Siamese LSTM 기반의 사이버 공격 그룹 분류
Cyber Attack Group Classification using Siamese LSTM

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

국방과학연구소
(changhee84@add.re.kr)

ABSTRACT

As cyber warfare intensifies, various computerized military facilities are seriously threatened. In addition, the scale of cyber attack is increasing to the campaign level, the existing cyber defense system is easily neutralized. To counter this, many researches are being conducted to model cyber attacks at the campaign level. In this paper, we propose a method to classify cyber attack group to help cyber defense. We express the cyber attack campaign as a sorted techniques, and we use the Siamese LSTM. Experimental results show that the proposed method can classify attack group with high accuracy.

Key Words : Cyber attack group. ATT&CK, Siamese network, Long short-term memory

1. 서론

사이버 공격이 개인의 수행하는 단순한 공격에서, 단체가 지원하는 복합적인 공격으로 진화함에 따라 전산화된 군사 시설이 크게 위협받고 있다. 특히 단체가 지원하는 APT(Advanced Persistence Threat)의 경우 기존의 사이버 방어 체계를 우회할 수 있도록 설계되는 경우가 많아 문제가 되고 있다. 기존에는 공격에 대응하기 위해 취약점을 찾아 미리 예방하는 방법이나[1], 공격자의 다음 행위를 예측하여 대응하는 방법이 연구됐다[2]. 단체가 지원하는 사이버 공격 그룹은 특정 임무가 주어지는 경우가 대부분이다. 예를 들면 “FIN6” 그룹은 POS기에 저장된 카드 정보를 탈취하는 등 금전적 이익을 얻기 위한 목적을 하고 있으며, “DarkHotel” 그룹은 스파이 활동을 목적으로 하고 있다[3]. 공격 그룹을 예측할 수 있다면, 최종 목표를 추정할 수 있고, 이에 따른 사이버 방어가 가능하다.

본 논문에서는 MITRE에서 공격 행위를 모델링한 ATT&CK을 기반으로 하여 사이버 공격의 공격자를 분류하는 연구를 진행하였다. 기존의 연구는 TTP에 영향성 점수를 매기고, 그룹 유사도 점수를 산출하는 방법을 이용하였는데, 정확도가 낮은 단점이 있었다[4]. 본 논문에서는 이를 보완하기 위해서 적은 데이터 셋에서도 우수한 성능을 보이는 Siamese Network와 순차적 행위 예측에 우수한 성능을 보이는 LSTM을 결합한 방법을 제안하였다.

2. MITRE ATT&CK

MITRE ATT&CK은 사이버 공격의 전술과 기술을 추상적으로 모델링하였다[3]. 전술 14개, 기술 188개, 서브기술 379개를 모델링하였는데, 본 논문에서는 전술과 기술만 사용하였다.

3. 데이터 셋 구축

과거 사이버 공격을 가장 빠르게 파악하는 방법은, 여러 보안 업체가 작성한 사이버 공격 분석 보고서를 읽는 것이다. 2006년부터 산재하여 있는 보고서나 블로그 글들을 모아둔 저장소가 존재한다[5]. 약 1,400여 개의 보고서가 저장되어있으며, 한 달에 10~20여개 정도 추가된다. 본 논문에서는 사이버 공격의 전술과 기술을 사용하기로 하였으므로, 이를 보고서로부터 추출할 방법이 필요하다. Valentine Legoy 연구진은 보고서에서 MITRE ATT&CK 전술 및 기술을 추출하는 연구 rcATT를 진행하였다[6].

우리는 1,400여 개의 보고서에 대해 rcATT알고리즘을 수행하였다. 너무 짧은 보고서의 경우 전술과 기술이 제대로 추출되지 않는 경우가 있고, 짧은 글의 경우 사이버 공격보다는 특정 기술이나 악성코드에 관해 설명한 경우가 많아, 기술의 개수가 5개 미만이고, 각 그룹의 표본이 5개 미만이면 데이터 셋에서 제외하였다. 현실성 있는 실험을 위해 2018년도까지 일어난 사이버 공격을 훈련 셋으로 하였으며 2019~2021년도에 일어난 사이버 공격을 테스트 셋으로 하였다. 표 1은 구축한 데이터의 일부 표본이다.

Table 1. 데이터셋 샘플

보고서	전술, 기술
Operation_Iron_Tiger ...	TA0001 TA0002 TA0006 TA0007 T1197 T1059..
Targeted_Attacks_On ...	TA0001 TA0002 TA0005 TA0011 T1059 T1203..
Operation_Daybreak ...	TA0001 TA0002 TA0005 T1189 T1203 T1559 ...

3. 제안하는 방법

3.1 전처리 및 도메인 변환

전처리 방식는 기존 연구[2]와 유사하게 진행하였다. 전술의 경우 14개이고, 대부분의 사이버 공격이 해당 전술들을 대부분 포함하고 있어 그룹 분류에 도움이 되지 않는다고 판단하였다. 따라서 기술의 순서를 정할 때 간접적으로 이용하고 삭제하였다.

기술의 순서는 각 공격 그룹마다 선호하는 방식이 반영되어 있다고 판단되어, MITRE ATT&CK에서 정의한 전술 순서대로 기술을 정렬하였다. 하나의 기술의 2개 이상의 전술에 포함되는 경우가 있는데, 이때에는 rcATT기술이 검출한 전술이 존재하면 이를 우선시하고, 없으면 앞선 전술을 사용하였다. 표 2는 표 1의 데이터를 처리한 결과이다.

Table 2. 도메인 변환 결과

보고서	정렬된 기술
Operation_Iron_Tiger ...	T1566->T1059->T1204->T1546->T1197->T1027..
Targeted_Attacks_On ...	T1566->T1059->T1203->T1057->T1102
Operation_Daybreak ...	T1189->T1566->T1203->T1559->T1204->T1036

3.2 학습 네트워크 설계

Siamese 네트워크는 적은 양의 데이터로도 학습 가능하도록 2개의 서브 네트워크를 이용하여 차이를 구하는 방식으로 설계되어 있다[7]. 전처리 및 도메인 변환이 끝난 데이터는 순서 있는 시퀀스이므로, 우리는 서브 네트워크를 LSTM을 이용하여 구성한 Siamese network를 그림 1과 같이 설계하였다.

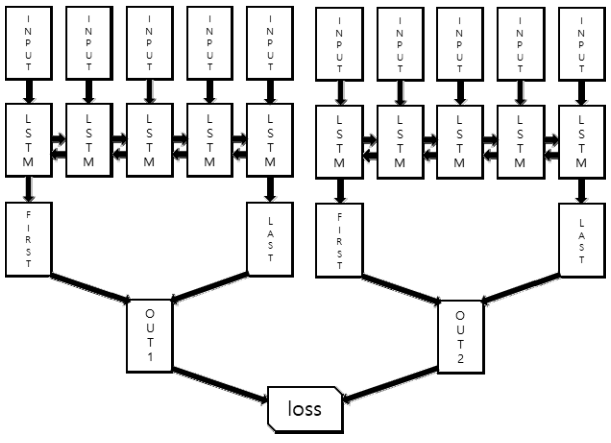

그림 1. Siamese-LSTM

기술의 경우 one-hot 인코딩을 수행하여 입력하였다. 데이터의 양이 적기 때문에, LSTM의 레이어는 2개로 제한하였고, 실제 기술의 흐름과 반대인 경우도 종종 있어서 양방향으로 설계하고, 쌍둥이 서브 네트워크를 똑같이 생성한 후, 두 네트워크에서 나오는 출력의 로스를 구하는 방식으로 학습을 진행하였다.

4. 실험 결과

Siamese 네트워크는 2가지 입력에 대해서 동일 여부를 판별해주는 주므로, 표본의 모든 데카르트 곱에 대해서 테스트하였다. 테스트 표본 개수는 18개였으며, 그림 2는 epoch에 따른 loss, 표 3은 정확도이다.

Table 3. 분류 정확도

	FN	FP	FN	TP	ACC
값	294	0	2	28	99.4%

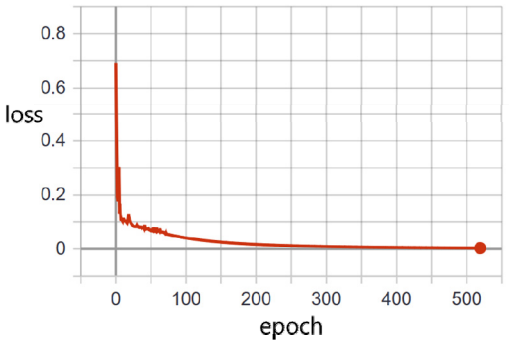

그림 2. epoch 에 따른 loss

5. 결론

본 논문에서는 MITRE ATT&CK을 기반으로 Siamese LSTM을 이용한 기술 사이버 공격 그룹 분류 기술을 제안하였다. 정확도 99.4%의 성능으로 분류 가능함을 실험으로 증명하였다. 향후 불안정한 네트워크를 안정시키는 연구를 진행할 예정이다.

References

[1] 김상수, 심신우, 임선영, 구성모, “사이버 위협 헌팅을 위한 사용자 행위 정보 기반 위협 우선순위 산정 기법 연구”, 한국통신학회논문지, 46(11) 1853-1861.

[2] 최창희, 신찬호, 신성욱, 서성연, 이인섭, “사이버 공격 행위 예측을 위한 딥러닝 학습 방법”, 한국군사과학기술학회 종합학술대회, 2021.

[3] MITRE ATT&CK, https://attack.mitre.org/, (accessed May, 2, 2022)

[4] 최창희, 신찬호, 신성욱, “TTP 정보 기반 사이버 공격 그룹 분류 기술”, 인터넷 정보학회 춘계 학술발표대회, 2021.

[5] APT & Cybercriminals Campaign Collection, https://github.com/CyberMonitor/APT_CyberCriminal_Campaign_Collections

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

[7] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov, “Siamese neural networks for one-shot image recognition.”, ICML deep learning workshop, Vol. 2, 2015.