EE209AS Course Project: House Occupancy Detection

Midterm Progress Report

Cong Jin(704205565)
ZhuoQi Li(504135743)

Abstract

In this project our objective is to infer the rough indoor activity of a room/house(if there is someone at home or not) by snipping and parsing the WiFi packets that are sent by IoT devices such as Google home or Amazon Echo, and by mobile devices. We furthur demonstrated that even by parsing the information contained in 802.11 frame header, we can create WiFi signature of different device and infer the activity within the house or home without decrypting WiFi packet payload.

Introduction

As the concept of Internet of things are prevailing these days, more and more family choose to use them in house as a center of control for other in-house IoT devices. For example, Google home can be associated with smart door lock or smart lamp for remote voice control. While these devices greatly facilitate our daily life, it also exposes security issues during device-to-device or device-to-cloud communication via network.

To be more specific, devices such as Google Home, or Amazon Alexa will need to be connected to the network in order to fully enforce it's functionality. Even though nowadays the WiFi packets are encrypted and are almost impossible to be decrypted, we still find a way to characterize different devices presented in the room by parsing the 802.11 frame headers which are not encrypted and can be analyzed automatically by tools such as whireshark. We demonstrate that by just using the information in the 802.11 frame header, we can form the WiFi signature of different device. After we gathering enough information and forming the signature of each device, we can compare each new snipped WiFi packet generated by unknown device and identify possible known device by generating signatures from the new packet and compare it with existing signatures. Lastly, by how devices are detected within certain period, we can infer the room occupancy.

Proposed method

After reading through several literatures on network traffic analysis, we propose two ways to reach our object: the first approach includes the use of classification algorithm such as Support vector machine(SVM) and K means clustering; the second method focus on extracting the contend of traffic ourselves instead of feed them into classifier, where we will selectively parse the traffic packet and form WiFi signature for each of detected device and perform occupancy detection based on the presence of those devices.

Current Progress

Week1-2: Confirm Project Object **------ Done**

Week 3: Read/Learn Related Resources/Literatures **----- Done**

Week 4: Kali Linux and WiFi packets collection environment/tools setup **----- Done**

Week 5: Traffic data collection**,** data pre-processing, and K-means clustering on data. -----**Done**

Week 6: WiFi data parsing and WiFi signature formation ----- **In progress**

Result & Discussion

So far we have successfully have our hardware setup and software and it's kali linux environment setup for traffic packet collection, to be more specific, we are able to collect traffic packets generated by google Home, Amazon echo and a mobile phone and are also able to identify the source of different packet by specific 802.11 header information we obtain from the packet. At last, we are also able to perform k-means clustering on the data we collected after filtering the data by MAC address and vectoring data. However, the result of clustering is not good enough for indoor presence detection as the input is still noisy and feature extraction is needed for a better data pre-processing which might generate a better result for classification.

After Midterm Plan

Week 7: WiFi data parsing and WiFi signature formation

Week 8-9: Indoor Occupancy detection using WiFi signature.

Week10: Tuning & Optimization

Related Work

In the paper “Is Anybody Home? Inferring Activity From Smart Home Network Traffic”, the researchers are trying to infer personal information by investigating device-to-device and device-to-cloud smart home network traffic. They are using traffic analysis techniques on network traffic generated by NEST thermostats to deduce information about the presence of residents and other events occurring within the property. They first collect traffic data and perform data filtering by MAC address and Organization Unique Identifier(OUI), after they have collected all the data they begin to learn the pattern of the data flow of each device by traffic classification and focusing on the connection size and correlation analysis as criterial for classification.

In the paper “Spying on the Smart Home: Privacy Attacks and Defenses on Encrypted IoT traffic, the researchers demonstrate that an ISP or other network observer can infer privacy sensitivity in-home activities even when the device use encryption. They further prove that even with encryption, smart home traffic attack is achievable with meta-data only. Their attack includes the use of DNS queries or device fingerprinting to identify smart home devices from network traffic and infer user activities from changes in the device traffic rate.

In the paper “Deep packet, A Novel Approach For Encrypted Traffic Classification Using Deep Learning”, the group propose a deep learning based approach for traffic classification that integrates both feature extraction and classification steps, their network is able to classify FTP and P2P network classes and can distinguish between VPN and non-VPN network thanks to the nature and extend of CNN. Also in the paper “Network Traffic Classification using Support Vector Machine and Artificial Neural Network” the group is exploring other methods in machine learning domain to perform an effective network traffic classification.

At last, to implement our design, we are following a tutorial called “Everything generates data: Capturing WiFi anonymous traffic using Raspberry Pi and WSO2 BAM” to collect WiFi packet in Kali Linux environment.