DHT 프로토콜 트래픽을 활용한 Mozi 봇넷 탐지 모델에 관한 연구

김대현[1], 이지수[2], 문종인[3], 박정우[4], 유동영[5]
[1 2 3 4]홍익대학교 소프트웨어융합학과 학부생
[5]홍익대학교 소프트웨어융합학과 교수
et9904@naver.com, wltn4752@naver.com, whddls5206@naver.com,
wjddn1815@gmail.com, ydy@hongik.ac.kr

A Study on the Mozi Botnet Detection Model Using DHT Protocol Traffics

Dea-Hyeon Kim[1], Ji-Sue Lee[2], Jong-In Mun[3], Jung-Woo Park[4],
Dong-Young Yoo[5]
[1 2 3 4 5]Dept. of Software and Communications Engineering, Hongik University

요 약

확장하는 디지털 인프라에 발맞추어 IoT 산업또한 점점 시장을 넓히고 있다. IoT 보안위협에 대한 대응준비는 아직 미흡하며 Mozi 봇넷 등 신규 IoT 봇넷의 등장과 증가하는 피해사례는 상황을 더욱 악화시키고 있다. 이에 본 논문에서는 Mozi 봇넷의 동작 원리를 기반으로 한 DHT 프로토콜의 흐름의 특징을 네트워크 로그에서 추출하고 이를 기계학습에 적용하는 탐지모델을 제안한다.

1. 서론

고도화되는 IT산업 속에서 디지털 인프라의 구축환경은 점점 넓어지고 있고, IoT 산업 또한 이에 발맞추어 점점 시장을 넓히고 있다. 하지만, IoT 보안위협은 계속 증가할 것으로 예상하나 대응준비는 미흡하다고 평가된다[1]. 최근 등장한 Mozi botnet은, 2022년 72개국의 1만 2,000여 대를 감염해 암호화폐 채굴 악성코드 유포를 위한 경유지로 활용하였다[2]. 이에 Mozi botnet 분석 연구가 활발히 진행되고 있으며, Teng-Fei Tu 외 5명은 Mozi 봇넷에 대한 종합적인 연구를 수행해 Mozi 봇넷의 통신 원리와 분산 해시 테이블 프로토콜 및 Mozi 봇넷의 확산 분포 현황 등을 다루었다[3]. 이에 본 논문에서는 DHT(Distributed Hash Table) 프로토콜 네트워크 트래픽을 분석하여 Mozi 봇넷의 특징을 추출해 기계학습에 적용하는 탐지모델을 제안한다.

2. Mozi botnet 특징 분석

2.1. Mozi botnet 기본 구조 분석

Mozi 봇넷은 P2P(Peer-to-Peer) 프로토콜 기반의 악성 코드로, “Botmaster”, “공용 DHT 노드”, “Mozi 노드”로 크게 구성한다. 공격자는 Botmaster를 통해 개인키로 암호화된 구성파일을 Mozi 노드에게 UDP로 전송하고, Mozi 노드들은 다른 모든 Mozi 노드와 구성파일을 동기화한다.

공격자는 Mozi 노드를 구성 파일의 동기화를 통해 제어한다. Mozi 구성파일은 일반 텍스트로 로컬에 저장되고, 노드의 ID, 정보 전송 방법 등 다양한 정보가 포함되어 있다. Mozi 노드는 구성 파일에 대해 서명확인을 수행하며, 서명 확인이 된 구성 파일에 대해서만 실행이 된다[3]. 서명 알고리즘은 ECDSA384 알고리즘을 사용하였고, 비대칭 XOR 암호화 방식을 통해 구성 파일을 암호화한다.

2.2. Mozi DHT 프로토콜 분석

DHT은 분산 해시 테이블이라 불리며, P2P 알고리즘으로 해싱을 통해 생성된 데이터들의 위치 정보를 시스템에 포함되어 있는 모든 노드들에게 균일하게 분산하기 위하여 고안된 lookup기법이다[4]. Mozi 봇넷은 독자적인 DHT 프로토콜을 갖추었으며, 이는 정상적인 DHT 프로토콜 네트워크 상에서 동작한다. Mozi 봇넷에 감염된 노드는 노드 ID에 관한 요청에 대해 노드 ID를 반환하거나 자신의 구성 파일을 반환하는 행동을 랜덤하게 보여준다[5]. 2.1장에서, 공격자는 구성 파일을 Mozi DHT 네트워크 상에 유포하고 이를 동기화하는 방식으로 관리를 한다고 설명하였다. 노드 ID 반환과 구성 파일 반환이

랜덤하게 이루어지므로, 구성 파일을 동기화 하기 위해서는 일반적인 DHT 노드보다 더 많은 이웃 리스트 요청을 송신해야한다. 이는 정상 DHT 노드와 Mozi DHT 노드를 구분할 수 있는 중요한 특징으로, 현재 노드의 감염 여부 및 통신하는 상대 노드의 감염 여부를 파악할 수 있는 지표가 된다.

3. 네트워크 트래픽 추출 설계

네트워크 환경에서 호스트에 접속하여 통신이 이루어질 때 로그가 발생한다. 한 노드에 Mozi 등록이 완료되면 Mozi 봇넷은 공개 DHT 노드를 통해 DHT find_nodes 요청을 UDP 프로토콜로 보내기 시작한다[3]. 이는 일반 DHT 로그와 Mozi 봇넷 DHT로그를 구별하는 큰 특징으로, 이러한 요청을 확인하기 위해 본 논문에서는 의심되는 노드의 UDP 패킷을 모니터링하고 DHT 프로토콜 패킷을 추출 및 시각화 하여 정상 패킷과 Mozi botnet 패킷 사이의 유사도를 파악한다.

4. 기계학습 기법 분석

이미지와 유사한 형태의 데이터를 처리하는 방법에는 기계학습 방법에는 CNN(Convolutional Neural Network) 모델을 사용한다. CNN은 입력 이미지에 대해 여러 개의 합성곱(Convolution) 연산을 수행하고, 그 결과를 다층 신경망에서 처리하여 출력을 만들어낸다. 이 과정에서 이미지의 특징을 추출하고, 이를 이용하여 이미지를 분류하거나 객체를 인식하는 등의 작업을 수행한다.

2.2장의 내용에서, Mozi 노드는 일반적인 노드보다 더 많은 통신과정이 이루어진다고 설명하였다. 때문에 정상적인 노드와는 달리, 특정 노드들에 대해 더 많은 DHT 이웃 요청 트래픽을 송신할 것이며, 이는 시각화 된 자료로 확인할 것으로 기대한다.

5. 실험 방법 및 탐지모델 설계

본 연구는 상기한 Mozi 봇넷의 특징과 네트워크 트래픽 탐지 및 기계학습 모델의 특징을 바탕으로 [그림 1]의 모델을 제안한다. 의심 노드의 패킷을 CapAnalysis를 통해 수집한 후, 이를 UDP 프로토콜 필터링해 DHT 통신 패킷만 추출한다. 추출한 패킷을 트래픽 시각화를 통해 이미지화 과정을 거친 후, CNN 모델을 통해 강화학습을 지도한다. 마지막으로, 학습이 끝난 후 모델을 적용하여 탐지율을 분석한다.

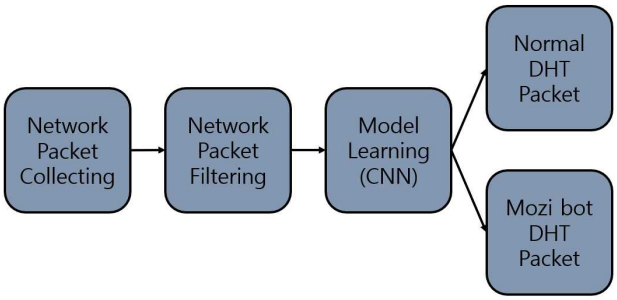

그림 1. 탐지모델 구성

6. 결론 및 향후 연구

본 연구에서는 Mozi 봇넷의 패킷 흐름의 특징을 동적 분석으로 탐지하는데 중점을 둔다. 또한 CNN 모델을 사용하여 강화학습을 통해 더욱 높은 탐지율을 가지며, 이는 향후 진행하게 될 실제 모델 구현에서 모델의 높은 탐지율을 긍정적으로 기대하고 있다. 단, 본 연구에서 진행하는 모델이 Mozi 봇넷이 가지는 여러 특성 중 하나만을 기반으로 탐지하기 때문에 최적의 모델을 구성하였다고 판단하기 어렵다. 따라서, 특정 패킷 길이, node필드의 hash값 등 Mozi 봇넷이 가지는 다른 특성을 바탕으로 탐지하는 모델에 관한 연구가 필요할 것이다. 추가적으로, 본 논문의 모델과 Opcode 및 API 등 정적 분석을 활용한 모델을 통합적으로 사용한 모델을 연구하면 더욱 향상된 연구 결과가 나올 것으로 기대한다.

참고문헌

[1] 민경식 외 2명, "2030 미래사회 변화 및 ICT 8대 유망기술의 사이버 위협 전망", KISA Insight, Vol.1, 2022

[2] 원병철, "전 세계 IoT 장비 1만 2,000여 대, 'Mozi 봇넷' 감염됐다", 보안뉴스, 2022

[3] Teng-Fei Tu 외 5명, "A comprehensive study of Mozi botnet", Wiley Periodicals LLC, 2022

[4] 이유진 외 2명, "효과적인 파일 공유를 위한 다중 링 기반 DHT 프로토콜", 추계종합학술대회, 한국해양정보통신학회, 2007, 506쪽

[5] Alex Turing 외 1명, "Mozi, Another Botnet Using DHT", Netlab360, 2019