

그래프 기반의 사이버 위협 분석을 위한 IOC 추출 검증

이주영¹, 한태현¹, 정혜란¹, 이태진²

¹호서대학교 컴퓨터공학부 학부생

²호서대학교 컴퓨터공학부 교수

jylee0741@gmail.com, 12lilliput@gmail.com, 0110jhr@gmail.com, kingecs0@gmail.com

Validation of IOC Extracts for Graph-based Cyber Threat Analysis

Ju-Young Lee¹, Tae-Hyun Han¹, Hye-Ran Jung¹, Tae-Jin Lee²

¹Dept. of Computer Science, Hoseo University, Student

²Dept. of Computer Science, Hoseo University, Professor

요 약

최근 그래프 기반 분석에 대한 연구가 활발히 진행되면서 이를 정보 보안 분야에 적용하려는 시도가 이루어지고 있다. 특히 GNN(Graph Neural Network)은 복잡한 네트워크 데이터를 모델링하고 관계를 분석하는 데 효과적이며, 악성 코드 탐지 등 사이버 공격에 대한 대응 능력을 향상시키는 데 활용할 수 있다. 하지만 GNN을 사용하기 위해서는 그래프의 노드가 될 IOC(Indicator of Compromise) 데이터가 필요하다. 본 논문에서는 IOC Extractor 중 하나인 Cyobstract를 통하여 위협 보고서로부터 IOC를 추출하는 방법과 이를 활용하여 그래프를 구축하고 분석할 방향을 제시한다.

1. 서론

최근 그래프 기반 분석에 대한 연구가 활발히 진행되면서 이를 사이버 보안 분야에 응용하려는 시도가 이루어지고 있다. 특히 GNN(Graph Neural Network)은 복잡한 네트워크 데이터 모델링과 관계 분석에 효과적이며, 악성 코드 탐지나 보안 이벤트 시각화 등 사이버 공격에 대한 대응 능력을 향상시키기 위해 다양한 방법으로 활용할 수 있다. 하지만 GNN을 사용하기 위해서는 그래프의 노드가 될 핵심 데이터가 필요하다. 여기에서 사이버 침해 지표인 IOC(Indicator of Compromise)가 중요한 역할을 한다. IOC는 악성 코드 감염 또는 기타 위협 활동을 감지하는 데 도움이 되는 정보로, 사이버 공격의 탐지, 분석, 대응을 위해 필수적으로 사용된다. 따라서 본 논문은 IOC Extractor 중 하나인 Cyobstract를 통해 사이버 위협 보고서로부터 IOC를 추출하는 방법을 소개하고, 이를 Graph로 구성하여 GNN에 적용 및 분석하는 방법을 제시하려 한다.

2. 관련 연구

근 십 년간 과학 기술의 발전으로 인해 사이버 공격이 급증하면서 분석해야 할 데이터의 양도 함께 증가했다. 따라서 많은 양의 정보로부터 신속하고 효율적으로 IOC를 추출할 필요성이 강조되고 있다.

정규 표현식은 IOC 추출의 효과적인 방법 중 하나로, 특정 패턴을 가진 문자열의 집합을 표현하는 데 사용하는 형식 언어이다. 예를 들어, 3개의 문자열 “Handel”, “Händel”, “Haendel”을 포함하는 집합은 패턴 “H(ä|ae?)ndel”으로 지정할 수 있다. 정규 표현식은 다양한 텍스트 처리 작업에 유용하며 원하는 패턴에 따라 직접 문자를 조합해 사용할 수 있어 특정한 패턴을 보여주는 IOC 추출에 효과적이다.

본 연구에서는 정규표현식을 사용하는 탐지 모듈 Cyobstract를 활용하여 IOC를 추출했다. Cyobstract는 SEI(Software Engineering Institute)의 CERT 부서가 국토안보부(DHS) 사고 보고서의 Dataset에 대해 수행한 탐색적 조사를 지원하기 위해 구축된 사고 대응(IR) 도구로서, 텍스트를 입력하면 IR에 대한 관련 정보를 출력한다. Cyobstract는 정밀도 92%, 재현율 96%, F1 score 94%로 높은 성능을 보여주고 있으며, 탐지 가능한 IOC는 ipv4addr, ipv6addr, ipv4range, ipv6range, ipv4cidr, ipv6cidr, asn, fqdn, email, filename, url, md5, sha1, sha256, ssdeep, filepath, regkey, useragent, cve, cc, isp, asnown, incident, malware, topic로 총 25종이다[1].

3. 제안 Framework

제안하는 연구의 전체 구조는 다음과 같다. 정규

