CIA 라벨 기반 TTP 분류 기술
Classifying TTP based on CIA labeling

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

* 국방과학연구소
(shinch2018@add.re.kr)

ABSTRACT

Cyber attacks on the national level are occurring in addition to attacks on individuals and groups. Not only the subjects of attack, but also the techniques and goals used have diversified, which make defenders inefficient to respond in terms of techniques. To deal with the inefficiency, the paper relabeled the TTP of the ATT&CK matrix based on the CIA triad of confidentiality, integrity, and availability as well as conducting a classification experiment to verify the labeling method.

Key Words : TTP, classification, BERT, MITRE ATT&CK

1. 서론

사이버 공격이 개인과 집단을 넘어 국가 단위로까지 수행됨에 따라 공격의 목적과 목표도 다양해지고 있다. MITRE ATT&CK[1]에서는 TA0040(Impact) 전략을 통해 공격자의 목표를 분류하였지만, 방어자 입장에서 TA0040의 각 기술을 일일이 대응하는 것은 비효율적이다. 본 논문은 MITRE ATT&CK를 정보보안의 3요소(기밀성, 무결성, 가용성) 관점에서 재라벨링(relabeling)하였다. 또한 CyBERT[2] 모델을 이용하여 분류실험을 진행함으로써, 이러한 라벨링 방법론이 타당함을 보인다.

2. 배경지식

2.1 MITRE ATT&CK

각 공격에 대한 수많은 보고서가 나옴에 따라 방어자들은 분석 보고서를 공유할 기준의 필요성을 느꼈다. 이에 MITRE에서는 사이버 공격을 TTP 단위로 분석할 수 있는 ATT&CK를 제시하였다. 현재 버전 11.0까지 나왔으며, 총 14개의 tactic, 191개의 technique, 386개의 sub-technique으로 이루어져 있다.

2.2 정보보안의 3요소 CIA

정보보안의 3요소에는 기밀성(Confidentiality), 무결성(Integrity), 가용성(Availability)이 있다. 기밀성은 권한이 있는 사람만 정보를 확인할 수 있어야 함을 의미하고 무결성은 권한이 있는 사람만 정보를 수정할 수 있음을 의미한다. 가용성은 권한이 있는 사람이 정보를 요구할 때, 언제나 열람할 수 있어야 함을 의미한다.

2.3 BERT, CyBERT

BERT[3]는 2018년 구글에서 공개한 사전훈련된 언어모델이다. Transformer[4]를 기반으로 만들어졌으며, 학습데이터로 영문 위키피디아 등을 사용하였다. 구글은 해당 모델을 Language modeling과 next sentence prediction 과제를 중심으로 학습시켰다.

CyBERT는 BERT 모델을 블로그 글, NVD[5], CVE[6] 등에 대해 재학습한 모델로, 사이버 보안 분야에 특화된 언어 모델이다.

3. 데이터 셋 및 모델

3.1 데이터 셋

MITRE ATT&CK의 설명문을 기반으로 각 기술을 정보보안의 3요소 관점에서 재라벨링하고자 한다. 따라서 데이터 셋은 (설명문, 라벨)로 이루어진다. 데이터 셋의 예시는 다음과 같다. 본 논문은 MITRE ATT&CK v10.0을 기준으로 데이터 셋을 구성하였다.

Table 1. 데이터셋 예시

Technique	설명문	라벨
T1531	Adversaries may interrupt availability of system	가용성
T1496	Adversaries may leverage the resources of co-opted	가용성
T1561	Adversaries may wipe or corrupt raw disk data on	무결성
T1565	Adversaries may insert, delete, or manipulate data in	무결성
T1052	Adversaries may attempt to exfiltrate data via a	기밀성
T1020	Adversaries may exfiltrate data, such as sensitive	기밀성

데이터 셋의 설명문을 CyBERT에 학습시키기 전에 일부 전처리 과정을 거친다. 설명문 내에 삽입된 코드나 인용문의 주소 등은 삭제 처리한다.

3.2 모델 및 파라미터

BERT는 huggingface의 transformers 라이브러리[7]내에 포함된 BertForSequenceClassification 클래스를 사용하였다. 사전학습된 CyBERT 가중치 값은 저자의 github[8]에서 구할 수 있다. 토크나이저도 마찬가지로 transformers 라이브러리의 BertTokenizer를 사용하였다. 그 외에는 모두 기본 파라미터를 사용하였다.

4. 실험 결과

4.1 환경 세팅

토크나이저는 최대 512 토큰을 가지게끔 설정하였다. 모델의 학습은 총 100 epoch 동안 진행하였으며, 배치는 20, optimizer는 AdamW, learning rate는 10^{-5}를 사용하였다. 실험에 사용한 기기는 i9-10980XE, RAM 128GB, GPU NVIDIA GeForce RTX 3090을 사용한다.

4.2 성능

총 43개의 데이터를 8:2 비율로 학습 및 테스트 데이터로 나누었다. 학습데이터와 테스트 데이터의 분포는 다음과 같다.

Table 2. 학습 및 테스트 데이터 수

	기밀성	무결성	가용성
학습	12	11	10
테스트	3	3	2
총	15	14	12

그림 1은 모델의 학습 중 loss 값을 보여준다. 여러 번 실험을 돌려본 결과, 평균적으로 20~50 epoch에서 validation loss가 감소하다가 증가하는 것을 볼 수 있다.

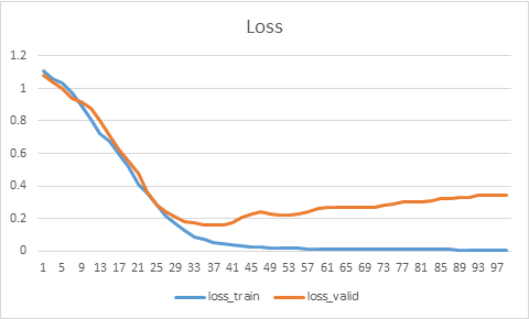

Fig. 1. Loss 그래프

그림 2는 실험결과이다. 정확도(Accuracy) 값과 F1 점수 모두 평균적으로 0.77~0.88 값을 보이며 이는 총 9개의 테스트 데이터 중 7~8개의 데이터를 맞춘다는 뜻이다. 즉, MITRE ATT&CK의 설명문을 토대로 TTP를 정보보안의 3요소를 기반으로 재라벨링하는 것이 합리적임을 증명한다.

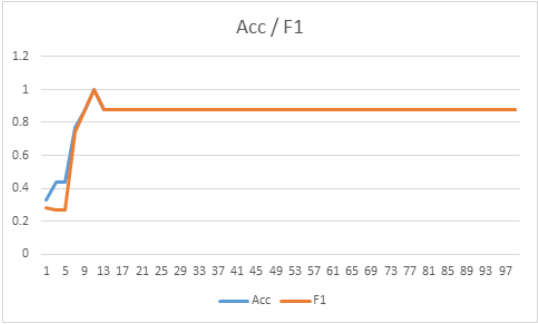

Fig. 2. Accuracy / F1 그래프

5. 결론

본 논문에서는 MITRE의 ATT&CK 중 공격목표와 관련 있는 일부 기술을 정보보안의 3요소를 기준으로 재라벨링하고 CyBERT를 이용한 분류 실험을 통해 합리성을 보였다. 실험 결과 총 3개의 라벨에 대해 약 90%의 정확도를 보이는 것으로 보아, 각 기술의 설명문을 토대로 라벨을 재라벨링하는 것은 충분히 합리적인 방법론이라 할 수 있다. 각 기술은 싱글라벨이지만 실제 보고서 및 공격사건은 여러 기술을 가지기 때문에 멀티라벨 문제가 된다. 앞으로는 공격 목표가 여러 개일 때, 문제를 어떻게 해결할지 등에 대한 연구를 진행할 예정이다.

References

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

[2] Ranade, et al, “CyBERT: Contextualized Embeddings for the Cybertesecurity Domain”, IEEE International Conference on Big Data, 2021.

[3] Jacob Devlin, Ming-Wei chang. Kenton Lee, and Kristina Toutanova, “BERT: Pre-training of deep bidirectional transformers for language undersatnding”, arXiv preprint arXiv:1810.04805, 2018.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need”, In Advances in Neural Information Processing Systems, pp. 6000-6010, 2017.

[5] NVD: Harold Booth, Doug Rike, and Gregory Witte. The national vulnerability database(nvd): Overview, Technical report, National Institute of Standards and Technology, 213.

[6] Mitre corporation. Common vulnerabilities & exploitations. https://cve.mitre.org

[7] https://huggingface.co

[8] https://github.com/priyankaranade1/CyBERT