그래프 기반의 사이버 위협 분석을 위한 IOC 추출 검증

이주영[1], 한태현[1], 정혜란[1] ,이태진[2]
[1]호서대학교 컴퓨터공학부 학부생
[2]호서대학교 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 Lee[1], Tae-Hyun Han[1], Hye-Ran Jung[1], Tae-Jin Lee[2]
[1]Dept. of Computer Science, Hoseo University, Student
[2]Dept. 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 Cyobstract을 통해 사이버 위협 보고서로부터 IOC를 추출하는 방법을 소개하고, 이를 Graph로 구성하여 GNN에 적용 및 분석하는 방법을 제시하려 한다.

2. 관련 연구

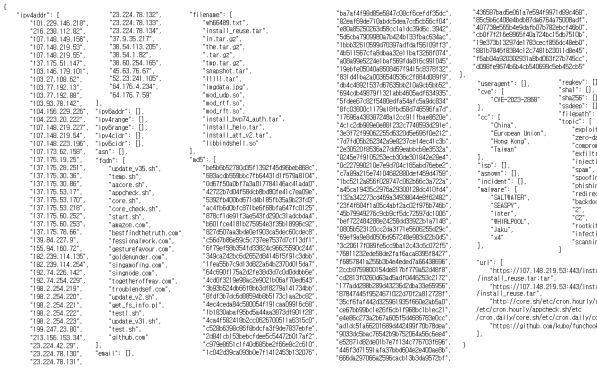
근 십 년간 과학 기술의 발전으로 인해 사이버 공격이 급증하면서 분석해야 할 데이터의 양도 함께 증가했다. 따라서 많은 양의 정보로부터 신속하고 효율적으로 IOC를 추출할 필요성이 강조되고 있다.

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



(그림 1) IOC 추출 결과

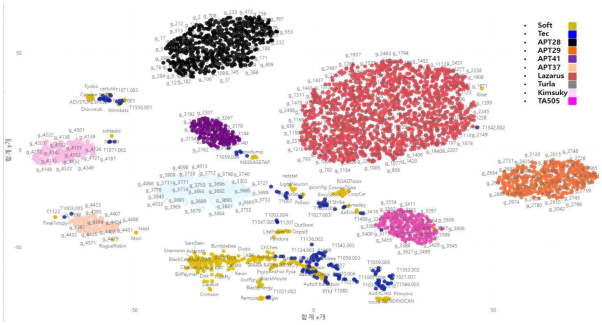
표현식을 기반으로 한 IOC Extactor를 사용하여 각종 위협 보고서로부터 IOC를 추출하고, 추출된 IOC 정보를 특정 포맷으로 정형화한 뒤 Graph로 표현한다. 이렇게 만들어진 IOC Graph를 GNN에 학습시키면 효율적인 공격 패턴 감지와 위협 분석이 가능해져 보안 담당자에게 위협 환경에 대한 넓은 인사이트를 제공할 수 있으며, 효과적인 사이버 보안 대응 전략을 구축할 수 있다. 본 논문에서는 그래프를 표현하기 위한 IOC를 추출하고, 이를 정형화하는 단계까지 진행하였으며 그래프 연계 시 활용 방안은 결론 및 향후 연구 방향에서 설명할 것이다.

4. 실험 및 결과

인터넷 데이터상에서의 IOC 추출을 위해 Mandiant 위협 보고서를 사용하였다[2]. 해당 보고서는 중국과 연계된 것으로 보이는 공격 그룹 UNC4841의 전술 및 기술에 관해 설명하고 있다. 보고서에는 IP 주소, FQDN, 해시 함수, 취약점, URL, 파일 이름 등 다양한 IOC 정보가 포함되어 있다. 수집한 보고서는 html 태그를 제거하고, 본문과 표에서 텍스트만 추출한다. 그다음, Cyobstract를 활용하여 IOC를 추출했다. 모듈이 탐지하지 못하는 IOC는 직접 정규표현식을 만든 후 추출을 진행했다. 추출된 결과는 (그림 1)과 같다. ipv4addr, fqdn, filename, url, md5, cve, cc, malware, topic이 모두 추출된 모습을 볼 수 있다.

5. 결론 및 향후 연구 방향

본 연구는 사이버 위협 보고서나 텍스트 데이터에 정규 표현식을 기반으로 한 IOC Extractor를 활용하여 IOC 핵심 정보를 추출하였다. 추출된 IOC는 클러스터링하여 (그림 2)와 같이 IOC간의 유사도를 확인할 수 있다. 이를 통해 특정 공격과 관련된 다른 공격에 대한 정보를 제공할 수 있고, 이전에는 알려



(그림 2) IOC를 통한 공격 그룹 클러스터링 예시

지지 않았던 위협 간의 연관성을 보여줄 수 있다.

또한, 이를 추후 GNN을 활용하여 그래프 형태로 시각화할 수 있다. 각 IOC를 Node로 나타내고, Node 간의 연결 관계를 Edge로 표현한다. 이를 통해 그래프 데이터 사이의 연관성을 고려할 수 있으며, 공격자의 행동 패턴과 전략을 더욱 자세하게 분석할 수 있어 신속하고 효율적으로 위협 대응 능력을 향상시킬 수 있다.

다만 특정 패턴이 없고 고유 명사로 이루어진 Malware와 공격자 등을 탐지하기 위해서는 추가적인 데이터가 필요하다. 따라서 향후 연구에서는 추가 조사를 통해 추출할 수 있는 IOC를 증가시키고, 추출한 IOC를 바탕으로 IOC Graph를 만들어 GNN에 학습시킨 후, 학습된 모델을 바탕으로 사이버 공격이나 위협을 예측하는 방법까지 연구하고자 한다.

Acknowledgements

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (RS-2023-00235509, ICT융합 공공 서비스·인프라의 암호화 사이버위협에 대한 네트워크 행위 기반 보안관제 기술 개발)

참고문헌

[1] Caballero, Juan, et al. "The rise of GoodFATR: a novel accuracy comparison methodology for indicator extraction tools." Future Generation Computer Systems 144 (2023): 74-89.

[2] Austin Larsen, John Palmisano, Mathew Potaczek, John Wolfram, Matthew Mcwhirt, “Barracuda ESG Zero-Day Vulnerability (CVE-2023-2868) Exploited Globally by Aggressive and Skilled Actor, Suspected Links to China”, Mandiant, https://www.mandiant.com/resources/blog/barracuda-esg-exploited-globally, (2023)