한국컴퓨터정보학회 동계학술대회 논문집 제32권 제1호 (2024. 1)

마약 범죄 추적을 위한 다크웹 상의 소셜미디어 유인 링크 수집체계 개발

박솔규[O], 김지연[*], 김창훈[*]
[O]대구대학교 컴퓨터소프트웨어학과,
[*]대구대학교 컴퓨터정보공학부
e-mail: {sol2677[O], jyk[*], kimch[*]}@daegu.ac.kr

Development of a Collection System of Bait Links to Social Media on Dark Web to Track Drug Crimes

Sol-Kyu Park[O], Jiyeon Kim[*], Chang-Hoon Kim[*]
[O]Dept. of Computer and Software Engineering, Daegu University,
[*]Division of Computer and Information Engineering, Daegu University

● 요 약 ●

다크웹(Dark Web)은 마약, 불법 촬영물, 해킹, 무기 등 불법 콘텐츠의 공유 및 거래가 이루어지는 인터넷 영역으로서 최근에는 소셜미디어와 연계된 형태로 범죄 양상이 변화하고 있다. 본 논문에서는 최근 국내외 사회 문제로 대두되고 있는 마약 범죄를 추적하기 위한 다크웹 수사 기술로서 다크웹 사용자를 소셜미디어로 유인하는 마약 정보 수집체계를 개발한다. 먼저 미국 마약단속국에서 공개한 대표적인 마약 용어 3개의 표준어 및 은어를 검색 키워드로 사용하여 마약 관련 다크웹을 수집하고, 수집된 다크웹을 크롤링하여 소셜미디어 계정 링크를 추출한다. 본 논문에서는 다양한 소셜미디어 중, 트위터 및 텔레그램 접속 링크를 수집하였으며 실험 결과, 접속 가능한 총 54개 다크웹 도메인의 9,046개 웹 페이지에서 트위터 유인 링크 567개, 텔레그램 유인 링크 118개를 추출하였다.

키워드: 다크웹(Dark Web), 소셜미디어(Social Media), 마약 유통(Drug Distribution), 크롤링(Crawling)

I. 서론

인터넷은 사용자의 접근성 및 가시성에 따라 크게 표면 웹(Surface Web), 딥웹(Deep Web), 다크웹(Dark Web)으로 구분된다. 표면 웹은 구글, 네이버와 같은 일반 검색엔진에 색인 되는 영역으로 전체 웹의 약 5%에 해당되고, 딥웹은 인증을 통해서만 접근이 가능한 메일, 데이터베이스 등의 영역으로 일반 검색엔진에 색인 되지 않는다는 특징이 있다. 딥 웹 일부에 해당되는 다크웹은 크롬(Chrome), 엣지(Edge), 사파리(Safari) 등과 같은 일반 웹 브라우저로는 접근이 어려운 영역으로 주로 불법적인 콘텐츠 및 서비스 거래를 목적으로 운영된다. 다크웹은 다중 암호화를 통해 어니언(Onion) 도메인을 라우팅할 수 있는 토르(The Onion Router), I2P(Invisible Internet Project)와 같은 전용 브라우저를 사용하여 접속 가능하다.

과거에는 다크웹의 강력한 익명성을 악용하여 마약, 불법 촬영물, 무기 등 불법 콘텐츠 및 제품이 다크웹에서 직접 거래 되었지만, 전 세계 소셜미디어 사용자가 증가하면서 다크웹과 소셜미디어가 연계된 형태로 범죄 양상이 변화하고 있다. 특히, 최근 소셜미디어를 통해 확산되고 있는 마약 유통의 경우, 국내외에서 심각한 사회문제로 대두되고 있다.

본 논문에서는 다크웹과 소셜미디어에 걸쳐 발생하는 마약 범죄를 추적하기 위하여 다크웹 사용자를 소셜미디어로 유인하는 정보 수집 기술을 개발한다. 먼저 미국 마약단속국에서 공개한 대표적인 마약 용어를 검색 키워드로 사용하여 마약 관련 다크웹을 'Torch' 검색엔진에서 수집하고, 수집된 각 다크웹을 크롤링하여 소셜미디어 관련 링크를 추출한다. 본 논문에서는 다양한 소셜미디어 중, 트위터 및 텔레그램 계정 링크를 다크웹에서 추출하였으며 추출된 링크를 통해 다크웹에서 소셜미디어로 사용자를 유인하는 마약사범을 추적할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 다크웹 크롤링 통해 범죄를 추적하는 관련 연구를 살펴보고, 3장에서 마약 유통을 위해 사용자를 소셜미디어로 유인하는 정보 수집체계를 설계한다. 4장에서는 설계된 수집체계를 기반으로 실제 크롤러를 개발하여 소셜미디어 유인 링크를 추출함으로써 제안된 수집체계의 성능을 검증한다. 마지막으로 5장에서는 결론 및 향후 연구계획을 제시한다.

II. 관련 연구

본 장에서는 다크웹 및 소셜미디어에서 발생하는 범죄 수사를 위해 수행된 기존 연구를 살펴본다.

다크웹 범죄 추적을 위한 연구로는 불법 다크웹 운영자를 추적하기 위하여 다크웹에서 표면 웹 관련 정보를 추출하고, 표면 웹의 도메인 정보를 활용하여 운영자 정보를 수집하는 연구[1], PGP(Pretty Good Privacy) 공개키 연관분석을 통해서 다크웹 범죄자를 추적하는 연구가 수행되었다[2].

소셜미디어상의 마약 범죄를 분석하는 연구로는 젊은 연령층이 주로 사용하는 소셜미디어 플랫폼에서 불법 마약 거래 현황 및 마약 유통 과정을 조사하는 연구가 진행되었다[3].

기존에 수행된 연구들은 다크웹 및 소셜미디어 중, 특정 플랫폼에 국한하여 분석한다는 점에서 두 플랫폼의 연계성을 추적하는 본 연구와는 차이점이 있다.

III. 다크웹 크롤링을 통한 소셜미디어 유인 링크 수집체계 설계

본 장에서는 다크웹 사용자를 소셜미디어로 유인하는 마약 범죄를 추적하기 위해 [그림 1]과 같이 총 4단계로 구성된 소셜미디어 유인 링크 수집체계를 설계하였다.

1단계에서는 미국 마약단속국에서 제공하는 33개의 마약 용어 중, 가장 대표적인 마약 Cocaine, Heroin, Marijuana를 검색 키워드로 선정하고, 2단계에서는 1단계에서 선정한 마약 용어 3개를 대표적인 다크웹 검색엔진 'Torch'에 입력하여 마약 관련 다크웹 URL을 수집한다. 3단계에서는 수집된 다크웹 URL에 접속하여 각 웹 페이지를 크롤링하고, 4단계에서는 크롤링 데이터에서 트위터 및 텔레그램 링크를 추출한다.

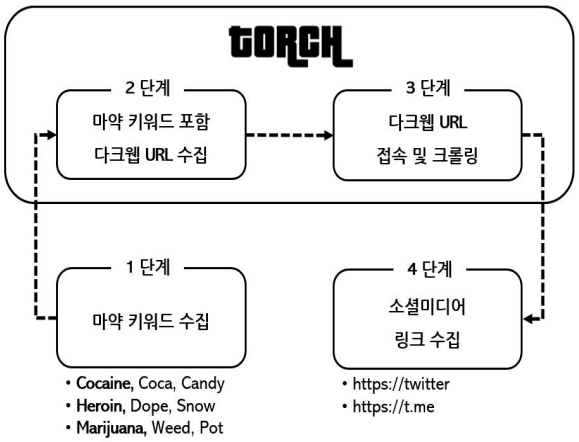

Fig. 1. 마약 범죄 추적을 위한 다크웹 상의 소셜미디어 유인 링크 수집체계

IV. 다크웹 크롤링을 통한 소셜미디어 유인 링크 수집 결과

본 장에서는 3장에서 설계한 다크웹 상의 소셜미디어 유인 링크 수집체계를 기반으로 마약 다크웹 크롤링을 수행하고, 크롤링 데이터에서 트위터 및 텔레그램 유인 링크를 추출한다. Cocaine, Heroin, Marijuana의 표준어 및 은어를 'Torch' 다크웹 검색엔진의 검색 키워드로 입력한 결과, 폐쇄 또는 크롤링 우회 기술이 탑재되어 접속이 불가능한 사이트를 제외하고, 54개의 다크웹을 수집하였다. 수집된 다크웹의 하위 웹 페이지를 포함하여 총 9,046개 웹 페이지를 크롤링한 결과, 트위터 유인 링크 567개, 텔레그램 유인 링크 118개를 추출하였다. 마약 용어별 추출된 트위터 및 텔레그램 유인 링크는 [표 1]과 같다.

Table 1. 마약 용어별 수집된 소셜미디어 유인 링크

마약 용어	검색 키워드	유인 링크	
		트위터	텔레그램
Cocaine	Cocaine(표준어)	0	0
	Coca(은어)	78	79
	Candy(은어)	296	31
Heroin	Heroin(표준어)	0	0
	Dope(은어)	144	5
	Snow(은어)	1	0
Marijuana	Marijuana(표준어)	4	0
	Weed(은어)	0	2
	Pot(은어)	44	1

Table 1에서 Cocaine, Heroin, Marijuana 중, Cocain과 관련된 다크웹에서 가장 많은 소셜미디어 유인 링크가 추출되었고, 표준어보다는 Coca, Candy와 같은 은어 사용 시, 유인 링크가 더 높은 빈도로 추출된 것을 확인할 수 있다. Heroin 및 Marijuana의 경우도 표준어보다 Dope, Pot과 같이 은어로 검색을 수행한 결과 더 많은 유인 링크가 추출된 것을 확인할 수 있다.

V. 결론

본 논문에서는 다크웹과 소셜미디어에 걸쳐 발생하는 마약 범죄를 추적하기 위한 사이버 수사 기술을 개발하기 위하여 다크웹에서 대표적인 소셜미디어인 트위터 및 텔레그램 유인 링크를 수집하는 다크웹 수사체계를 개발하였다. 먼저 미국 마약단속국에서 공개한 대표적인 마약 용어 Cocaine, Heroin, Marijuana를 다크웹 검색엔진 'Torch'의 검색 키워드로 입력하여 마약 관련 다크웹을 검색한 결과 총 9,046개 웹 페이지에서 트위터 유인 링크 567개, 텔레그램 유인 링크 118개가 추출되었다.

3개의 키워드 중, Cocaine 관련 다크웹에서 가장 많은 유인 링크가 추출되었고, Marijuana 관련 다크웹에서 가장 적은 유인 링크가 추출되었다. 특히, 마약 용어를 표준어로 검색한 다크웹보다 은어로 검색한 다크웹에서 소셜미디어 유인 링크가 상대적으로 높은 빈도로

추출된 것으로 보아, 마약 범죄 수사를 위해서는 다양한 은어를 사전에 확보하는 것이 필요함을 알 수 있다.

본 논문에서 제시한 다크웹 상의 소셜미디어 유인 링크 수집체계는 다크웹 사용자를 소셜미디어로 유인하는 마약 범죄를 추적하는 데에 활용할 수 있으며 향후에는 본 논문에서 수집한 유인 링크의 정확성 검증을 위한 소셜미디어 크롤링 연구를 수행할 계획이다. 또한, 마약뿐 아니라, 불법 촬영물, 불법 무기 거래, 해킹 등 다양한 사이버 범죄 추적에 본 연구를 활용할 수 있도록 제안된 수집체계를 확장하고, 다양한 사례연구를 통해 신뢰성을 강화할 계획이다.

ACKNOWLEDGEMENT

이 논문은 과학기술정보통신부·경찰청이 공동지원한 '폴리스랩2.0 사업(www.kipot.or.kr)'의 지원을 받아 수행된 연구 결과입니다. [과제명: 다크웹 범죄 예방을 위한 능동형 다크웹 정보 수집 및 분석·추적 기술 개발 / 과제번호 : RS-2023-00244362]

REFERENCES

[1] Jin Pil-geun. "Study on dark web data collection and analysis methods for forensic investigation." Domestic Master's Thesis Korea University Graduate School of Information Security, 2021. Seoul

[2] Jaejin Kim (2019). Dark Web Crime Investigation Using PGP Public Keys Cues. Journal of Digital Forensics , 13(4), 219-230.

[3] I. Hylén, "The illegal drug trade on social media : Does social media increase the use of drugs among young people?," Dissertation, 2023.