블랙리스트 식별 정보에 따른 딥러닝 기반 익스플로잇 킷 탐지 성능 비교

김보선, 최정우, 이민성, 백의준, 김명섭

고려대학교

{boseon12, choigoya97, min0764, pb1069, tmskim}@korea.ac.kr

Performance Comparison of DL-based Exploit Kit Detection According to Blacklist-Identification

Boseon Kim, Jeong-Woo Choi, Min-Seong Lee, Ui-jun Baek, Myung-Sup Kim

Korea University

요 약

익스플로잇 킷은 취약점 공격 도구로, 브라우저를 통해 사용자 시스템과 통신하여 취약점을 식별하고 악성 코드를 사용하여 사용자 시스템에 악성 소프트웨어를 전달하고 실행한다. 기존의 블랙리스트 기반 탐지 방법의 경우, EK 개발자가 악의적으로 host나 ip와 같은 블랙리스트 기반 식별 정보를 동적으로 변조하거나 정상 트래픽과 유사하게 위장할 경우, EK 트래픽을 정상 트래픽으로 잘못 판단해 대처하지 못할 가능성이 존재한다. 블랙리스트 기반 식별 정보가 위장 혹은 변조될 가능성과 신종 EK 트래픽을 탐지하기 어렵다는 가능성을 두고 해당 시나리오를 가정하여 분류 모델이 블랙리스트 기반으로 정상과 비정상 트래픽을 분류하는지 실험하였고, 실험 결과 블랙리스트 기반 식별 정보 없이는 오탐률이 증가할 가능성이 있다는 것을 알 수 있었지만 큰 차이를 보이지 않았다. 또한 정상으로 위장하거나 변조된 EK 트래픽을 제대로 탐지할 수 있는지 알고자 하여 EK 트래픽의 식별 정보를 정상 트래픽과 유사하게 변조하여 테스트 하였고, 실험 결과 식별 정보가 변조된 EK 트래픽을 탐지하기 어렵다는 것을 알 수 있었다.

Ⅰ. 서 론

익스플로잇 킷(Exploit Kit)은 취약점 공격 도구로, 브라우저를 통해 사용자 시스템과 통신하여 취약점을 식별하고 악성 코드를 사용하여 사용자 시스템에 악성 소프트웨어를 전달하고 실행한다. 이는 응용 프로그램의 취약점을 공격하는 취약점 코드 및 악성코드 유포 사이트 등에 대한 정보를 가지고 있으며, 해당 도구를 이용하여 자동으로 웹 사이트 취약점을 공격하게 된다. EK는 매우 정교하고 자동화되어 사이버 범죄용으로 암시장에서 거래되고 있으며, 사이버 범죄자들이 이용하기 쉽고 구하기도 간편하여 현재까지도 공격 도구로 사용되고 있다. 대표적인 EK 탐지 방법은 블랙리스트 기반 탐지 방법이다. 그러나 EK 개발자가 host나 ip와 같은 블랙리스트 기반 식별 정보를 동적으로 변조하고 위장할 수 있는 탐지 기술 회피 방법을 보유하고 있어, 블랙리스트를 피해갈 수 있다. EK 개발자가 악의적으로 블랙리스트 기반 식별 정보를 변조하거나 위장할 경우, EK 트래픽을 정상 트래픽으로 잘못 판단해 탐지하지 못할 가능성이 높다.

우리는 블랙리스트 기반 탐지 방법이 변조된 EK 트래픽과 신종 EK 트래픽에 대해 대처하지 못하고 오탐할 수 있다는 시나리오에 기반을 두어 실험을 진행하였다. 두 가지 실험을 통해 실험 모델이 블랙리스트 기반으로 트래픽을 탐지하는지 알고자 하며, 정상 트래픽으로 위장한 EK 트래픽을 올바르게 탐지하는지 알고자 한다. 첫 번째는 Raw 데이터 셋을 사용한 분류 실험과 Raw 데이터에서 블랙리스트 기반 식별 정보를 제거한 데이터 셋을 사용한 분류 실험 결과에 대해 비교한다. 이를 통해 분류 실험 모델이 블랙리스트 기반 식별 정보를 기반으로 학습을 하는지 알 수 있다. 두 번째는 Raw 데이터 셋을 사용한 분류 실험과 Raw 데이터 셋에서 EK 트래픽을 정상 트래픽으로 위장하여 분류 실험하여 비교한다.

본 논문은 2장에서 관련 연구를 제시하고 3장에서 정상 트래픽과 비정상 트래픽(EK) 분류 실험 및 실험 결과를 기술한다. 마지막으로 4장에서는 결론 및 향후 연구를 제시한다.

Ⅱ. 관련 연구

익스플로잇 킷(Exploit Kit)은 피싱 공격 툴 중 하나로, 시스템의 보안 취약점을 이용해 악성 코드를 유포시키기 위해 사용하는 해킹 도구이다[1]. EK 감염 경로는 그림 1과 같다.

▸ Contact : 사용자가 정상적인 웹 사이트나 비정상적인 웹 사이트, 이메일, 메신저를 이용한다.

▸ AD Network : 웹 사이트에 익스플로잇 킷으로 리다이렉션하는 악성 코드를 주입하거나 광고에 악성 코드를 주입한다. 이메일을 통해 사용자를 속여 악의적인 링크를 열도록 한다.

▸ Redirect : 사용자가 비정상적인 웹 사이트를 이용하거나 광고를 누르면 해당 브라우저는 중간 페이지를 통해 리다이렉션 된다. 리다이렉션은 공격 최종 URL을 가린다.

▸ Exploit : 시스템 취약성을 식별한 후 익스플로잇 킷을 실행한다. 취약성이 발견되면 시스템을 공격한다.

▸ Infect : 공격이 성공하면 랜섬웨어와 같은 다양한 악성 프로그램을 가져와 실행한다.

본 논문은 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구 (No. 20008902, IT비용 최소화를 위한 5채널 탐지기술 기반 SaaS SW Management Platform(SMP) 개발) 이고, 2021년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과(2021RIS-004)임.

표 2 블랙리스트 기반 식별 정보 유무에 따른 분류 실험 결과

모델	데이터 셋									
	(#1) Raw					(#2) Raw - 블랙리스트 기반 식별 정보 12개				
	학습		테스트		차이	학습		테스트		차이
	정확도	편차	정확도	편차	(학습-테스트)	정확도	편차	정확도	편차	(학습-테스트)
2D-CNN	99.98	0.026	99.70	0.053	0.28	99.86	0.051	98.92	0.185	0.94

Ⅲ. 본론

본 장에서는 수집한 데이터 셋으로 정상 트래픽과 비정상 트래픽의 분류 실험을 진행하였다.

A.데이터셋

정상 트래픽은 직접 수집한 이메일, 카카오톡, 크롬 트래픽이다. 비정상 트래픽은 EK 트래픽이며, Neutrino-EK, Angler-EK, Nuclear-EK 등 총 17종의 EK 트래픽이다[2]. 데이터 셋 설명은 표 1과 같다.

표 1 데이터 셋 구성

분류	플로우		패킷	
	개수	%	개수	%
정상	3036	49	61801	86
비정상	3139	51	10453	14
	6175	100	72254	100

B.실험-블랙리스트기반식별정보유무

첫 번째 실험은 분류 모델이 블랙리스트 기반 식별 정보를 통해 학습을 하는지 알고자하여 실험을 진행하였다. 2D-CNN을 이용하여 정상 트래픽과 비정상 트래픽을 분류하였고, 총 두 가지 데이터 셋을 사용한 분류 실험 결과를 비교하였다. 첫 번째는 Raw 데이터를 사용한 분류 실험(#1)이며, 두 번째는 Raw 데이터에서 블랙리스트 기반 식별 정보를 모두 제거한 분류 실험(#2)이다. 블랙리스트 기반 식별 정보는 블랙리스트에 정의될 수 있으며, EK 개발자가 변조 및 위장할 수 있는 식별 정보를 의미한다. 우리는 본 실험에서 블랙리스트 기반 식별 정보로 총 열두 가지를 정의하였다. ip, port, ethernet, host, location, server, user-agent, referer, get/post, date, accept, accept_language가 있다. 실험 결과는 표 2와 같다. Raw 데이터를 사용한 분류 실험 결과(#1)는 정확도가 높고 편차가 작다. Raw 데이터에서 블랙리스트 기반 식별 정보를 모두 제거한 분류 실험 결과(#2)는 #1에 비해 정확도가 낮고 편차가 크다. 해당 결과를 통해 #1은 블랙리스트 기반의 분류 모델을 만들었을 가능성이 존재한다고 유추할 수 있다. #1의 경우 #2에 비해 알려지지 않은 신종 EK나 위장 트래픽을 탐지하지 못할 가능성이 존재한다. 그러나 블랙리스트 기반 식별 정보의 유무에 따른 결과가 차이가 크지 않은 것을 알 수 있다.

C.실험-블랙리스트기반식별정보변조여부

두 번째 실험은 블랙리스트 기반 식별 정보를 제거하지 않고 해당 정보가 위변조 된 상황을 가정하여 실험을 진행하였다. 기존과 동일하게 정상과 비정상 트래픽 모두 Raw 데이터 셋으로 학습하였다. 본 실험은 비정상 데이터가 정상인 척 위장한 상황을 가정할 때, 제대로 탐지가 가능한지를 알기 위한 실험이기 때문에 테스트 할 때는 정상 트래픽은 Raw 데이터 셋을 사용하였고, 비정상 트래픽은 Raw 데이터의 블랙리스트 기반 식별 정보를 정상 트래픽과 유사하게 변조하여 사용하였다. 우리는 열두 가지 블랙리스트 기반 식별 정보 중 여덟 가지 식별 정보를 변조하였다. 비정상 데이터의 ip, port, user-agent, referer, host, date, accept_language, accept을 정상과 유사하게 변경하여 테스트를 하였다. 실험 결과는 표 3과 같다. 정상과 비정상 Raw 데이터를 학습했을 때 높은 정확도와 작은 편차를 보였지만, 위장 및 변조한 데이터로 테스트 했을 때 낮은 정확도와 큰 편차를 보였다. 해당 결과를 통해 Raw 데이터를 사용하여 학습할 경우 변조되거나 위장한 비정상 트래픽을 탐지하기 어렵다는 것을 알 수 있다.

표 3 블랙리스트 기반 식별 정보 변조 여부에 따른 분류 실험 결과

모델	학습		테스트		차이
	정확도	편차	정확도	편차	(학습-테스트)
2D-CNN	99.99	0.006	66.81	5.156	33.18

Ⅳ. 결론

우리는 익스플로잇 킷과 같은 공격 트래픽의 식별 정보가 변조되거나 정상으로 위장되는 상황을 가정하여 정상 트래픽과 비정상 트래픽의 분류 실험을 진행하였다. 첫 번째 실험은 Raw 데이터 셋을 사용한 분류 실험과 Raw 데이터에서 블랙리스트 기반 식별 정보를 제거한 데이터 셋을 사용한 분류 실험이다. Raw 데이터 셋의 분류 실험 결과(#1)가 Raw 데이터에서 블랙리스트 기반 식별 정보를 제거한 데이터 셋의 분류 실험 결과(#2)보다 정확도가 비교적 높고 편차가 작다. 해당 결과를 통해 모델이 블랙리스트 기반 식별 정보를 이용하여 학습하고, 블랙리스트 기반 식별 정보 없이는 오탐률이 증가할 가능성이 있다고 유추할 수 있지만 큰 차이를 보이진 않는다. 식별 정보는 위변조가 가능하며, EK 개발자는 이를 악용하여 정상 트래픽으로 위장할 수 있다. 본 상황을 가정하여 두 번째 실험을 진행하였고 비정상 Raw 데이터의 식별 정보를 정상 트래픽과 유사하게 임의로 변조하여 정상 트래픽과 비정상 트래픽을 분류하였다. 해당 실험은 블랙리스트 기반 식별 정보를 제거하지 않고 기존 Raw 데이터를 사용하여 탐지 모델을 만들 경우, 블랙리스트 기반 식별 정보를 위변조한 비정상 트래픽을 탐지할 수 있는지 알고자 실험하였다. 변조된 비정상 트래픽으로 테스트한 결과, 기존의 Raw 데이터를 사용한 분류 실험 결과에 비해 정확도가 낮은 것을 알 수 있다. 이는 Raw 데이터를 사용하여 EK 탐지 모델을 만들 경우, 변조 혹은 위장한 EK 혹은 신종 EK를 탐지하기 어렵다는 것을 의미한다.

본 논문은 비정상 트래픽에서 블랙리스트 기반 식별 정보의 위변조 및 신종 EK를 탐지하기 위해 새로운 EK 분석 방법이 필요하다는 것을 의미한다. 향후 연구에서 블랙리스트 기반 식별 정보를 제거하고 변조 및 위장되거나 알려지지 않은 신종 EK를 탐지할 수 있는 모델을 구축할 계획이다.

참 고 문 헌

[1] BURGESS, Jonah, et al. LSTM RNN: detecting exploit kits using redirection chain sequences. Cybersecurity, 2021, 4.1: 1-15.

[2] Malware-Traffic-Analysis.net[Website]. (https://www.malware-traffic-analysis.net/).