사이버 공격 캠페인 유사도 분석을 위한 TTP 시퀀스 임베딩
Analyzing Cyberattack Campaign Similarity via TTP Sequence Embedding

이인섭* · 신찬호* · 신성욱* · 서성연* · 최창희*
Insup Lee* · Chanho Shin* · Sunguk Shin* · Seongyun Seo* · Changhee Choi*

* 국방과학연구소
(dlstjq0711@add.re.kr)

ABSTRACT

As cyberattacks became more sophisticated and intelligent, APT attacker groups emerged and expanded on campaign-scale attacks. To analyze these cyberattack campaigns, numerous studies have been made to model the campaigns as TTP sequences using cyber kill chains. Recently, deep learning has been widely exploited for TTP sequence detection, which makes embedding sequences into appropriate vectors to calculate similarity an important issue. In this paper, we propose a TTP sequence embedding method using TF-IDF and calculate cosine similarity amongst the vectors.

Key Words : Cyber attack campaign, embedding, TF-IDF, TTP sequence, ATT&CK,

1. 서론

정보통신이 급속도로 발전함에 따라 많은 편의를 가져다주었지만, 공격 표면이 확대됨에 따라 사이버 공격 및 보안 위협 역시 증대되었다. 이러한 사이버 공격이 점차 고도화되어 Advanced Persistent Threats (APT) 공격을 수행하는 그룹들이 전 세계적으로 등장하였고 이에 대응하는 것은 국가차원에서도 중요한 문제이다. APT 공격을 효과적으로 분석하기 위해, 미국 MITRE에서는 사이버 킬 체인으로 공격 캠페인을 모델링한 매트릭스인 ATT&CK [1]를 발표한 바 있다.

최근 딥 러닝이 많은 주목을 받게 되면서 도메인에 맞게 데이터를 임베딩 하는 연구와 [2, 3] 딥 러닝을 활용해 공격을 탐지하는 연구가 많이 수행되고 있다. 하지만 아직 캠페인 규모의 공격에 대한 탐지에 대한 연구는 초기 단계이며, TTP 시퀀스로 표현된 캠페인 데이터에 대해 효과적인 임베딩 방법을 찾는 것은 중요한 문제이다. 본 논문에서는 TF-IDF를 활용해 사이버 공격 캠페인 데이터를 임베딩하고, 캠페인 간 유사도를 계산하는 방안에 대해 제안한다.

2. 사이버 공격 캠페인 데이터

2.1 MITRE ATT&CK

MITRE에서 만든 ATT&CK 프레임워크는 사이버 킬 체인의 단계를 자체적으로 정리한 프레임워크이다. 공격 사례들을 바탕으로 공격자가 이용한 악성 행위에 대해 공격 방법 (Tactics), 기술 (Techniques), 절차 (Procedures) 등 TTP를 구조화 하는 것에서 시작되었으며, 일관된 공격 패턴에 대한 분석을 기반 하여 TTPs 정보를 매핑 해 공격자의 행위를 식별한다.

2.2 Dataset preprocessing

rcATT (reports classification by Adversarial Tactics and Techniques) [4]는 TTP 태깅 알고리즘으로, 입력 받은 텍스트 형식의 cyber threat reports에 대해 tactics 와 techniques를 예측한다. rcATT는 태깅 결과를 JSON 혹은 STIX 형식으로 저장하므로 딥러닝 모델의 학습데이터로 쉽게 사용 가능하다.

본 연구에서는 rcATT로 태깅 한 1490개의 사이버 캠페인 공격 분석 리포트를 기본 데이터 셋으로 사용한다, 이 중 시퀀스 길이가 최소 3 이상인 462개 캠페인에 대해서만 실험을 수행한다. TTP 시퀀스로 표기된 데이터는 표 1과 같으며, 특정 캠페인을 나타내는 technique 순서는 무작위로 배치한다.

Table 1. 도메인 변환 결과

Cyberattack Campaign	Sequence
11	T1066 → T1064 → T1027
35	T1066 → T1108 → T1045 → T1110
335	T0002 → T0003 → T0004 → T0005 → T1053 → T1106 → T1117 → T1059 → T1015 → T1034

3. 유사도 분석을 위한 사이버 공격 캠페인 임베딩

3.1 사이버 공격 캠페인 임베딩

TF-IDF는 여러 문서들이 존재하는 문서 집합에서 특정 단어가 갖는 중요도를 표현하는 통계적 수치 기법이다. TF (Term Frequency)는 주어진 문서에서 특정 단어가 발생하는 빈도를 의미하며, IDF (Inverse Document Frequency)는 특정 단어가 발생하는 문서 수의 역을 의미한다. TF-IDF는 TF와 IDF를 곱한 수치이며, 단어가 특정 문서에서 자주 등장하고 다른 문서

에서는 등장 빈도가 낮을수록 높은 값을 가진다.

본 연구에서는 캠페인 데이터, 즉 TTP 시퀀스와 테크닉의 관계를 TF-IDF의 문서와 단어로 치환하여 접근한다. 다시 말해서, 사이버 캠페인 임베딩 시 테크닉이 특정 캠페인에 자주 등장할수록 그리고 적은 캠페인에 등장할수록 중요도가 높다는 전제하에 임베딩을 수행한다. 4개의 테크닉으로 이루어진 캠페인 데이터에 대한 TF-IDF 임베딩 예시는 표 2와 같다.

Table 2. TF-IDF를 활용한 TTP시퀀스 임베딩

Cyberattack Campaign	Embedding Results	
35	TTP Sequence	T1066 → T1108 → T1045 → T1110
	Embedded Vectors	(0.54391, 0.46437, 0.42502, 0.55485)

다음으로, 임베딩 된 TTP 시퀀스를 2차원 벡터공간에 투영하여 시각화하여 분석한다. 높은 차원의 복잡한 데이터를 2차원으로 차원 축소하는 대표적 기법인 T-SNE를 사용하며 시각화된 결과는 그림 1과 같다. 캠페인 기법이 고도화됨에 따라 공격그룹의 패턴이 잘 드러나지 않고 데이터가 복잡하므로, 시각화된 벡터들이 산재되어 명확한 정보가 도출되지는 않음을 확인할 수 있다. 그러나 일부 캠페인벡터들은 응집되어 있으며, 이들에 대한 유사도 분석을 수행한다.

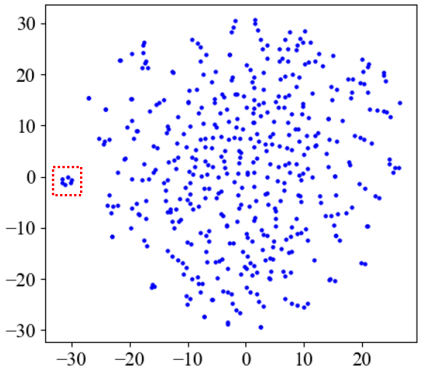

Fig. 1. 임베딩 된 사이버 공격 캠페인에 대한 2차원 T-SNE 시각화

3.2 캠페인 유사도 분석

본 절에서는 임베딩 된 사이버 공격 캠페인 벡터 간 유사도를 구하고 시각화하는 과정을 기술한다. 가장 단순한 유사도인 코사인 유사도를 활용하며 범위를 조정해주기 위해 지수함수를 취하여 다음과 같이 구한다.

$$similarity = \exp(\cos(\theta)) = \exp\left(\frac{A \cdot B}{\|A\|\|B\|}\right)$$

각 캠페인 데이터 간 유사도를 시각화하기 위해 모든 캠페인 간 코사인 유사도를 계산하여 462x462 매트릭스 형태의 유사도 히트 맵을 작성한다. 두 캠페인 간 유사도가 높을수록 히트 맵 상에서 파란색으로 표기되며, 유사도가 낮을수록 흰색으로 표기된다. 시각화 된 히트 맵은 그림 2와 같다. 462개의 캠페인 중 가장 유사도가 높게 측정된 캠페인 쌍 중 하나는 (Campaign332, Campaign335) 이었으며, 두 캠페인 벡터 간 유사도는 2.4586으로 측정되었다.

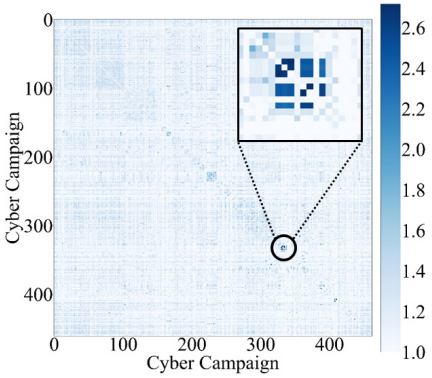

Fig. 2. Cosine similarity 기반 캠페인 데이터 유사도 matrix

유사도 분석 결과를 검증하기 위해 추가적으로 데이터에 대한 수동 분석을 수행한다. 데이터 전처리 전 원본 캠페인 분석 리포트의 내용 중 일부는 표 3과 같다. 쉘 실행이라는 공통된 내용에 대해 다른 방식으로 서술하고 있으나 TTP tagging에 해당 내용이 반영되므로, 캠페인 데이터를 TTP 시퀀스 임베딩 한 벡터 중 유사도가 높은 대상은 실제 원본 데이터의 내용도 유사함을 확인할 수 있었다.

Table 3. 수동분석 결과

Cyber Campaign	Descriptions
332	(생략) ... Malware for remote control (Remote Access Tool/Trojan RAT) has a function to **execute shell commands** from a remote environment ... (생략)
335	(생략) ... After CreateProcess, the next most commonly used API where software restriction policy is envorced is **ShellExecute** ... (생략)

4. 결론

본 연구에서는 딥 러닝 기반 사이버 공격 캠페인 탐지를 위해, 캠페인 데이터를 TF-IDF로 임베딩 한 후 코사인 유사도 기반 캠페인 유사도를 계산했다. 임베딩 된 벡터와 유사도 결과를 시각화한 후, 결과를 추가 검정하기 위해 유사도가 높게 측정된 캠페인에 대한 원본 데이터 정보를 활용했다.

References

[1] MITRE ATT&CK, https://attack.mitre.org

[2] M. ring, A. Dallmann, D. Landes, and A. Hotho, “Ip2vec: Learning similarities between ip addresses,” in Proc. of IEEE International Conference on Data Mining Workshops, 2017.

[3] E. L. Goodman, C. Zimmerman, and C. Hudson, “Packet2vec: Utilizing word2vec for feature extraction in packet data,” arXiv preprint arXiv:2004.14477, 2020.

[4] rcATT, https://github.com/vlegoy/rcATT