

다크웹 아동 음란물 추적을 위한 연관규칙 기반 키워드 수집체계

김진경^o, 김지연*, 김창훈*

^o대구대학교 컴퓨터정보공학부(컴퓨터공학전공),

*대구대학교 컴퓨터정보공학부

e-mail: wlsrud1470@daegu.ac.kr^o, {jyk, kimch}@daegu.ac.kr*

Keyword Collection System based on Association Rules to Track Pornography of Children on Dark Webs

Jin-Gyeong Kim^o, Jiyeon Kim*, Chang-Hoon Kim*

^oDept. of Computer Engineering, Daegu University,

*Division of Computer and Information Engineering, Daegu University

● 요약 ●

다크웹을 통한 마약, 금융거래, 해킹 등 사이버 범죄가 증가하면서 다크웹 상의 범죄 추적을 위한 사이버 수사 필요성이 증대되고 있다. Tor와 같은 다크웹 접속 브라우저는 강력한 익명성을 제공하기 때문에 주로 다크웹 운영상의 취약점 분석, 악성코드를 활용한 합정수사 기법이 실효성 높은 다크웹 수사 기술로 간주된다. 그러나 사이트 개설 및 폐쇄가 빈번하게 발생하는 다크웹의 특성상 최신 범죄 정보를 수집하기 위해서는 방대한 다크웹 정보를 실시간 수집하고, 능동적으로 검색 키워드를 확장할 수 있는 고도화된 크롤러 기술 개발이 필요하다. 본 논문은 다양한 다크웹 사이트 중, 아동 음란물 사이트를 크롤링을 통해 수집하고, 수집된 텍스트의 연관 분석을 통해 검색 키워드를 확장하는 수집 체계를 제안한다.

키워드: 다크웹(Dark Web), 아동 음란물 (Child Pornography), 연관규칙(Association Rule)

I. 서론

다크웹(Dark Web)은 강력한 익명성을 제공하는 웹 브라우저를 통해 접속할 수 있으며 3개의 노드를 거쳐 IP 추적이 불가능하도록 설계된 Tor(The Onion Router) 브라우저가 가장 대표적으로 사용된다. 다크웹에서는 마약 유통, 아동 음란물 유포, 해킹 등 다양한 범죄가 발생되고 있지만, 강력한 익명성으로 인해 범죄 정보 수집 및 추적이 어려움이 있다.

본 논문에서는 다크웹의 다양한 범죄 중, 아동 음란물을 유포하는 사이트를 식별하고, 범죄 정황을 수집하기 위한 검색 키워드를 능동적으로 확보할 수 있는 키워드 수집체계를 제시한다.

아동 음란물 사이트를 수집하기 위해서는 ‘서비스 이용자’들이 사용하는 다양한 은어를 사전에 확보하고, 이를 다크웹 검색 키워드로 사용하여야 한다. 본 논문에서는 아동 음란물과 관련된 은어를 사용하여 대표적인 다크웹 검색엔진인 ‘Torch’의 Seed 검색 키워드로 사용하여 다크웹 사이트를 수집하고, 수집된 다크웹 텍스트의 연관 분석을 통해 다크웹 사이트 수집에 효과적인 추가 키워드를 확보하는 수집 체계를 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 다크웹 및 아동 음란물 수사 기술 동향을 살펴보고, 3장에서 다크웹 아동 음란물 사이트 수집체계를 설계한 후, 4장에서 결론 및 향후 연구계획을 제시한다.

II. 관련 연구

본 장에서는 다크웹 수사를 위해 사용되는 추적 기술 및 연구 동향에 대해서 살펴보고 아동 음란물 프로파일링 시스템을 이용한 차단 및 수사 방안에 대해 살펴본다.

한국어 다크웹 수사 연구로는 마약 다크웹에서 추출한 PGP(Pretty Good Privacy) 공개키와 아이디를 분석하여 다크웹 범죄자 단서를 확보하는 연구가 진행되었다[1].

아동 음란물 추적을 위한 사이버 수사는 다양한 은어를 사용하여 웹 게시물을 검색하는 연구가 주로 수행되었고, 경찰 프로파일링 시스템에 보관되어 있는 아동 음란물의 MD5(Message-Digest algorithm 5), SHA-1(Secure Hash Algorithm 1) 등 해시값을 이용하

여 파일의 해시값을 기준으로 전송을 차단하고, 이용자에게 경고 메시지를 보내는 수사 기술이 연구되었다[2].

III. 다크웹 키워드 정의 및 수집 체계

다크웹 아동 음란물 관련 Seed 키워드로는 ‘Lolita’와 ‘Pedomom’를 사용한다. ‘Lolita’는 어린 소녀에게 비정상적인 성욕을 느끼는 소아성애자를 의미하며, ‘Pedomom’은 자녀에 대한 사랑이 아닌 자녀를 성욕의 대상으로 생각하는 소아성애자를 의미한다.

본 논문에서는 상기 Seed 키워드를 사용하여 아동 음란물 다크웹을 수집하는 수집체계를 [그림 1]과 같이 제안하였다.

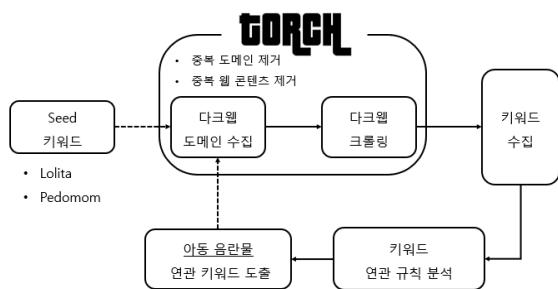


Fig. 1. 연관규칙 기반 다크웹 크롤링 키워드 수집체계

본 논문에서 제안하는 아동 음란물 다크웹 수집체계는 대표적인 다크웹 검색엔진인 ‘Torch’를 사용하여 아동 음란물 다크웹을 검색하고, 중복 도메인 제거, 중복 웹 콘텐츠 제거를 통해 연관규칙 분석을 위한 최종 키워드를 수집한다. 이후, 수집된 키워드를 연관규칙 기반으로 분석하여 아동 음란물 다크웹 검색을 위한 추가 검색 키워드를 도출할 수 있다.

IV. 결론 및 향후 연구

본 논문에서는 아동 음란물 다크웹 크롤링을 위한 키워드 수집체계를 제안하였다. 다크웹 크롤링은 범죄에 사용되는 언어를 Seed 키워드로 사용하였고, 수집된 다크웹 텍스트의 연관규칙 분석을 통해 아동 음란물 다크웹 수집을 위한 추가 검색 키워드를 추가 확보하였다. 본 논문의 결과는 마약 유통, 금융 범죄, 해킹 등 다양한 다크웹 상의 범죄를 추적하는 데에 활용될 수 있으며 향후에는 본 논문에서 추가 확보한 키워드를 활용하여 능동적으로 다크웹 크롤링을 수행할 수 있는 고도화된 사이버 수사 기술을 개발할 계획이다.

ACKNOWLEDGEMENT

이 논문은 과학기술정보통신부·경찰청이 공동 지원한 ‘폴리스랩 2.0사업(www.kipot.or.kr)’의 지원을 받아 수행된 연구결과입니다. [과제명 : 다크웹 범죄 예방을 위한 능동형 다크웹 정보 수집 및 분석·추적 기술 개발 / 과제번호 : RS-2023-00244362]

REFERENCES

- [1] Jaecin Kim.(2019).Dark Web Crime Investigation Using PGP Public Keys Cues.Journal of Digital Forensics ,13(4),219-230.
- [2] Sang Duk Lee.(2018).Blocking and investigation of child pornography using child pornography profiling system of National Police Agency.Journal of Digital Forensics ,12(3),95-105.
- [3] Park, Woong Shin.(2019).Review of new investigative techniques techniques child pornography - Focusing on the investigation of child pomography on the dark web -.Law Review,19(3),51-83.