoo)

2017.8 고려사이버대학교 소프트웨어공학 학사
2021.2 숭실대학교 컴퓨터학과 석사
2023.2-현재 숭실대학교 컴퓨터학과 박사 과정
2015.11-현재 ㈜비욘드테크 수석연구원
<주관심분야> 저작권 보호 및 이용활성화

이재청(Jae-Chung Lee)

1996.02 서울과학기술대학교 전자계산학과 학사
2012.07-현재 (주)비욘드테크 이사
<주관심분야> 저작권 보호 및 이용활성화

장세영(Seyoung Jang)

2018.2 평생교육원 학점은행 졸업
2021.6 숭실대학교 컴퓨터학과 석사
2023.2-현재 숭실대학교 컴퓨터학과 박사 과정
<주관심분야> 저작권 보호 및 이용활성화

박병찬(Byeongchan Park)

2015.2 평생교육원 학점은행 졸업
2018.2 숭실대학교 컴퓨터학과 석사
2023.8 숭실대학교 컴퓨터학과 박사
2023.9-현재 숭실대학교 초빙교수
<주관심분야> 저작권 보호 및 이용활성화

김석윤(Seok-Yoon Kim)

1980.2 서울대학교 전기공학과 졸업
1990.2 University of Texas at Austin
Dept. of ECE 석사
1993.8 University of Texas at Austin
Dept. of ECE 박사
1982-1987 ETRI 연구원
1993-1995 모토로라(Austin, Tx)
책임 연구원
1995-현재 숭실대학교 교수
<주관심분야> 시스템설계방법론, 저작권보호기술

김영모(Youngmo Kim)

2003.2 대전대학교 컴퓨터공학과 졸업
2005.2 대전대학교 컴퓨터공학과 석사
2011.2 대전대학교 컴퓨터공학과 박사
2012-현재 숭실대학교 교수
<주관심분야> 저작권 보호 및 이용활성화