ISSN(Print) : 1976-5304, ISSN(Online) : 2713-931X
https://kdfs.jams.or.kr, DOI : 10.22798/kdfs.2022.16.2.116

사이버 위협정보 검색 서비스를 활용한 불법사이트 실 IP 주소 추적 기법

황 예 성*, 한 재 혁**, 이 상 진***
고려대학교 정보보안학과 (대학원생)*, 고려대학교 정보보호대학원 (박사과정)**, (교수)***

Real IP address tracking techniques for illegal sites using Cyber Threat Intelligence search services

Yeseong Hwang*, Jaehyeok Han, Sangjin Lee*****
Dept. of Information Security, Korea University (Graduate Student)*,
School of Cybersecurity, Korea University (Doctor's Courser), (Professor)*****

요 약

최근 국내에서 미디어 시장의 규모가 커져가면서 미디어의 저작권을 위반하는 불법사이트의 수가 지속적으로 증가하고 있다. 이러한 증가 추세에 따라 국가에서 불법사이트를 근절시키기 위해 노력하고 있으나, 불법사이트는 URL을 지속적으로 변경하거나 리다이렉션(redirection)하는 등의 여러 방식으로 단속을 회피하고 있다. 하지만 근본적으로 불법사이트를 차단시키기 위한 정책을 진행하기 위해서는 서버의 물리적 위치를 발견해야 하는데 대부분의 불법사이트는 CDN (Content Delivery Network) 서비스를 사용하여 실 IP 주소를 감추고 있다. 본 연구에서는 사이버 위협정보 검색 서비스를 활용하여 불법사이트의 실 IP 주소를 획득하고, 이를 기반으로 Whois 정보 및 배너그랩 정보 등을 이용하여 불법사이트의 운영자를 추적하고 근본적으로 차단하는 방식을 제안한다. 제안한 기법을 이용하여 900개의 불법사이트를 정제하여 획득한 565개의 불법사이트를 대상으로 IP 주소를 수집한 결과 총 64개의 불법사이트 URL에 해당하는 789개의 IP 주소를 획득할 수 있었다. 따라서 사이버 위협정보 검색 서비스를 이용한 불법사이트의 실 IP 주소 획득으로 불법사이트의 근본적인 차단에 기여할 수 있을 것이다.

주제어 : 웹 크롤링, 불법사이트, 사이버 위협 정보(CTI), 콘텐츠 전송 네트워크(CDN), 배너그랩, IP 주소, Censys, Shodan, CriminalIP

ABSTRACT

Recently, the number of piracy sites has been continuously increasing as the size of the domestic media market has grown. As such, the government is putting effort to eradicate illegal sites, but these sites are avoiding crackdowns in various ways, such as changing or redirecting URLs. In order to crack down on the illegal sites, the physical location of the server must be identified, but most of illegal sites hide their actual IP addresses by using Content Delivery Network (CDN) services. In this, we propose a method of identifying the actual IP Address of illegal sites using Cyber Threat Intelligence search services. We also propose a method of tracking down the operator as well as shutting down the site by using an actual IP address. In our study, we collected IP addresses from 565 illegal sites selected from 900 illegal sites, and a total of 789 IP addresses corresponding to 64 illegal sites URLs were obtained. In this way, we expect the actual IP address of illegal sites by using the Cyber Threat Intelligence search service can contribute to the blocking of illegal sites.

Key Words : Web Crawling, Illegal Sites, Cyber Threat Intelligence(CTI), Contents Delivery Network(CDN), Banner grabbing, IP address, Censys, Shodan, CriminalIP

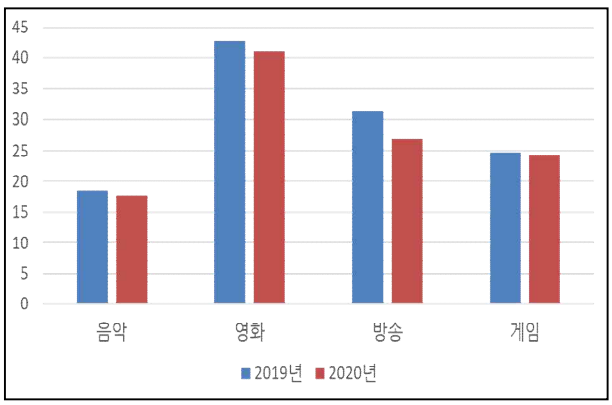
※ 이 논문은 2022년도 정부(문화체육관광부)의 재원으로 한국저작권보호원의 지원을 받아 수행된 연구임 (No 2021, 저작권 특화 디지털포렌식 전문 인력 양성사업)
- Received 14 February 2022, Revised 15 February 2022, Accepted 27 June 2022
- 제1저자(First Author) : Yeseong Hwang (Email : hys1734@gmail.com)
- 교신저자(Corresponding Author) : Sangjin Lee (Email : sangjin@korea.ac.kr)

Ⅰ. 서 론

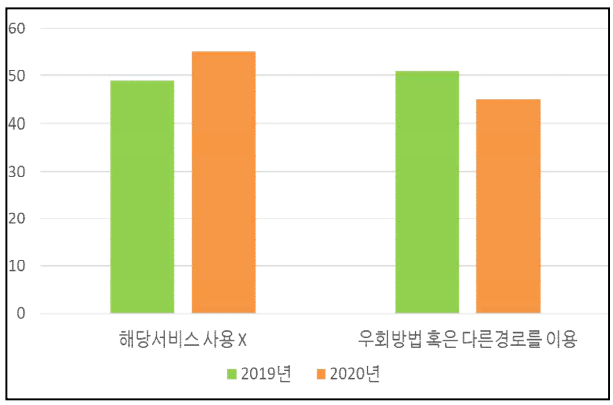
정보통신의 발전과 확산은 정치·경제·사회·문화 등 수많은 분야에 대한 접근을 편리하게 하였다. 특히 문화 분야에서 전 세계의 다양한 미디어 콘텐츠에 접근할 수 있는 기회가 생겼다. 그로 인해 해외 OTT 서비스(Over-The-Top Media service)를 활용한 영상 콘텐츠의 국내 유입이 지속적으로 진행되고 있으며, 웹툰의 국내 시장규모가 해를 거듭할수록 증가하고 있다 [1][2]. 반면에 미디어 시장규모가 커짐과 동시에 미디어 콘텐츠의 저작권을 위반하는 불법사이트가 기승을 부리고 있다. 이와 관련하여 한국저작권보호원에서 발표한 2021년 저작권 보호 연차보고서[3]에 따르면 2019년에 비해 2020년 불법복제물 이용률이 〈Figure 1〉과 같이 소폭으로 감소하였지만 불법복제물을 무단으로 유통하는 불법사이트 접속 차단에 대해서 해당 서비스를 이용을 하지 않는다는 답변과 다른 경로 또는 우회 방법을 찾는다고 답변한 사용자는 2019년도와 비교하였을 때 〈Figure 2〉와 같이 비슷하다. 이는 불법사이트에 대한 수요가 지속적으로 있음을 나타내며 근본적인 차단정책이 필요함을 알 수 있다. 이와 같은 현상을 근본적으로 차단하기 위해 불법사이트가 운영되고 있는 서버의 실 IP 주소를 획득하는 것이 필요하다. 하지만 대부분의 불법사이트는 CDN (Contents Delivery Network) 서비스를 이용하고 있어 불법사이트의 IP 주소를 조회하였을 때, 실 IP 주소 대신 CDN 서비스에 사용되는 IP 주소를 반환하고 있어 실 IP 주소 획득에 어려움을 겪고 있다.

이를 해결하기 위해 본 논문에서는 불법사이트의 HTML 및 내부 디렉토리 구조를 분석하고, 그 결과를 기반으로 사이버상에서 존재하는 모든 자원들의 데이터를 수집하여, 이를 기반으로 데이터를 가공하여 새로운 정보를 제공하는 것을 의미하는 사이버 위협정보 (Cyber Threat Intelligence; CTI)를 활용하여 불법사이트가 운영되고 있는 실 IP 주소를 추적하는 기법을 소개한다. 또한 기존 차단방식을 우회하며 운영되고 있는 불법사이트를 효과적으로 추적하기 위해서 여러 가지 CTI 검색 서비스에서 제공되는 정보를 비교하고 활용할 수 있는 방안을 제시하며, 불법사이트가 게시되어 있는 웹 서버의 실 IP 주소와 운영자 정보를 추적한 사례를 소개한다.

Ⅱ. 관련 연구 및 배경



〈Figure 1〉 The rate of use of illegal copies in 2020



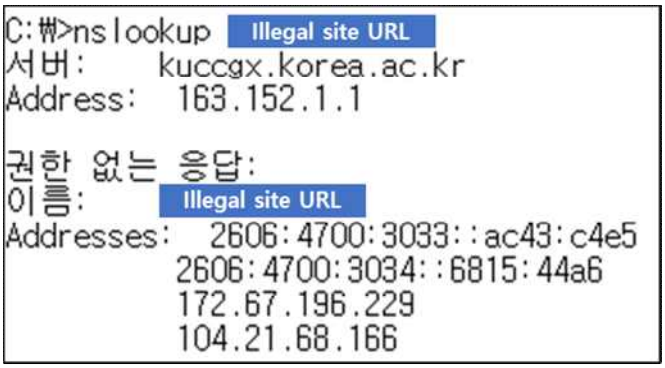
〈Figure 2〉 The effect of blocking illegal sites

2.1 불법사이트의 특징

불법사이트는 온라인상에서 유통되는 불법 정보 및 유해 정보를 포함하는 사이트로써, 불법사이트 유형으로는 저작권 위반 스트리밍, 웹툰, 음란물, 도박, 마약 거래 등이 있다. 국가에서는 불법 또는 유해정보를 갖고 있는 불법사이트로의 접속을 차단하기 위해 DNS (Domain Name Service) 차단 방식, SNI (Server Name Indication) 필드 차단 방식 등 여러 가지 차단 방식을 운영하고 있다. 하지만 불법사이트는 URL을 불규칙하게 변경하거나, 다양한 URL을 갖고 있어 기존의 DNS 차단 방식을 우회하여 계속해서 운영되고 있다.

또한 대부분의 불법사이트는 웹사이트에 사용되는 콘텐츠의 효율적인 전달을 위해 사용하는 CDN 서비스를 사용하고 있다. 〈Figure 3〉과 같이 CDN 서비스를 사용하는 불법사이트는 서버가 운영되고 있는 실 IP 주소를 반환하는 것이 아닌 CDN 서비스의 IP 주소를 반환하고 있다. 5개의 링크모음사이트를 시작점(Seed)으로 하여 해당 사이트에 게시되어 있는 불법사이트 URL을 카테고리 별 크롤링하여 900개의 불법사이트 URL을 대상으로 수집한 사이트를 대상으로 CDN 서비스의 사용 여부를 확인한 결과, 63개의 사이트는 실 IP 주소를

반환하고 837개의 사이트는 CDN 서비스의 IP 주소를 반환하고 있다 [4]. 이는 불법사이트 차단정책을 진행하기 위한 실 IP 주소 추적 관점에서 불법사이트의 IP 주소 획득을 위해 IP 주소를 조회하였을 때 실 IP 주소를 감추고 CDN 서비스의 IP 주소를 반환하여 실 IP 주소 추적이 어렵다는 것을 의미한다. 또한 불법사이트는 국내에 위치한 서버를 사용하는 것이 아닌 해외에 거점을 두고 있는 서버를 사용하고 있다. 따라서 국내 사법관할권이 미치지 않는 곳에 있기에 처벌 및 수사가 어렵다.



〈Figure 3〉 An example of search result retrieved by the nslookup command

2.2 배너그랩 기법을 이용한 정보 수집

네트워크에 연결된 시스템의 각 포트에 사용하는 서비스 정보를 획득하는 기법으로 알려진 배너그랩은 사이버 위협정보 수집 연구에 사용되고 있다. 전 세계 네트워크를 대상으로 배너그랩을 이용하여 시스템 정보를 수집하는 Shodan과 Censys의 스캔 방식을 비교하고 효과적인 포트 스캔 방식을 제시하였다 [5]. 또한 배너그랩을 이용하여 시스템 정보를 수집하는 크롤링 (crawling) 도구를 이용하여 취약점이 있는 서버들을 파악하고 수많은 사용자들이 공격에 노출된 것을 확인한 후, 이를 기반으로 국가별 위험 지표를 도출하고 앞으로의 위협을 예측하였다 [6]. [5][6]에서는 배너그랩을 사용하여 효율적인 사이버 위협정보를 수집하는 방법을 제시하였지만, 상용되고 있는 사이버 위협정보 서비스와 비교하였을 때 지표 및 방법론만 제시되었을 뿐, 수집한 데이터의 활용 방안에 대해서 제시하지 못했다. 본 연구에서는 전 세계 네트워크를 대상으로 사이버 위협정보 데이터를 제공하는 서비스를 비교하고 불법사이트의 실 IP 주소 추적에 적합한 데이터를 제공하는 서비스를 이용하여 불법사이트 차단을 위한 정보를 수집 방안을 제시한다.

2.3 사이버 위협정보 검색 서비스 비교

CTI는 사이버 상에서 수집할 수 있는 모든 자원들의 정보로써, 사전에 수집한 사이버 위협정보를 가공하여 유의미한 데이터 셋을 사용자에게 제공한다. 이 외에도 배너그랩, Whois 정보, 특정 URL의 출처를 추적하기 위한 정보를 획득하는데 사용되고 있다. 이를 기반으로 불법사이트의 특징을 이용하여 실 IP 주소 획득 및 운영되고 있는 서버의 위치 등 불법사이트 차단에 사용될 수 있는 유의미한 정보를 얻을 수 있다. 사이버 위협정보 서비스를 이용하여 불법사이트의 실 IP 주소 추적 및 차단에 활용될 수 있는 정보는 서버가 운영되고 있는 회사 이름과 Local Address 등을 식별할 수 있는 Whois 정보, DNS에서 URL과 매칭되는 실 IP 주소, 서버에 열려있는 포트의 서비스 정보를 알려주는 배너그랩 정보, HTTPS 통신에서 활용되는 인증서 정보 등의 데이터를 활용할 수 있다. 본 절에서는 Shodan, Censys, CriminalIP 3개의 서비스를 소개하고, 사이버 위협정보에서 사용되는 주요 지표 4가지를 〈Table 1〉과 같이 비교하였다.

Shodan은 IP 주소를 기반으로 전 세계 네트워크를 검색하여 인터넷에 연결된 서버, 인터넷 전화, IoT (Internet of Things) 기기 등의 호스트 정보를 수집하고 이를 가공하여 사용자에게 제공한다. 특히, 특정 포트에 사용되고 있는 서비스 정보, 호스트 OS 버전, 웹 엔진의 버전 등을 수집하는 기술로 알려진 배너그랩에 특화되어 있는 서비스이다 [7]. Censys는 인터넷에 연결된 호스트 및 네트워크 정보를 조회하여 보안 취약점을 가진 시스템을 찾아주는 검색 엔진이다. 네트워크 스캔에 최적화된 Zmap 방식을 활용하여 전 세계 IP 주소를 대상으로 자주 사용되는 포트는 1일 단위로, 자주 사용되지 않는 포트는 10일 단위로 스캔하여 사용자 검색 시점에 IP 주소의 상태 정보를 제공한다 [8]. CriminalIP는 전 세계 IP 주소를 대상으로 배너그랩을 이용하여 최근 2주 이내의 데이터를 주기적으로 수집하고 해당 데이터를 기반으로 IP 주소, 도메인, 취약점 등의 키워드를 활용하여 사용자가 얻고자 하는 정보를 제공하는 검색엔진 서비스이다 [9].

각 사이버 위협정보 서비스마다 제공하는 데이터와 특징이 다르다. 따라서 사이버 위협정보 검색 서비스의 특징을 1) 배너그랩 정보, 2) 실 IP 주소 정보, 3) Whois 정보, 4) 인증서 정보, 총 4가지 종류의 정보가 제공되는지 여부를 기준으로 비교하고, 불법사이트의 실 IP 주소 추적에 적합한 서비스 및 운영자 추적에 필요한 Whois 정보 및 배너그랩 정보를 검증할 수 있는 서비스를 선정하였다. 먼저 네트워크에 연결된 시스템의 열려있는 포트에 운영 중인 서비스 정보를 획득할 수 있는 배너그랩 정보를 획득할 수 있는지 비교해보면, Shodan, Censys, CriminalIP 3개의 서비스 모두에서 획득할 수 있었다. 다만, 각 서비스마다 네트워크를 스캔하는 방식 및 주기가 다르기에 제공하는 배너그랩 정보는 일부 상이하였다. URL나 키워드를 검색할 경우에 실 IP 주소의 획득 여부를 살펴보면, Shodan에서는 URL 검색을 지원하지 않아 다른 검색 옵션으로는 실 IP 주소 획득이 불가능했다. CriminalIP의 경우 URL 검색 결과 CDN 서비스에 활용되는 IP 주소를 반환하므로 실 IP 주소를 획득할 수 없었다. 하지만 Censys에서 URL 검색을 진행한 결과, CDN 서비스의 IP 주소가 아닌 실 IP 주소를 반환하는 것을 확인하였다. 웹사이트가 운영되는 위치 및 회사 이름을 식별할 수 있는 Whois 정보는 모든 서비스에서 사용할 수 있었다. 사이버 위협정보 검색 서비스마다 제공하는 Whois 정보에서 상이한 부분이 있으므로 다수의 서비스를 비교해서 분석할 필요가 있다. 웹사이트 간 암호화된 연결을 위해 사용되는 인증서 정보는 모든 서비스에서 획득할 수 있다. 사이버 위협정보 검색 서비스 특징을 비교한 결과를 기반으로 불법사이트의 실 IP 주소 추적에 적합한 검색 서비스와 검증에 필요한 검색 서비스를 선정하기 위해 URL 검색 가능 서비스, IP 주소 검색 가능 서비스로 분류한다. 실 IP 주소 추적에 적합한 검색 기능은 URL 검색 가능 서비스이기에, 서비스를 사용하여 실 IP 주소를 획득하고 IP 주소 검색 가능 서비스는 실 IP 주소를 이용하여 획득한 정보를 검증하는 형태로 사용한다.

4개의 지표로 각 서비스를 비교한 결과는 〈Table 1〉과 같이 정리할 수 있다. 불법사이트의 실 IP 주소 추적에 적합한 사이버 위협정보 검색 서비스는 URL 검색만으로 4가지 지표를 모두 획득할 수 있는 Censys로 확인되었다. 또한 Shodan과 CriminalIP는 실 IP 주소를 이용하여 Whois 정보 및 배너그랩 정보 검색이 가능하므로 Censys의 Whois 정보 빛 배너그랩 결과와 비교 분석 및 검증에 적합한 서비스로 선정할 수 있었다.

〈Table 1〉 Comparative features of CTI services for the Real IP address tracking

URL Searching service				
Services	Banner grabbing info	Type of IP addr.	Whois info	Certificate Info
Censys	Y	Real IP addr.	Y	Y
CriminalIP	Y	CDN IP addr.	Y	Y

IP Address Searching Service				
Services	Banner grabbing info	Type of IP addr.	Whois info	Certificate Info
Censys	Y	Real IP addr.	Y	Y
CriminalIP	Y	CDN IP addr.	Y	Y
Shodan	Y	-	Y	Y

Ⅲ. 불법사이트 운영자 정보 추적 프로세스

3.1 불법사이트 운영자 추적 필요성 및 프로세스 소개

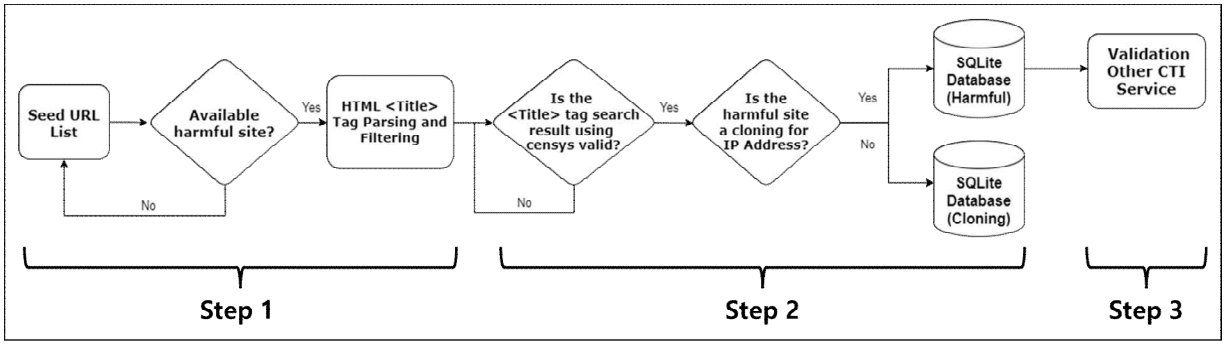
사용자들은 특정 웹페이지에 접근하기 위해 IP 주소를 직접 입력하기보다 URL을 입력하는 경향이 있는데, URL을 IP 주소로 변환시키는 역할은 DNS (Domain Name Server)가 수행한다 [10]. DNS에 "nslookup", "ping", "tracert"와 같은 명령어를 동작시키면 URL과 매칭된 IP 주소를 얻는 것이 일반적이다. 반면에 불법사이트는 웹에 사용되는 콘텐츠들을 효율적으로 전달하기 위해 여러 노드를 가진 네트워크에 콘텐츠를 저장하여 제공하는 CDN을 통해 웹서비스를 제공한다. CDN을 사용하고 있는 웹사이트는 서버의 IP 주소를 나타내는 것이 아니라, CDN 서비스에서 할당된 IP 주소를 나타낸다. 따라서 위의 명령어를 사용하면 웹사이트의 실 IP 주소를 획득할 수 없고 CDN 서비스에 사용되는 IP 주소가 반환되어 실제 웹 서버가 운영되고 있는 IP 주소와 다른 것을 알 수 있다.

불법사이트 특징에서 설명한 것처럼 대다수의 불법사이트는 CDN 서비스 중 하나인 Cloudflare를 사용하고 있다. 따라서 "nslookup", "ping", "tracert" 3가지 명령어로 불법사이트의 IP 주소를 획득하면 〈Table 2〉의 IP 주소 중 하나를 얻게 된다 [11]. 따라서 CDN 서비스를 사용하지 않는 일부 불법사이트를 대상으로는 실 IP 주소 값을 획득할 수 있으나, CDN 서비스를 사용하는 불법사이트는 추적이 불가능하다. 따라서 CDN 서비스를 이용하는 불법사이트의 숨겨진 실 IP 주소를 획득하기 위한 방법이 필요하다.

이에 본 논문에서는 CDN을 사용하는 불법사이트 웹서버의 실 IP 주소를 획득하고 이를 기반으로 불법사이트 차단정책을 진행하기 위한 운영자 추적 3단계 과정 기법을 〈Figure 4〉와 같이 제시한다. 가장 먼저 불법사이트의 도메인 이름을 수집하고, 현재 불법사이트의 접속가능 여부를 판단한다. 그리고 사이버 위협정보 검색 서비스에 활용할 수 있도록 수집된 불법사이트 정보 관련 데이터를 재구성한다 (Step 1). 이후 사이버 위협정보 검색 서비스 Censys 검색 옵션과 정제된 데이터를 이용하여 불법사이트의 실 IP 주소 획득 및 불법사이트의 운영자 추적과 관련된 Whois 및 배너그랩 정보를 획득하고 국내와 해외에서 운영되고 있는 IP 주소를 분류하고 획득한 IP 주소로 불법사이트에 접근하였을 때 Cloning 사이트인지 실제 불법사이트인지 비교 분석한다 (Step 2). 마지막으로 획득한 불법사이트의 실 IP 주소를 이용하여 다른 사이버 위협정보 검색 서비스에서 제공하는 데이터 비교 분석 및 검증을 통해 불법사이트의 실 IP 주소 및 운영자 관련 정보를 도출하여 불법사이트의 근본적 차단에 사용한다 (Step 3).

〈Table 2〉 Examples of Cloudflare's IP address version 4 and 6

IP address v4			IP address v6	
103.21.244.0/22	103.22.200.0/22	103.31.4.0/22	2400:cb00::/32	2606:4700::/32
104.16.0.0/13	104.24.0.0/14	108.162.192.0/18	2803:f800::/32	2405:b500::/32
131.0.72.0/22	141.101.64.0/18	162.158.0.0/15	2405:8100::/32	2a06:98c0::/29
172.64.0.0/13	173.245.48.0/20	188.114.96.0/20	2c0f:f248::/32	
190.93.240.0/20	197.234.240.0/22			



〈Figure 4〉 Process of tracking operators of illegal sites using Cyber Threat Information

3.2 불법사이트 연결 URL 후보군 선별: Step 1

불법사이트 URL 목록은 5개의 링크모음사이트를 시작점(Seed)으로 하여 해당 사이트에 게시되어 있는 불법사이트 URL을 카테고리 별 크롤링하여 900개의 불법사이트 URL 후보군을 수집하였다. 수집된 불법사이트 연결 URL을 Censys에 사용되는 검색 옵션 중 HTML 검색에 활용할 수 있도록 재구성하는 것이 필요하므로 불법사이트 URL 후보군을 정제하는 과정이 필요하다. 먼저 수집 시점의 불법사이트 접속 가능 여부를 확인하고 접속되지 않는 불법사이트들을 후보군에서 제거하는 1차 정제작업을 진행하였다 [4]. 다음으로 Censys HTML title tag 검색 옵션을 활용하기 위해, 정제된 불법사이트 URL 후보군을 대상으로 사이버 위협정보 검색 서비스에 활용되는 HTML 내 〈Title〉 Tag를 크롤링하여 획득한다. 획득한 결과를 대상으로 서비스에 정상 접속하였을 때 나타내는 〈Title〉 Tag의 특징인 Cloudflare, Cloudflare 에러 메세지를 알려주는 521: Web Server is Down, 522: Connection Timeout 등의 문자열과 웹사이트 비정상적인 접근을 알려주는 403 Forbbiden, 404 Not found와 같은 상태코드 문자열을 필터링 리스트로 구성하여 불법사이트 후보군을 정제하였다. 모든 정제과정을 진행한 결과, 정제된 불법사이트 URL 후보군의 갯수는 총 565개를 수집할 수 있었다.

3.3 Censys 검색을 통한 실 IP 주소 수집 및 검증: Step 2

정제된 불법사이트 URL 후보군을 대상으로 Censys에서 제공하는 Python 버전 API의 다양한 검색 쿼리 옵션 중 HTTP와 관련된 3가지 옵션을 이용하여 불법사이트의 실 IP 주소 획득을 시도하였다. 검색 시 반환되는 페이지의 수를 설정하는 per page 항목은 다양한 결과를 얻기 위해 maximum 값인 100으로 설정하여 검색하였다.

첫 번째 검색옵션은 HTML body에 있는 값들을 검색하는 〈services.http.response.body〉를 이용하여 불법사이트가 갖고 있는 특정 값을 검색하였다. 검색 결과 불법사이트 외 링크모음사이트 및 주소안내사이트 등 다른 웹사이트 내 존재하는 불법사이트의 title 값이 포함되어 함께 결과로 반환하기에 불법사이트 후보군에 해당하는 실 IP 주소 획득 자동화가 불가능하였다. 다음으로 HTML Tag 내부에 있는 값을 검색하는 〈services.http.response.html tags〉을 이용하여 불법사이트가 갖고 있는 특정 HTML Tag 값을 검색하였다. 검색 결과, 〈services.http.response.body〉 옵션 검색과 유사하게 불법사이트 외 링크모음사이트 및 주소안내사이트 내 존재하는 불법사이트 값을 결과로 반환하기에 검색한 불법사이트에 해당하는 실 IP 주소 획득 자동화가 불가능하였다. 마지막으로 HTML 〈Title〉 Tag 검색에 활용되는 쿼리 〈services.http.response.html title〉 검색 옵션을 이용하였다. 위의 두 가지 검색 옵션과는 다르게, 불법사이트의 〈Title〉 Tag를 특정하여 검색하였기에 특정 불법사이트의 해당하는 실 IP 주소를 획득할 수 있었다. 위 결과를 바탕으로 〈services.http.response.html title〉검색 옵션을 자동화하여 정제된 불법사이트 URL 후보군을 대상으로 실 IP 주소 획득을 시도하였다.

2022년 2월 8일을 기준으로 불법사이트 실 IP 주소 획득 결과, 총 64개의 불법사이트 URL에 해당하는 789개의 IP 주소를 획득할 수 있었다. DNS 특성상 1개의 URL에 다수의 IP 주소가 할당될 수 있기 때문에 불법사이트의 URL보다 훨씬 많은 IP 주소를 획득하였다. 획득한 IP 주소를 대상으로 Censys가 제공하는 데이터 셋을 이용하여 Whois 정보를 조회하였다. 이를 국가별로 분류한 결과, 총 34개의 국가가 확인되었고 해외에 위치하는 IP 주소로 식별된 불법사이트는 48개, 국내에 위치하는 IP 주소로 식별된 불법사이트는 총 16개가 발견되었다. 이 중 국내에 위치하는 IP 주소로 식별된 불법사이트를 대상으로 IP 주소에 HTTPS 혹은 HTTP 프로토콜로 웹사이트에 직접 접근한 결과, 실제로 접근되는 불법사이트가 있는 반면에 불법사이트와 유사한 형태를 띄는 가짜 불법사이트, 즉 Cloning 사이트로 접근하는 사례가 발견되었다.

먼저 Censys를 이용하여 획득한 IP 주소를 이용하여 웹사이트에 접근한 결과 불법사이트의 실 IP 주소임을 확인했다. URL로 접근하였을 때와 실 IP 주소로 접근하였을 때 HTML 구조를 비교한 결과 동일한 구조를 갖고 있음을 확인했다. 또한 웹 프록시를 설정하여 네트워크 패킷을 확인한 결과, 서버를 구성하는 엔진과 웹 구성 언어가 모두 동일한 것을 확인했다. 마지막으로 IP 주소로 불법사이트에 접근하여 로그인, 회원가입 등의 다양한 기능에 접근하였을 때 불법사이트의 URL로 변경되는 것으로 확인되었다.

다음으로 발견된 IP 주소 중 국내 IP 주소를 사용하는 불법사이트에 접근한 결과, 일부 경우는 실제 불법사이트의 IP 주소가 아닌 Cloning 사이트의 IP 주소가 나타나는 경우를 발견하였다. 불법사이트의 Cloning 여부를 확인한 방식은 여러가지가 있었다. new****.com 대상으로 Cloning 사이트 여부를 비교한 결과, 실제

〈Table 3〉 Tracking results illegal sites IP addresses in South Korea

Process					Findings
Subject	URL	Real IP addr.	Whois info.	Cloning sites	
Step 1	Given URLs (Illegal sites by crawling) : 900	-	-	-	**The number of URLs** - Before refining: 900 - After refining: 565
Step 2	Before Tracing : 565 After Tracing : 64 (Censys searching)	▶Tracing Result : 789 *Repeated IP addresses are included.	**S, Korea: 78** Overseas: 711 (Based IP addr.)	-	**The number of real IP addr.** - Tracing result: 78
Step 3	-	-	For S.Korea - Before Validation : 78 - After Validation : 63	Validated Results : 16 - is_Cloning? 8 - Real IP addr.? 8	**Illegal sites with real IP addr.** - Before Validation: 64 (S. Korea: 16, Overseas: 48) **- After Validation: 8** (S. Korea: 8)

new****.com의 경우 Nginx + PHP으로 구성된 것을 확인했다. 반면에 Cloning 사이트로 판단되는 IP 주소로 판단되는 웹사이트의 경우, Express + node.js으로 구성된 것을 확인했다. 이는 웹사이트를 구성하는 엔진의 특성으로 추가 검증이 가능하였다. Apache와 Nginx를 사용하는 웹사이트의 경우 Host를 바꿔주면 내부의 VirtualHost 설정에 따라서 라우팅이 되는데, 불법사이트를 Cloning한 웹사이트는 Host value를 랜덤한 값으로 변경하여도 정상 작동하는 것으로 확인되어 clone 여부를 판단할 수 있었다. 또한 IP 주소로 접근한 Cloning 사이트는 JS코드 내 실제 URL로 접근한 new****.com에 존재하지 않는 setTimeout() 함수가 존재하여, 일정 시간이 지나면 특정 유튜브 영상으로 redirection되는 코드가 확인되었다. 마지막으로 불법사이트를 Cloning한 웹사이트의 IP 주소에는 new****.com 뿐이 아니라 week****.com, aki***.cc, met***.com라는 도메인 주소도 호스팅 하는 것으로 확인되어 불법사이트를 Cloning한 사이트 임을 확인할 수 있었다. 따라서 Censys를 활용하여 불법사이트의 IP 주소를 획득하였어도 실 IP 주소가 아닌 Cloning한 웹사이트의 IP 주소가 획득될 수 있다.

3.4 CTI 내 Whois 정보 비교를 통한 실 IP 주소 정보 획득: Step 3

3.2절에서 Censys의 HTML 검색 옵션을 이용하여 불법사이트의 IP 주소 리스트 및 Whois 정보를 획득했다. IP 주소 획득은 Censys에서만 가능하여 하나의 서비스에 의존할 수 밖에 없다. 하지만, Shodan과 CriminalIP를 이용하면 획득한 IP 주소의 운영자 추적에 용이한 Whois 정보 검색 및 배너그랩 정보 획득이 가능하다. 따라서 운영자 추적의 정확도를 기존 결과보다 높게 얻기 위해 Shodan API의 IP 주소 검색 옵션과, CriminalIP API의 IP 주소 검색 옵션을 이용하여 Censys에서 획득한 IP 주소를 검색했고, 두 서비스 검색 결과를 기반으로 각 서비스 별 Whois 정보 및 배너그랩 정보 교차 검증을 했다. Censys에서 획득한 국내에 위치한 IP 주소 78개를 기반으로 하여 Whois 정보를 각 서비스 별로 비교한 결과, Shodan은 Censys 검색 결과와 73개가 동일한 것으로 확인되었고, CriminalIP는 63개의 정보가 동일한 것으로 확인되었다. Shodan과 CriminalIP의 Whois 정보 검색 결과가 Censys의 결과와 다른 값들을 대상으로 비교 분석한 결과, Shodan은 1개의 IP 주소에 대한 정보가 없는 것으로 확인되었고, 4개의 IP 주소에 대한 국가 정보는 동일하나, 운영 중인 회사 정보가 없는 것으로 확인되었다. CriminalIP는 15개의 IP 주소에 대한 국가 정보가 모두 한국이 아닌 미국으로 확인되었다. 따라서, 같은 IP 주소를 갖고 있어도 사이버 위협정보 검색 서비스마다 IP 주소를 스캔하는 주기와 방식이 달라 각 서비스가 제공하는 정보가 다른 것을 확인했다. 각 사이버 위협정보 서비스 Whois 검색을 비교 분석하여 공통적인 결과를 취합하였을 때, 국내에 위치하는 IP 주소 총 63개를 획득하였다. 불법사이트를 대상으로 〈Figure 4〉의 각 프로세스 과정을 분석한 결과 〈Table 3〉의 결과를 얻을 수 있었다. 연구 결과를 살펴보면 불법사이트 추적으로 끝나는 것이 아니라 결과를 기반으로 타 사이버 위협정보 검색 서비스 정보와 비교 분석 및 검증하여 오탐을 최소화 하는 것이 필요하다.

Ⅳ. 결 론

국내 수사기관의 불법사이트 차단 노력에도 불구하고 불법사이트의 CDN 서비스 사용, 예측할 수 없는 주기의 URL 변경 등 기존의 차단 방법을 우회하며 운영되고 있어 불법사이트 수사에 어려움을 겪고 있다. 본 논문에서는 사이버 위협정보 검색 서비스를 사용하여 CDN 서비스로 감춰진 불법사이트의 실 IP 주소를 추적하는 방안을 제시하였다. 또한 획득한 불법사이트의 실 IP 주소를 사용하여 불법사이트가 운영되고 있는 서버의 위치 및 운영자 정보를 나타내는 Whois 정보, 불법사이트가 운영되고 있는 서버의 포트에서 운영되고 있는 서비스의 정보를 나타내는 배너그랩 정보를 획득하였다.

이를 기반으로 전체 불법사이트 URL 900개를 정제하여 획득한 565개를 검색한 결과, 64개의 불법사이트의 실 IP 주소가 확인되었고, 64개를 대상으로 Whois 정보를 조회한 결과, 국내에 위치한 IP 주소로 식별된 불법사이트는 16개, 해외에 위치한 IP 주소로 식별된 불법사이트는 48개임을 확인했다. 국내에 위치한 IP 주소를 대상으로 실 IP 주소로 접근해본 결과, 실제 불법사이트의 IP 주소가 아닌 불법사이트를 Cloning한 사이트의 IP 주소도 발견했다. 또한 Censys 검색을 통해 획득한 국내 IP 주소를 타 사이버 위협정보 서비스와 비교 분석 및 검증한 결과 총 8개의 국내 불법사이트의 실 IP 주소를 획득했다.

위의 결과처럼 국내에 위치하고 있는 IP 주소를 기반으로 운영되는 불법사이트의 실 IP 주소를 획득하였을 때, 국내에서 차단 및 처벌이 가능하다. 하지만 해외에 위치하고 있는 IP 주소를 기반으로 운영되는 불법사이트는 국내 사법관할권이 미치지 않기에 차단 및 처벌이 어렵다. 불법사이트의 실 IP 주소 조회 결과, 해외에 위치하는 IP 주소 갯수가 국내에 위치하는 IP 주소 갯수보다 훨씬 많은 것을 확인할 수 있다. 따라서 국가 간의 국제 공조 수사를 통해 불법사이트를 차단하려는 움직임이 필요하다. 또한 본 연구에서 사용한 Censys 쿼리 검색 외 다른 쿼리 옵션 검색 및 획득한 데이터를 사용하여 수집하지 못한 불법사이트의 실 IP 주소를 획득하는 연구와 불법사이트 Cloning 여부를 판별하는 자동화된 기술 연구가 필요하다.

참 고 문 헌 (References)

[1] Sanghyeok Han, "Evaluation of competition in the broadcasting market in 2020", Korea Communications Commission, 2020.

[2] Youngjun Kim, "Cartoon Industry Withe Paper 2020", Korea Creative Content Agency, 2020

[3] Korean Copyright Commission, "2021 Copyright Protection Annual Report", 2021.

[4] S. Choo, Y. Hwang, and S. Lee, Methods for Collecting illegal Websites Using Web Crawling, Journal of Digital Forensics 15 (2021) 127 - 138.

[5] S. Im, S. Shin, B. Roh, and J. Lee, Scan Modeling and Performance Analysis for Extensive Terminal Information Identification, The Journal of Korean Institute of Communications and Information Sciences 42 (2017) 785 - 790.

[6] H. Kang, H. Kim, H. Lee, and S. Lee, Study on Collecting Server Information through Banner Grabbing, Journal of The Korea Institute of Information Security and Cryptology 27 (2017) 1317 - 1330.

[7] "What is Shodan?", Shodan, Available: https://help.shodan.io/the-basics/what-is-shodan, 2022.07.07. confirmed.

[8] "Censys Internet Scanning Intro", Censys, Available: https://support.censys.io/hc/en-us/articles/360059603231, 2022.07.07. confirmed.

[9] "CriminalIP GLOBAL BETA SERVICE OPEN", CriminalIP, Available: https://security.criminalip.com/, 2022.07.07. confirmed.

[10] "Domain Name System (DNS)", Microsoft, Available: https://docs.microsoft.com/en-us/windows-server/networking/dns/dns-top, 2022.07.07. confirmed.

[11] "nslookup", Microsoft, Available: https://docs.microsoft.com/en-us/windows-server/administration/windows-commands/nslookup, 2022.07.07. confirmed.

저 자 소 개

황 예 성 (Yeseong Hwang)
준회원

2014년 3월 ~ 2020년 8월: 고려대학교 컴퓨터 정보학과 졸업
2020년 9월 ~ 현재: 고려대학교 정보보호대학원 석사과정 재학
관심분야 : 디지털 포렌식, 침해사고대응, 사이버 보안

한 재 혁 (Jaehyeok Han)
준회원

2011년 2월: 서울시립대학교 수학과 졸업
2016년 2월: 고려대학교 정보보호대학원 정보보호학과 공학석사
2016년 3월 ~ 현재: 고려대학교 정보보호대학원 정보보호학과 박사과정
관심분야 : 디지털 포렌식, 정보은닉, 파일시스템

이 상 진 (Sangjin Lee)
평생회원

1989년 10월 ~ 1999년 2월 : ETRI 선임 연구원
1999년 3월 ~ 2001년 8월 : 고려대학교 자연과학대학 조교수
2001년 9월 ~ 현재 : 고려대학교 정보보호대학원 교수
2008년 3월 ~ 현재 : 고려대학교 디지털포렌식연구센터 센터장
2017년 3월 ~ 현재 : 고려대학교 정보보호대학원 원장
관심분야 : 디지털 포렌식, 심층암호, 해시함수