마르코프 체인 모델 기반 암호화된 악성 트래픽의 패밀리별 분류

배국태, 최수빈, 노희준
고려대학교 응용수리과학부 사이버보안전공
{920ktbae, subin0630, hjroh}@korea.ac.kr

Classifying Encrypted Malware Traffic by Family Based on Markov Chain Model

Kuktae Bae, Subin Choi, and Heejun Roh
Cyber Security Major, Division of Applied Mathematical Sciences, Korea University

요 약

인터넷에서 사용자 간 주고받는 데이터의 보안에 대한 중요성이 대두되면서, 많은 웹 응용이 통신보안을 제공하기 위해 설계된 프로토콜인 Transport Layer Security(TLS)를 사용하고 있다. 그러나 악성 행위를 하는 공격자도 이를 악용해 악성코드를 암호화하여 전송하고 있으며, 기존의 방법으로는 탐지하지 못하거나 분류의 일반성이 떨어져 한계가 발생한다. 본 논문에서는 멀웨어의 암호화된 악성 트래픽 분류 문제를 해결하기 위해, 마르코프 체인(Markov chain)모델에 기반을 둔 암호화된 악성 트래픽의 패밀리별 분류를 논의한다. 공개된 악성 트래픽 pcap 파일을 바탕으로, pcap 파일 안에 존재하는 TLS 플로우에 대한 TCP 페이로드의 길이열을 수집하고 이를 세분화된 멀웨어 패밀리명을 기준으로 분류하였다. 분류한 패밀리들을 바탕으로 마르코프 모델과 최대우도 추정 기법을 결합하여 멀웨어 패밀리 분류를 시도해 본 결과, TLS 플로우를 생성한 멀웨어들을 패밀리별로 구분 가능함을 알아낼 수 있었다.

Ⅰ. 서 론

오늘날 인터넷에서는 종단 간 보안 및 사용자 사생활 보호가 필수 요구사항이 되며 네트워크 상의 트래픽에서 암호화된 트래픽의 비율이 높아지고 있다. Let's Encrypt에 따르면, Firefox 브라우저에서 HTTPS가 사용된 웹페이지의 비율은 2020년 6월 기준 84%에 달했다.[1] 한편으로, 악성행위를 하는 공격자도 보안 관제자가 암호화된 정보를 쉽게 볼 수 없다는 점을 악용하여 악성 코드를 암호화하여 전송하고 있다. SonicWall Capture Labs의 2020 사이버 위협 보고서[2]에 의하면, TLS 및 SSL 암호화 표준 기반의 HTTPS를 거친 멀웨어의 공격이 전년 동기 대비 27.3% 늘어났다고 보고했다. 트래픽이 암호화됨에 따라 기존의 심층 패킷 조사(DPI) 또는 시그니처(signature) 기반의 위협 분석 기법은 멀웨어 탐지에 적용할 수 없게 되었다[3]. 이러한 문제를 극복하기 위해 TLS 플로우 및 네트워크적 특징을 중심으로 멀웨어 분류를 시도한 연구[3]가 있었으나, 이러한 기존 연구는 악성 행위와 트래픽 간의 연관성에 집중하기보다는 단순히 특징 집합에 대해 머신러닝 기법에 적용하기 때문에 훈련 집합에 따라서 성능이 저하될 수 있다는 문제점이 발생한다.

한편, [4]에서는 전문가가 작성한 사고 보고서를 바탕으로 TLS 암호화된 악성 플로우를 분류한 뒤, 악성 플로우를 이루는 TCP 페이로드(payload)의 길이 열에 대한 마르코프 체인(Markov chain) 기반 모델을 정립하고, 최대우도(Maximum Likelihood) 추정으로 악성 행위를 판단하는 기법의 가능성을 논의했다. 해당 논문은 각 멀웨어의 동작이 결정하는 페이로드 길이 열만으로도 악성 트래픽을 분류할 수 있음을 보여준 것에 큰 의미가 있다. 하지만 아쉽게도 Exploit Kit으로 분류된 멀웨어의 다양성으로 정확도가 매우 낮은 문제가 있었다.

본 논문은 이전 연구에서 멀웨어 분류의 낮은 세밀성이 마르코프 체인 모델의 확률 값에 영향을 주는 문제를 극복하기 위해, 악성 행위를 기반으로 멀웨어를 보다 세분화해 분류한 뒤 마르코프 체인 모델에 적용하고자 한다. 그리고 이를 적용한 결과값을 바탕으로 최대우도 추정의 결과값을 확인하는 실험을 통해 추론을 검증함으로서 악성 트래픽을 패밀리별로 분류 가능함을 논하려 한다.

Ⅱ. 데이터 수집 및 처리 방법

본 연구에서는 데이터의 일관성을 유지하기 위해 보안 분야의 한 전문가가 2013년 6월 18일부터 2020년 5월 29일까지 공개한 악성 트래픽 pcap파일과 사고 보고서[5]를 수집하였다. 사고 보고서는 악성 트래픽이 담긴 pcap파일을 보안 분석 도구인 Wireshark 등을 활용해 분석하였다. Pcap파일 내에 존재하는 20,127개의 TLS 플로우에 대한 TCP 페이로드의 길이열을 수집한 후, TCP 페이로드의 길이를 총 10개의 범주로 나누어 마르코프 체인 기법을 적용했다.

또한 사고 보고서를 통해 확인한 악성 행위에 따라 악성 TLS 플로우를 가지는 멀웨어를 지정 후 세분화하여 각 플로우에 패밀리 명으로 레이블(label)을 부여하였다. 세분화하는 과정에서 수가 너무 적거나, 다양한 공격기법을 이용하는 Exploit Kit 행위의 특징으로 인해 제대로 분류할 수 없다고 판단한 경우, 해당하는 멀웨어를 others 레이블에 포함하였다.

본 논문은 [4]에서의 방법을 참고하여 TCP 페이로드의 길이를 수집하였으나, 레이블을 부여할 때 행동에 따른 광범위한 분류가 아닌 세분화된 패밀리별로 레이블링을 했다는 점에서 차별점을 두었다.

Ⅲ. 제안 기법 및 실험 결과

우선 Exploit kit 행위와 관련된 TLS 플로우를 패밀리별로 분류하기 위해, 2절에서 언급한 데이터 처리 기법을 적용해 멀웨어 패밀리 명에 따라 Θ={Seamless, Spelevo, Styx, others}로 세분화해 레이블링하였다. 그리고 마르코프 모델과 최대우도 추정 기법을 결합해 판정 결과를 얻었다. 그림 1은 각 레이블에 대한 데이터셋에 대해 제안 기법이 판정한 혼동행렬(confusion matrix)을, 클래스 별 판정 정확도를 쉽게 비교할 수 있도록 막대 그래프로 나타낸 것이다. 여기서 가로축은 입력으로 주어진 데이터셋의 레이블이며, 세로축은 해당 분류의 정확도(accuracy)를 나타낸다. 각 클래스는 98.20%, 52.12%, 98.96%, 8.34%의 정확도를 갖는 것으로 나왔다. 패밀리 명으로 분류한 세가지 클래스의 정확도 결과는, 멀웨어 행동을 기반으로 세분화된 분류가 전제된다면 악성 트래픽의 다양성이 높은 Exploit Kit 플로우도 패밀리별로 분류해낼 수 있음을 시사한다. 한편, others 클래스의 경우 매우 낮은 정확도를 보이는데, 이는 데이터를 정제하는 과정에서 세분화해 패밀리를 분류하지 못해서 발생한 것으로 판단된다.

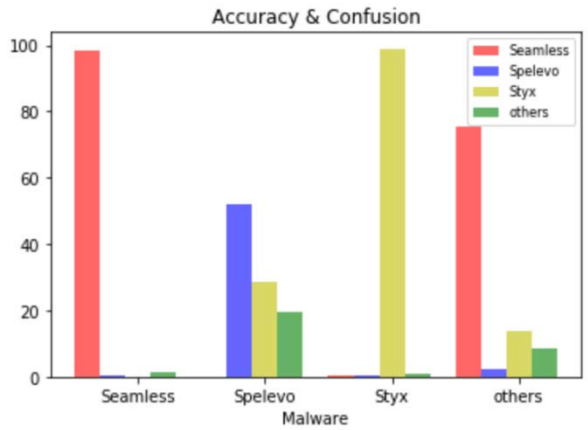

그림 1. Exploit Kit 의 암호화된 트래픽 분류결과.

앞선 실험을 바탕으로 다른 종류의 멀웨어 패밀리들도 분류 가능한지 확인하기 위해 추가적인 실험을 진행했다. Exploit Kit행위와 관련된 멀웨어 패밀리 데이터와 Malspam행위에 포함된 패밀리인 Hancitor 데이터, 그리고 Ransomware행위에 포함된 Ursnif 데이터를 가지고 2절의 형식처럼 Θ={Spelevo, Styx, Hancitor, Ursnif}로 레이블링을 진행했다. 판정 결과는 그림 2처럼 나왔고, 각 클래스의 정확도는 48.65%, 70.90%, 77.62%, 65.41%를 갖는 것으로 나왔다. 해당 결과는 다른 행위를 하는 TLS 플로우 데이터에 대해 각 패밀리별로 세분화해서 레이블링을 진행하면, 악성 트래픽을 멀웨어 패밀리별로 분류할 수 있음을 시사한다.

본고에서는 멀웨어의 행동을 기반으로 패밀리 명에 맞춰 악성 TLS 플로우 데이터셋을 수동으로 분류하고 레이블링했다. 하지만 이는 마르코프 체인 모델이 성공적으로 동작하기 위해 매번 전문가에 의한 세밀한 정제가 필요함을 나타낸다. 따라서 후속 연구에서는 데이터 정제를 자동화하는 방법을 연구할 예정이다.

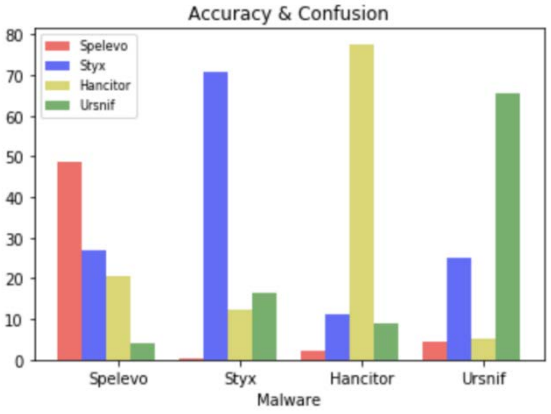

그림 2. 제안 기법의 암호화된 트래픽 분류결과.

Ⅳ. 결론

본 연구는 멀웨어 행동을 기반으로 패밀리 명에 맞춰 악성 TLS 플로우 데이터셋을 세분화해 레이블링 한 뒤 마르코프 체인 모델과 최대우도 추정 기법을 적용하면 멀웨어 패밀리별 분류를 정확하게 해낼 수 있다는 것을 확인하였다. 본 연구는 악성 TLS 플로우에서 TCP 페이로드의 길이 열 만을 특징으로 사용했지만, 멀웨어 행동을 기반으로 세분화된 분류가 전제된다면 보다 정확히 멀웨어를 패밀리별로 분류해낼 수 있음을 확인하였다.

ACKNOWLEDGMENT

This work was supported by KISTI.

참 고 문 헌

[1] Let's Encrypt Stats, "Percentage of Web Pages Loaded by Firefox Using HTTPS," 2020, (https://letsencrypt.org/stats/).

[2] SONICWALL, "2020 SONICWALL 사이버 위협 보고서," 2020.

[3] B. Anderson, and D. McGrew, "Identifying Encrypted Malware Traffic with Contextual Flow Data," in *Proc of ACM AISec* co-located with *ACM CCS*, October 2016.

[4] 조영복, 류희정, 김동언, 신원근, 고명훈, 노희준, "암호화된 악성 트래픽 분류를 위한 마르코프 체인 모델의 가능성 분석," *한국통신학회 학술대회논문집*, pp. 469-470, 2019.

[5] B, Duncan, "Malware Traffic Analysis [Online]," 2019, (http://malware-traffic-analysis.net/).