

머신러닝 기반 IoT 장치를 식별하는 기법 분석

2023 한국차세대컴퓨팅학회

안석현, 박민수, 조성제, 김홍근

INDEX

01

서론

02

배경 지식

03

머신 러닝 기반 IoT 장치 식별 연구 비교

04

논의

05

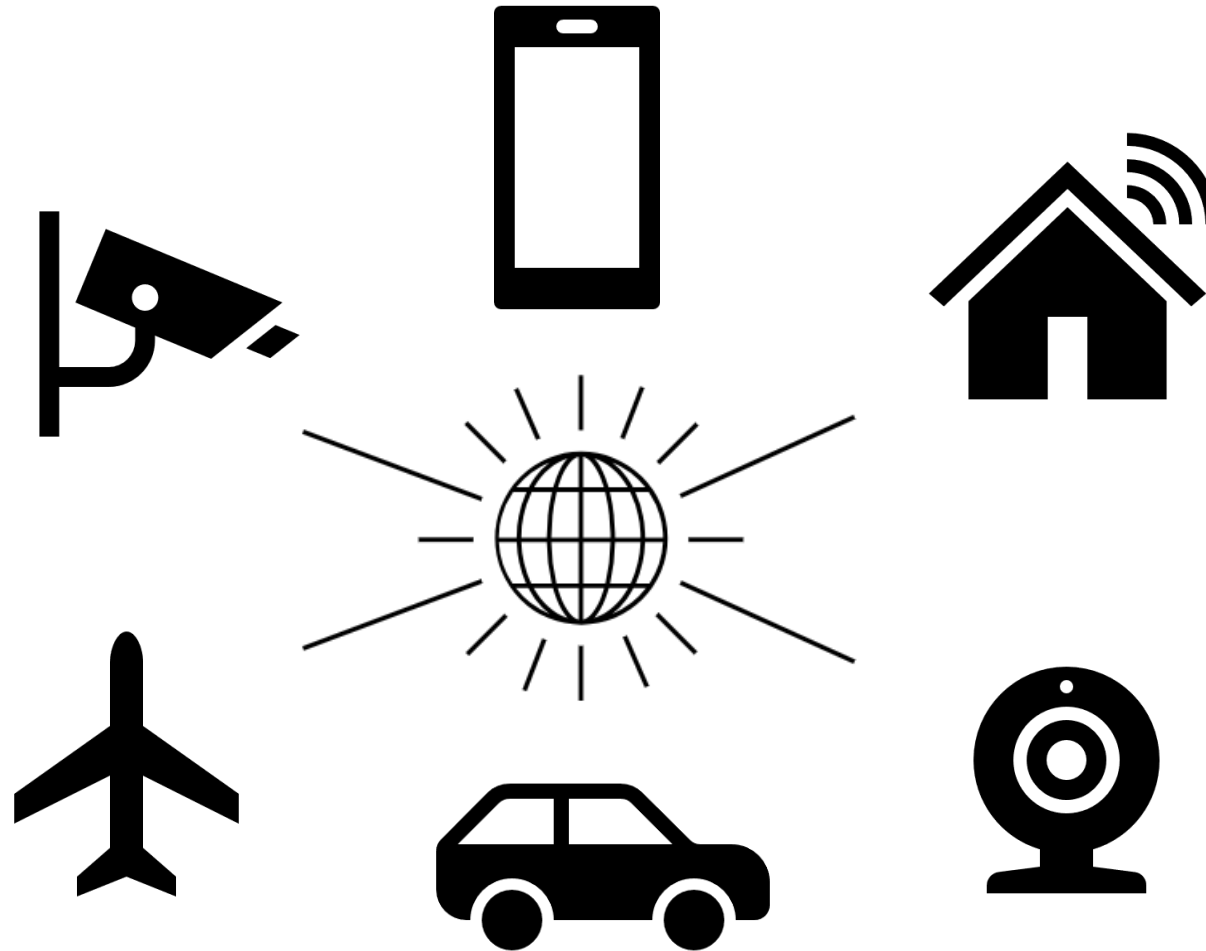
결론 및 향후 연구

01

서론

1. 서론

- ❖ 인터넷 기술의 발전으로 IoT(Internet of Things) 기술이 발전
- ❖ 자동차, 의료, 건축 등 다양한 분야에서 활용되고 있음



1. 서론

❖ IoT 장치의 문제점

- 제한적인 자원
- 장치의 가용성에 영향을 주지 않기 위한 최소한의 보안 솔루션
- IoT 장치가 많을수록 관리가 어려움
- 기존의 장치가 인터넷에 연결되면서 공격 표면(Attack surface)의 증가

❖ IoT 장치가 악의적인 활동에 노출될 경우 발생할 수 있는 피해

- 다양한 분야에서 사용되고 있어 경제적 피해 및 인명 피해가 발생할 수 있음
- 스마트 홈과 스마트 빌딩 등의 거주 공간에 있는 장치는 사생활 침해가 발생할 수 있음



02

배경 지식

2. 배경지식

❖ IoT 장치를 식별하기 위해 특징(Fingerprinting) 정보 생성 방법

■ 수동적 방법(passive)

- 장치와 직접적인 상호작용 없이 특징 정보를 추출
- IoT 장치의 가용성에 영향을 주지 않지만, 정확한 특징 정보를 얻기 어려우며 시간이 많이 소요된다.

TCPDUMP & LIBPCAP



■ 능동적 방법(active)

- 장치와 직접적인 상호작용을 통하여 특징 정보를 추출
- IoT 장치의 가용성에 영향을 끼쳐 장치가 멈추거나 오작동할 수 있음



2. 배경지식

- ❖ IoT 장치 식별을 위해 다양한 논문에서 머신 러닝을 기반으로 IoT 장치 식별을 시도
- ❖ LSTM(Long Short-Term Memory)
 - RNN 모델의 한 종류, 장기적인 시간 간격에 걸친 데이터 처리에 장점이 존재
 - 장기적으로 의존성을 가진 패턴 학습에 강점이 있는 모델
- ❖ CNN
 - 이미지 특징 정보 추출에 장점을 가짐
- ❖ LSTM-CNN
 - LSTM & CNN 결합하여 시퀀스 데이터에서 특징을 추출하고 장기적인 패턴을 학습하여 자연어 처리, 음성 인식, 동영상 분석 등의 분야에서 사용

2. 배경지식

❖ RF(Random Forest)

- 앙상블(Ensemble Learning)의 한 종류, 여러 개의 의사 결정 트리를 생성하고 각각의 트리들의 예측 결과를 조합하여 최종 예측을 수행
- 하이퍼 파라미터 조정으로 과적합 방지와 의사 결정 트리의 특성으로 이상치 및 노이즈 등에 대한 강하다는 장점이 존재

❖ HMM(Hidden Markov Models)

- 관찰 변수 집합에서 숨겨진 변수의 시퀀스를 추정할 수 있는 확률적 그래픽 모델
- 네트워크 트래픽 추적의 반복 가능한 패턴을 감지

❖ Logit Boost

- 앙상블 학습 알고리즘, 분류 오차에 따라 샘플의 가중치 증가
- 올바르게 분류된 샘플의 가중치 감소시켜 모델을 조정하고 예측 성능을 증가시킴

03

머신 러닝 기반 IoT 장치 식별 연구 비교

3. 머신 러닝 기반 IoT 장치 식별 연구 비교

[8] Bai et.al “Automatic Device Classification from Network Traffic Streams of Internet of Things”

- 15개의 IoT 장치(Smart Things, TP-Link Smart plug, Amazon Echo...)를 대상으로 네트워크 트래픽을 19일 동안 수집
- 트래픽 양(traffic volume), 패킷 길이, 네트워크 프로토콜, 트래픽 방향 등의 특징 정보를 사용
- 장치의 종류가 다양해질 경우 정확하게 구별하지 못했다.

[9] Thangavelu et.al “DEFT: A Distributed IoT Fingerprinting Technique”

- 16개의 IoT 장치(Echo dot, Smart remote, Camera...) 를 대상으로 네트워크 트래픽을 7일동안 수집
- DNS, HTTP, TLS 등 10개의 프로토콜에서 특징 정보 추출하여 사용
- 실험에 사용된 16개의 장치외에 다른 장치에서는 식별이 불가능할 수 있음

3. 머신 러닝 기반 IoT 장치 식별 연구 비교

[10] Network Traffic Modeling For IoT-device Re-identification

- IoT Sentinel(31 개) & IEEE TMC 2018(28 개) 데이터 셋 사용
- 네트워크 패킷 헤더, 암호화된 트래픽 등을 특징 정보로 사용

[13] Ensemble machine learning approach for classification of IoT devices in smart home

- 41개의 IoT 장치(Camera, Smart bulb, Smart Things...)를 대상으로 네트워크 트래픽 수집
- IP 주소, 포트 번호, 사용된 프로토콜 등의 특징 정보를 사용
- 앙상블을 이용한 방법으로 리소스가 제한적인 IoT 장치에서 사용이 어려울 수 있음

3. 머신 러닝 기반 IoT 장치 식별 연구 비교

❖ 머신 러닝 기반 장치 식별 관련 연구 비교

동향	Bai et al. [8]	Thangavelu et al. [9]	Najari et al. [10]	Cvitić et al. [13]
데이터 셋 공개 여부	공개	비공개	공개	비공개
IoT 장치 개수	15개	16개	59개	41개
분석 방법	패킷 길이, 네트워크 프로토콜 등	DNS, HTTP, TLS 등의 프로토콜	네트워크 패킷 헤더, 암호화된 트래픽 등	IP 주소, 포트 번호, 네트워크 프로토콜 등
모델	LSTM-CNN 등	Random Forest 등	LSTM 등	Logit Boost 등
수동/능동	수동적 방법(Passive)			
정확도	74.8%	98%	99%	99.7%

04

논의

4. 결론 및 향후 연구

- ❖ 네트워크에 연결된 IoT 장치들을 식별하기 위해 여러 기존 연구에서 머신 러닝을 사용하고 있음
 - IoT 장치 식별을 하기 위해 네트워크 프로토콜, 패킷 헤더 정보 등을 주요 특징 정보로 사용
- ❖ IoT 장치와 다른 장치들을 많이 사용하는 스마트 홈, 스마트 빌딩과 같은 환경에서 머신 러닝 기반 IoT 장치를 식별하는 것은 제약이 있을 수 있음
- ❖ 실시간으로 IoT 장치를 제어하는 경우 가용성에 영향을 줄 수 있음

04

결론 및 향후 연구

4. 결론 및 향후 연구

- ❖ 향후 IoT 장치 식별을 위해 우선적으로 파악해야할 특징 정보와 머신 러닝 모델에 대한 정보 및 한계점을 제공
- ❖ 다양한 분야에서 IoT 장치의 사용이 증가하고 있으며, 장치의 종류 또한 다양해지고 있다.
- ❖ 자원이 제한적인 IoT 장치에 대한 네트워크 보안 및 개인 정보 보호 문제에 대한 연구가 필요

Q&A
