암호화 트래픽 분석을 통한 Multi-Modal Signature 기반의 악성 행위 탐지

박지태, 최정우, 김주성, 유경민, 신창의*, 김명섭

고려대학교, 국방기술품질원*

{pjj5846, choigoya97, jsung0514, rudals2710, tmskim}@korea.ac.kr, *superego99@dtaq.re.kr

Malicious Behavior Detection based on Multi-Modal Signatures through Encrypted Traffic Analysis

Jee-Tae Park, Jeong-Woo Choi, Ju-Sung Kim, Gyeong-Min Yu, Chang-Yui Shin*, Myung-Sup Kim

Korea Univ., *Defense Agency for Technology and Quality

요 약

최근 네트워크 기술의 발전과 고도화에 따라 네트워크 내 다양한 악성 행위가 발생하고 있다. 네트워크 악성 행위는 개인뿐만 아니라 기업 및 단체를 대상으로 큰 피해를 줄 수 있기 때문에 악성 행위를 탐지 할 수 있는 방법에 대해 연구가 수행되고 있다. 대표적인 방법으로 시그니처 기반의 방법이 있으며, 이는 사전에 악성 트래픽에 대한 공통 시그니처 및 패턴을 기반으로 악성 행위를 탐지하는 방법이다. 하지만 암호화 트래픽의 사용이 점차적으로 증가하고, 악성 행위가 점차적으로 고도화 되면서 공통 시그니처를 추출하기 어려워졌다. 본 논문에서는 암호화 트래픽을 사용하는 악성 행위를 대상으로 Multi-Modal Signature 기반의 악성 행위 탐지 방법을 제안한다. 제안하는 방법은 암호화된 TLS 트래픽을 대상으로 트래픽의 헤더, 통계 정보와 암호화되지 않은 TLS Handshake과정에서 SNI (Server Name Indication) 정보를 활용하여 악성 행위의 공통 시그니처를 추출한다. 본 논문에서는 실제로 수집된 악성 트래픽에 제안하는 방법을 적용하여 대상 악성 행위에 대한 공통 시그니처를 추출한다.

Ⅰ. 서 론

최근 과학 기술의 발전과 네트워크 사용량의 급증에 따라 네트워크 환경 규모는 점차적으로 복잡해지고 있다. 네트워크 관리 연구는 네트워크 내 원활한 서비스 제공과 효율적인 리소스 관리, 성능 최적화 뿐만 아니라 네트워크 내 발생하는 여러 가지 악성 행위를 탐지하고 신속하게 대응하는 보안 측면에서도 중요한 역할을 한다[2]. 기존에는 사전에 악성 행위 트래픽의 헤더 및 페이로드를 분석하여 공통 시그니처를 추출하고, 이를 기반으로 악성 행위를 탐지하였다. 하지만 암호화 트래픽의 보급화로 데이터가 암호화되어 전송되기 때문에, 탐지 성능이 크게 낮아졌다. 또한, 악성 행위도 점차적으로 복잡해지고, 고도화되고 있으며, 암호화 트래픽을 이용한 악성 행위들이 증가하고 있다. 암호화 트래픽에 대한 분석은 대부분 머신러닝 및 딥러닝을 활용한 학습 기반의 방법을 사용하며, 특히 딥러닝을 적용한 분석 방법은 암호화 트래픽을 대상으로도 높은 정확도를 보이지만, 높은 성능을 도출하기 위해 많은 양의 라벨링 된 데이터가 필요하다는 한계점이 있다. 또한, 많은 시간과 리소스가 소모되기 때문에 실시간 악성 행위 탐지하는 네트워크 보안 분야에 적용하기에는 부적합하다.

본 논문에서는 기존 연구 방법 중 하나인 시그니처 기반의 분석 방법을 제안하며, 트래픽의 다양한 특성을 반영하여 여러 가지 트래픽 정보를 활용한 Multi-Modal Signature를 사용한다. 제안하는 방법은 암호화 트래픽을 대상으로도 적용 할 수 있도록 트래픽 헤더 및 통계 정보와 통신 과정 중 암호화 되지 않은 TLS Handshake에서 SNI 정보를 사용한다.

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

본 논문의 구성은 본 장의 서론에 이어 2장 관련 연구에서 악성 행위 탐지 연구 동향에 대해 설명한다. 이 후, 3장 본론에서 제안하는 시스템과 실제 악성 행위 트래픽으로 부터 생성된 시그니처에 대해 설명하고, 4장에서 결론 및 향후 연구를 제시하며 논문을 마친다.

Ⅱ. 관련 연구

네트워크 보안 분야에서 악성 행위 탐지 연구는 오래전부터 수행되어 왔다. 초기에는 악성 행위에 대한 트래픽 분석을 통해 악성 행위의 특징을 시그니처로 정의하고, 이를 기반으로 악성 행위를 탐지하는 시그니처 기반의 분석 방법이 수행되었다[1]. 하지만 기존 시그니처 기반 분석 방법은 악성 행위 별 공통된 헤더 및 페이로드 특징을 활용하여 시그니처를 정의하였지만, 암호화 트래픽을 대상으로는 적용 할 수 없다는 한계점이 있다.

학습 기반의 분석 방법은 사용하는 알고리즘에 따라 머신러닝 기반과 딥러닝 기반으로 나누어진다[1,2]. 학습 기반의 분석 방법은 복잡하고 고도화된 공격을 대상으로도 높은 성능으로 탐지 할 수 있지만, 높은 성능을 위해 많은 양의 정상, 악성 행위 트래픽 데이터가 필요하다[1]. 악성 행위 트래픽은 보안 회사 혹은 연구소 이외의 일반 연구자들이 얻기 힘들기 때문에 많은 연구에서는 CICDS2017, UNSW-NB15와 같은 공개된 데이터 셋을 사용한다[1]. 하지만 공개된 데이터 셋은 정상 트래픽 양에 비해 부족하며, 데이터 불균형 문제는 탐지 성능을 저하 시킨다.

Ⅲ. 본론

(1) 시스템 구조

제안하는 시스템 구조는 그림 1에 나타나있으며, 크게 두 가지 세부 모듈로 구성되어 있다. 시그니처 기반의 분석 방법은 악성 행위 트래픽의 공통

특징을 분석한다. 악성 행위 별로 공격 방법, 대상 패턴 등이 다르기 때문에 트래픽 특징도 악성 행위 별로 달라진다. 따라서 시그니처 기반 분석 방법을 적용하기 위해서는 사전에 대상 악성 행위를 정하고, 대상 악성 행위에 대한 트래픽만 수집하여 시그니처를 추출해야 한다. 시그니처 생성 과정에서 여러 가지 악성 행위 트래픽이 섞이면 잘못된 시그니처가 추출될 수 있다. 악성 트래픽은 [3]으로부터 수집하였으며, 악성 행위별로 데이터 셋을 분류 하였다.

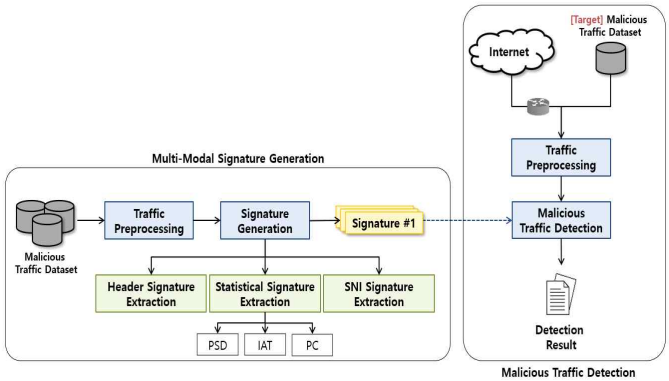

그림 1. 시그니처 생성 및 악성 트래픽 탐지 시스템 구조

먼저 Multi-Modal Signature 생성 모듈은 전처리 과정과 시그니처 생성 과정으로 구성된다. 전처리 과정은 입력으로 들어오는 대상 악성 행위 트래픽 셋을 대상으로 수행된다. 수집된 악성 트래픽은 패킷 단위의 pcap 형태로 되어있으며, 시그니처를 생성하기 위해 5-tuples 정보가 같은 패킷들의 집합인 플로우 단위로 변환한다. 시그니처 생성 모듈은 입력으로 들어오는 플로우 단위의 트래픽을 대상으로 헤더, 통계, SNI 시그니처를 추출한다.

헤더 시그니처는 대상 악성 행위에서 공통적으로 발생하는 헤더 정보를 나타내며, 주로 고정된 포트 번호를 사용하는 악성 행위를 탐지하는데 사용된다. 통계 시그니처는 대상 악성 행위의 플로우에서 추출할 수 있는 여러 가지 통계 정보를 나타내며, 제안하는 방법에서는 PSD (Packet Size Distribution), IAT (Packet Inter Arrival Time), PC (Packet Count)를 사용한다. SNI 시그니처는 암호화 통신에서 악성 도메인과 관련된 SNI 정보를 시그니처로 정의한다. 각각의 시그니처를 활용한 악성 행위 탐지 방법은 과거에 높은 성능을 보였지만, 악성 행위들이 점차적으로 고도화되면서 개별 시그니처의 성능 저하와 한계점이 나타났다. 제안하는 방법은 세 가지 시그니처를 모두 적용하여 기존의 개별 시그니처 적용 방법의 한계점을 해결하는 것을 목표로 한다.

악성 트래픽 탐지 모듈에서는 일반 응용으로부터 수집된 정상 트래픽과 대상 악성 행위 트래픽을 함께 넣고, 시그니처 생성 모듈과 동일한 전처리 과정을 수행한다. 이 부분에서 정상, 악성 트래픽이 라벨링되어 있으며, 각 트래픽 정보를 활용하여 탐지 성능을 평가한다. 다음으로 플로우 기반의 트래픽과 생성된 시그니처를 입력으로 악성 트래픽을 탐지하고, TP, FP, FN, TN을 활용하여 Recall, Precision을 계산한다.

(2) 시그니처 분석

제안하는 방법을 실제 악성 행위 트래픽에 적용하여 시그니처를 추출하였다. 대상 악성 행위는 Qakbot으로 선정하였으며, [3]에서 10 가지 Trace의 트래픽 셋을 수집하였다. Qakbot은 악성 코드를 활용하여 정보 탈취, 피싱 공격 등의 악성 행위를 수행한다. Qakbot는 유사한 악성 행위 패턴을 보인다. 예를 들어 악성 행위 중에 대용량 데이터를 탈취하고 전송하는 경향을 보이며, 이는 패킷 사이즈, 개수 등의 유사한 통계적 특징이 나타난다.

표 1은 Qakbot을 대상으로 추출된 시그니처 정보를 나타낸다. 헤더 시그니처는 기본적으로 TCP 443 포트의 암호화 트래픽을 사용하기 때문에 추출되지 않았다. PSD는 플로우 별로 가변적인 정수형 벡터 값을 가지며, TLS Handshake 과정을 포함하는 1~7 번째 패킷의 평균을 나타낸다. 이 중에서 1~3 번째 패킷 사이즈는 고정된 값이 도출된다. IAT는 플로우 별로 최소, 최대, 평균으로 나타내며, PC는 플로우 내 평균 패킷 수를 나타내었다. SNI 시그니처는 Qakbot에서 추출되지 않았으며, 악성 행위 특성 상 시스템을 감염시키기 위해 HTTP를 사용하기 때문으로 판단된다. Qakbot은 악성 행위 특성상 주로 통계 정보에 대한 공통 특징이 있는 것으로 보이며, 악성 트래픽 대한 공통 특징은 악성 행위에 따라 달라질 수 있다.

표 1 Qakbot에 대한 시그니처 정보

대상 악성 행위	시그니처 종류		시그니처 내용
Qakbot	Header		*None*
	Statistical	PSD	[66, -58, 54, 571, -54, -1430, -1514]
		IAT	Min : 10~20 ms
			Max : 500 ms
			Mean: 100~200 ms
		PC	Mean : 50~100
	SNI		*None*

Ⅳ. 결론 및 향후 연구

본 논문에서는 서론에서 네트워크 트래픽 분석 연구 배경과 악성 행위 탐지의 필요성에 대해 설명하고, 관련 연구에서 기존의 여러 가지 연구 방법들을 소개한다. 기존 연구 방법 중 시그니처 기반 방법의 한계점을 해결하기 위해 Multi-Modal Signature 기반의 악성 행위 탐지 방법을 제안한다.

제안하는 시스템은 시그니처 생성과 악성 트래픽 탐지 모듈로 구성되어 있으며, 대상 악성 행위에 대해 세 가지 트래픽 정보를 활용한다. 시그니처 생성 모듈에서는 헤더, 통계, SNI에 대한 공통 패턴을 여러 가지 시그니처를 추출한다. 또한 추출된 시그니처를 활용하여 악성 행위를 탐지한다.

하지만 제안하는 방법은 기존 시그니처 기반 방법의 한계점을 해결 할 수 있으며, 암호화 트래픽을 대상으로 적용 가능하지만, 여러 가지 한계점이 있다. 첫 번째로 본 논문에서는 실제 악성 트래픽에 대한 시그니처를 추출하지만, 이를 기반으로 탐지 성능에 대한 검증이 필요하다. 두 번째로, 제안하는 방법을 포함한 시그니처 기반 방법은 트래픽 패턴이 변할 경우, 성능이 크게 저하된다.

따라서 향후 연구로 제안하는 방법을 개선하여 트래픽 패턴이 변하더라도 적용 할 수 있는 자동 시그니처 업데이트 매커니즘에 대해 연구 할 예정이며, 추가적으로 학습 기반 방법을 함께 적용하여 한계점을 해결 할 예정이다. 또한, 추출된 시그니처를 활용하여 악성 트래픽 탐지 실험을 통해 제안하는 방법에 대한 타당성을 검증 할 예정이다.

참 고 문 헌

[1] Wang, Z., Fok, K. W, and Thing, V. L. “Machine learning for encrypted malicious traffic detection: Approaches, datasets and comparative study”. Computers & Security, 113, 2022, 102542.

[2] 전덕조, 박동규. “머신 러닝 (Machine Learning) 기법을 활용한 암호화된 TLS 트래픽 내 악성코드 탐지 기법”. 한국정보기술학회논문지, 19(10), 2021, pp. 125-136.

[3] https://www.malware-traffic-analysis.net/index.html