딥러닝 기반 익스플로잇 킷 탐지 모델의 입력 데이터 설계에 관한 연구

김보선, 최정우, 이민성, 백의준, 김명섭

고려대학교

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

A Study on the Input Data Design of DL-based Exploit Kit Detection Model

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

Korea University

요 약

인터넷 기술이 발전하며 다방면으로 사용자들에게 여러 이점을 가져와 긍정적인 효과를 제공하지만, 인터넷 사용량이 급속도로 증가함에 따라 웹 페이지를 통한 감염 피해 사례도 늘어나고 있다. 인터넷의 보안 취약성을 악용한 사례 중 하나는 익스플로잇 킷(EK)이다. EK 탐지 방법 중 블랙리스트 기반 탐지 방법은 정보가 변경 혹은 위장된 경우와 알려지지 않은 신종 EK에 대해 취약하여 공격에 대처하지 못하거나 오탐할 확률이 매우 높다. 그래서 우리는 기존 연구에 주로 사용되는 Raw 데이터 셋과 Raw 데이터에서 블랙리스트 기반 식별 정보를 제거한 데이터 셋을 사용하여 실험하였고, Raw 데이터 셋을 사용한 실험은 높은 정확도와 작은 편차를 보였지만 Raw 데이터에서 블랙리스트 기반 식별 정보를 제거한 데이터 셋을 사용한 실험에서는 낮은 정확도와 큰 편차를 보였다. 실험 결과를 통해 블랙리스트 기반 식별 정보가 없이는 오탐률이 증가할 가능성이 있다는 것을 알 수 있다. 우리는 블랙리스트 기반 식별 정보 없이 EK를 탐지할 수 있는 방법에 대한 연구 필요성을 강조한다.

Ⅰ. 서 론

인터넷 기술이 발전하며, 다방면으로 사용자들에게 여러 이점을 가져와 긍정적인 효과를 제공하고 있다. 그러나 인터넷 사용량이 급속도로 증가함에 따라 웹 페이지를 통한 감염 피해 사례도 늘어나고 있다. 인터넷의 개방적인 네트워크 특성과 프로토콜의 보안 취약성을 악용한 사례 중 하나는 익스플로잇 킷(Exploit Kit)이다. EK는 사용자 시스템 속 프로그램의 취약점을 이용해 광고, 이메일 첨부 파일, 웹 페이지 접속 등으로 악성 코드를 유포하는 자동화된 공격 도구이다[1]. 국내 감염 피해를 입힌 주요 EK로는 블랙홀(Blackhole), 앵글러(Angler), 리그(Rig), 뉴트리노(Neutrino) 등이 있다. EK 탐지에 관한 연구가 지속되고 있고, 대표적인 방법으로는 특정 패턴이 지속적으로 사용되는 것을 감지하여 미리 정의된 시그니처를 검사하는 방법(블랙리스트 기반 탐지)이 있다. 그러나 최근 EK는 도메인 생성 알고리즘(Domain Generation Algorithm)과 고도화된 난독 기술, 백신을 우회하는 기능 등을 업데이트하여 IP나 헤더 정보를 변경하거나 호스트 위장 가능성이 존재한다. 도메인 생성 알고리즘은 수많은 도메인 주소를 동적으로 생성하는 알고리즘으로 공격 차단을 우회하는 방법이다. 사용자가 광고 배너를 눌렀을 때, 수백 개의 리다이렉션을 통해 실제 광고 페이지가 아닌 EK 랜딩 페이지로 접속하게 되고 랜딩 페이지의 URI는 계속 변경되며 패턴을 도출하기 어려워져 블랙리스트 기반 탐지로는 EK를 탐지하기에 어려움이 존재한다. 또한 블랙리스트 기반 탐지를 통해 EK를 탐지할 경우, 알려지지 않은 신종 EK에 대해 취약하여 공격에 대응하지 못하거나 탐지하지 못할 확률이 매우 높다. 따라서 블랙리스트 기반의 식별 정보를 사용하지 않는 EK 탐지 모델이 필요하다.

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

우리는 정상 트래픽과 비정상 트래픽(EK)의 분류 모델이 블랙리스트 기반으로 EK 탐지를 하는지 알고자, 기존 연구에 주로 사용되는 Raw 데이터와 Raw 데이터에서 블랙리스트 기반 식별 정보를 제거한 데이터를 사용하여 실험을 진행한다. 두 실험 결과를 비교 및 분석하고자 한다.

Ⅱ. 본론

본 장에서는 정상 트래픽과 비정상 트래픽 분류 실험 과정을 설명한다.

A. 데이터 수집

실험에 사용된 데이터는 직접 수집한 응용 트래픽(정상 트래픽)과 멀웨어 감염과 관련된 EK 트래픽(비정상 트래픽)이며, 자세한 내용은 표 1과 같다[2].

표 1 데이터 셋 정보

프로토콜	분류 대상	이름	플로우	
			개수	%
HTTP	Normal (0)	E-mail	1876	49
		Kakaotalk		
		Chrome		
	Abnormal (1)	Styx-EK	1961	51
		Neutrino-EK		
		Gondad-EK		
		Sibhost-EK		
		Sweet-Orange-EK		
		Blackhole-EK		
		DotkaChef-EK		
		Fiesta-EK		
		Nuclear-EK		
		Angler-EK		
		Rig-EK		
		FlashPack-EK		
		Null-Hole-EK		
		Zemot-EK		

표 2 입력 데이터 정보 및 실험 결과

번호	제거 식별 정보		학습		테스트		차이 (학습-테스트 정확도)
	계층	필드	정확도	편차	정확도	편차	
#1	-	-	99.98	0.04	99.42	0.16	0.56
#2	IP	ip	99.92	0.04	99.10	0.15	0.82
#3	TCP	port	99.95	0.07	99.01	0.23	0.94
#4	Ethernet	ethernet	99.75	0.11	97.76	0.27	1.99
#5	HTTP	host	99.99	0.03	99.22	0.22	0.77
#6	HTTP	location	99.97	0.05	99.12	0.26	0.85
#7	HTTP	server	99.97	0.05	99.12	0.23	0.85
#8	HTTP	user-agent	99.91	0.04	98.87	0.20	1.04
#9	HTTP	referer	99.96	0.05	99.02	0.25	0.94
#10	HTTP	get/post	99.98	0.04	99.07	0.23	0.91
#11	HTTP	date	99.97	0.05	99.02	0.20	0.95
#12	HTTP	accept	99.98	0.04	99.07	0.20	0.91
#13	HTTP	accept_language	99.96	0.05	99.07	0.35	0.89
#14	HTTP	#5~13	99.88	0.07	98.82	0.26	1.06
#15	IP, TCP, Ethernet, HTTP	#2~13	99.74	0.11	97.73	0.29	2.01

B. 데이터 전처리

수집한 Pcap 파일을 JSON 파일로 변환하고 플로우 단위의 데이터로 저장한다. 플로우마다 크기가 다르기 때문에 플로우의 1~3번째 패킷을 추출하고, 부족할 경우 패딩 처리를 하여 동일한 크기의 이미지 형태로 변환한다.

C. 실험

본 연구에서는 2D CNN을 사용하여 정상 트래픽과 비정상 트래픽을 분류하였다. 정상 트래픽과 비정상 트래픽을 섞어 만든 데이터 셋 중 50%는 학습용 데이터로 사용하였고, 나머지 50%의 데이터 셋은 테스트용 데이터로 사용하였다. 과 적합을 방지하기 위해 EarlyStopping이라는 콜백 함수를 사용하여 적절한 시점에 학습을 조기 종료하도록 하였다.

실험은 총 15가지로 구성되어 있으며, 데이터 셋에 포함된 식별 정보에 따라 각각 실험을 진행하였다. 실험 결과는 표 2에 나타내었다. 각 실험은 10번 반복하였고, 학습 정확도와 테스트 정확도는 최대 정확도를 평균 낸 값이다. #1은 Raw 데이터 셋의 분류 실험 결과이고, #2부터 #13은 정상 트래픽과 비정상 트래픽을 학습할 때, 블랙리스트 기반으로 학습될 가능성이 있는 식별 정보를 Raw 데이터에서 각각 제거한 결과이다. 예시로 #2는 Raw 데이터에서 IP 계층의 ip 정보를 제거한 데이터 셋의 실험 결과이며, #3은 Raw 데이터에서 TCP 계층의 port 정보를 제거한 데이터 셋의 실험 결과이다. 비정상 트래픽의 경우, 가상 환경에서 데이터 셋이 수집되었기 때문에 #4는 Raw 데이터에서 ethernet 정보를 제거한 실험 결과이다. #5부터 #13은 Raw 데이터에서 HTTP 계층의 host, location, server, user-agent, referer, get/post, date, accept, accept_language를 각각 제외한 데이터 셋의 실험 결과이다. #14와 #15는 블랙리스트 기반 식별 정보를 조합하여 제거하여 실험하였다. #14는 Raw 데이터에서 HTTP 계층의 9가지 정보(#5~#13)를 제거한 실험 결과이고, #15는 Raw 데이터에서 블랙리스트 기반 모든 식별 정보(#2~#13)를 제외한 실험 결과이다.

실험 결과, 트래픽 분류에 주로 사용되는 Raw 데이터 셋을 사용한 실험(#1)이 높은 정확도와 작은 편차를 보이는 것을 알 수 있다. 또한 Raw 데이터 셋에서 블랙리스트 기반 식별 정보를 모두 제거한 실험(#15)은 비교적 테스트 정확도가 낮고 편차가 크며 학습 정확도와 테스트 정확도의 차이가 큰 것을 알 수 있다. #1과 #15의 실험 결과를 통해 해당 모델은 블랙리스트 기반의 분류 모델을 만들었을 가능성이 크다. Raw 데이터 셋을 사용한 #1의 경우 호스트 정보를 위장하거나 알려지지 않은 신종 공격에 대해 취약하며 우리의 가정대로 분류 모델이 해당 식별 정보를 사용하여 EK를 탐지한 경우라면 #15는 #1에 비해 오탐률이 증가할 가능성이 있다. 호스트 정보가 지속적으로 변경되거나 위장하는 상황, 신종 공격을 대비하여 위장 데이터 셋의 위험 가능성을 낮추고자 블랙리스트 기반 식별 정보를 제거한 실험(#15)은 정확도가 낮게 나옴을 알 수 있고, 기존 연구 방법인 Raw 데이터 셋을 사용한 실험(#1)은 블랙리스트 기반으로 학습을 진행하였음을 알 수 있다.

Ⅲ. 결론

우리는 EK가 지속적으로 업데이트되어 호스트 정보를 변경하거나 위장하여 블랙리스트 기반 탐지 모델이 EK를 탐지하기 어려워졌고 알려지지 않은 신종 EK에 대해 대처하지 못하고 오탐할 수 있다는 가정에 기반하여 실험을 진행하였다. 본 논문은 Raw 데이터 셋과 Raw 데이터에서 블랙리스트 기반 식별 정보를 제거한 데이터 셋을 이용하여 15가지 실험을 진행하였다. Raw 데이터 셋을 사용한 실험 결과는 99.42%의 높은 분류 정확도와 작은 편차를 보였고, Raw 데이터에서 블랙리스트 기반 식별 정보를 제거한 데이터 셋을 사용한 실험 결과는 비교적 낮은 분류 정확도와 큰 편차를 보였다. 이를 통해 블랙리스트 기반 식별 정보가 없이는 오탐률이 3.9배 증가함을 알 수 있었고, Raw 데이터 셋을 사용할 경우 해당 모델은 블랙리스트 기반의 탐지 모델을 만들 가능성이 크다는 것을 알 수 있었다. 이는 호스트 정보를 변경하거나 위장한 EK를 탐지하지 못하며 신종 EK를 탐지하지 못할 가능성이 크다.

우리는 EK를 비롯한 다양한 악성 트래픽 데이터 셋을 추가로 수집하고 블랙리스트 기반 학습에 사용될 수 있는 식별 정보를 제거한 정상 트래픽 및 비정상 트래픽을 통계 특징 기반 머신러닝 또는 CNN 기반 딥러닝을 통해 강인한 EK 탐지 모델을 구축하는 연구를 수행할 계획이다.

참 고 문 헌

[1] QIN, Yan, et al. An exploit kits detection approach based on http message graph. IEEE Transactions on Information Forensics and Security, 2021, 16: 3387-3400.

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