Detecting Malicious IoT Devices

Alex Korzyniowski & Jason Vettese

Department of Computer Science, University of New Hampshire

University of New Hampshire

Summary

- ❑ Wanted to investigate the malicious behavior of common IoT devices
- ❑ Chose to test Smart Assistants like Amazon Echo since they are very common in households
- ❑ Created a methodology for analyzing IoT traffic with machine learning
- ❑ Packet captures on an isolated segregated network
- ❑ Machine learning to flag malicious traffic patterns
- ❑ Implemented Regression Decision Tree using the Gini index
- ❑ Investigated the security of popular IoT devices

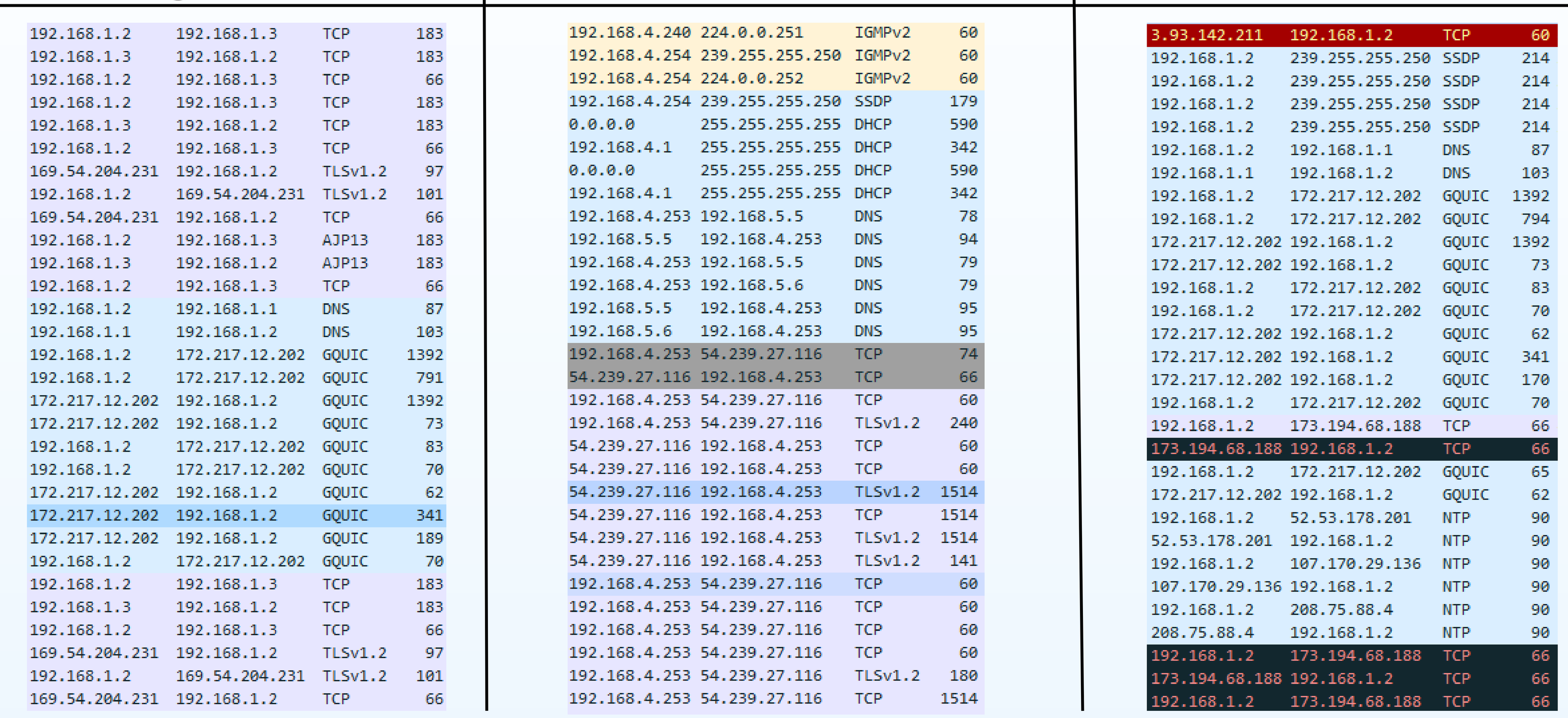
Background

- ❑ IoT devices have grown quickly in popularity
- ❑ IoT devices have become a common attack vector
- ❑ There are few alternatives available for assessing IoT network traffic
- ❑ Compromised IoT devices put users at risk of stolen personal information and inadvertent involvement in cybercrimes are not properly secured
- ❑ The purpose of this project is to provide users with a method to flag potentially malicious traffic patterns.

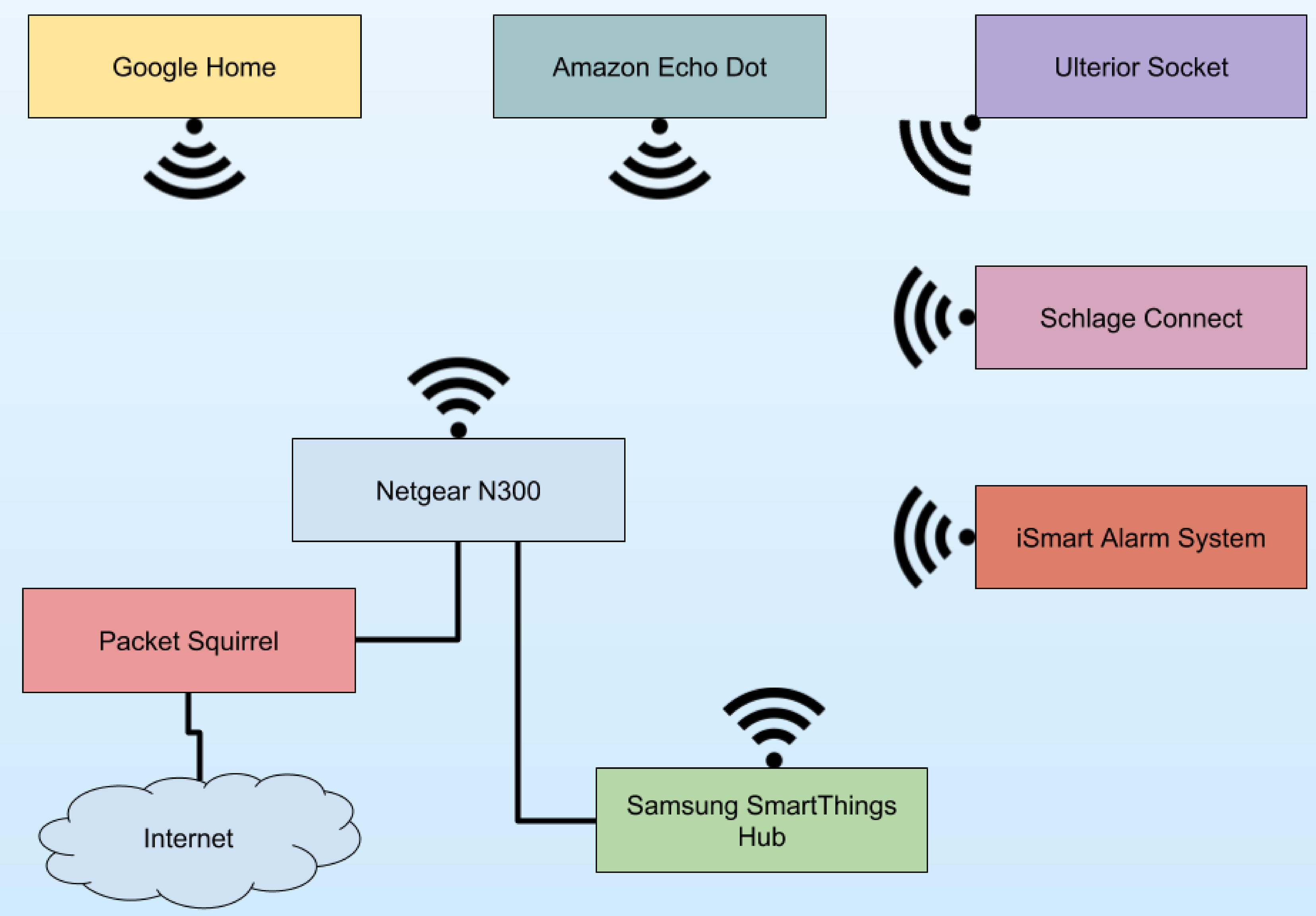
Analysis

- ❑ Created Network to capture data
- ❑ Network comprised of Wireless Router and Zigbee Hub
- ❑ Captured both Wireless and Ethernet traffic
  - ❑ Wireless traffic was found to be not helpful
- ❑ Packet Squirrel was used to intercept data going inbound and outbound
- ❑ Machine Learning model was trained using a continuous dataset
- ❑ NetFlow data is primarily categorical data
  - ❑ Source IP, Destination IP, Protocol, and Packet Size
  - ❑ No deep packet inspection to reduce computational complexity
- ❑ Translation of IPs into geographic coordinates

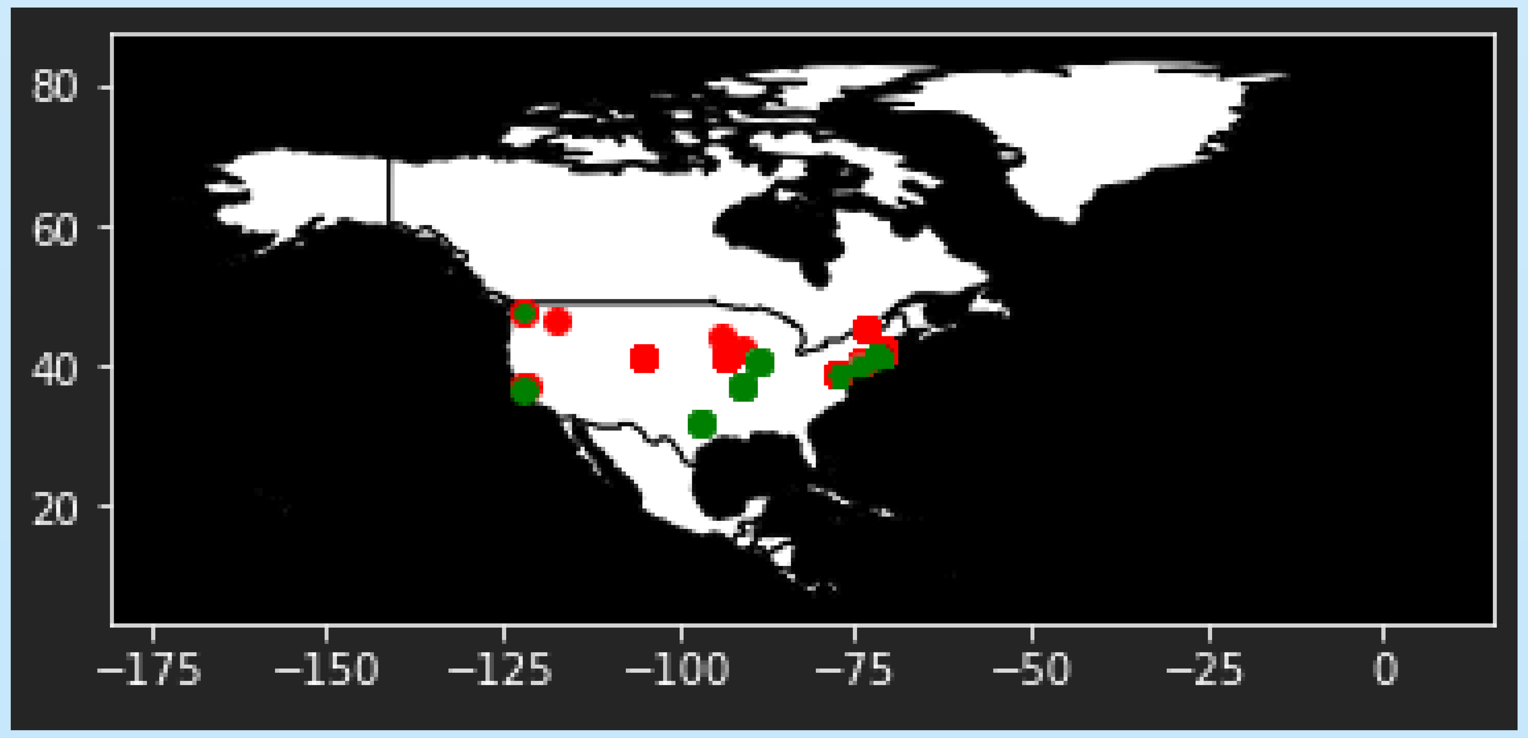
Google Home | Amazon Alexa | Schlage Smart Lock



Network Capture Model

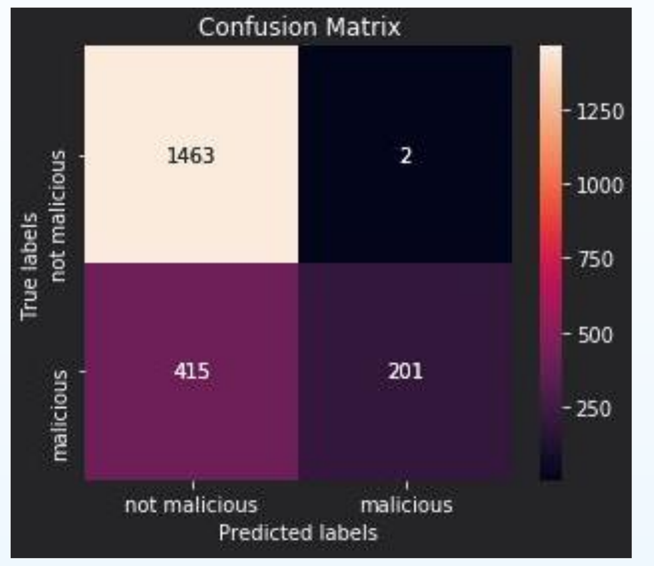


Google Home Geographic Data



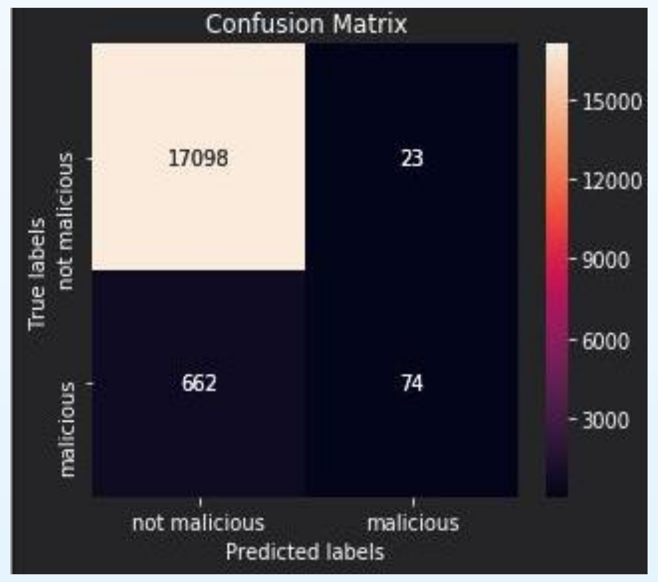
Results

Google Home:



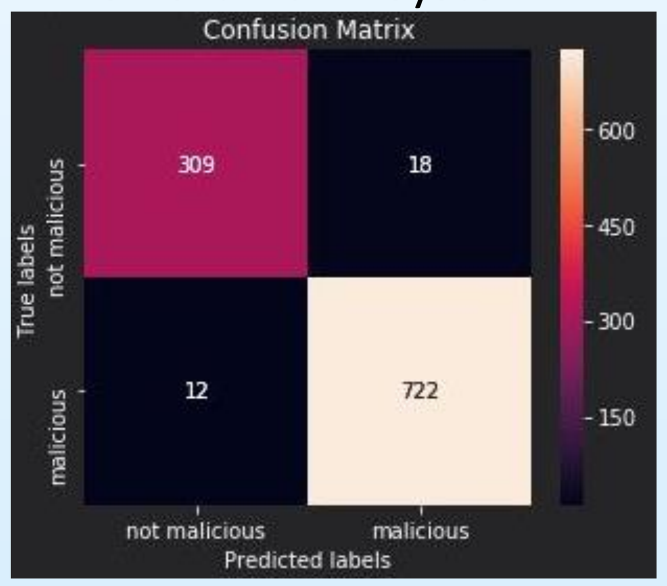
Google Home Model Accuracy: **74.75%**

Amazon Echo:



Amazon Echo Model Accuracy: **96.16%**

iSmart Alarm System:



iSmart Alarm Model Accuracy: **97.17%**

Discussion

The results that were attained from the model seem to make sense. It is interesting that the model was not able to predict the Google Home as accurately as the Echo or iSmart Alarm. This result is most likely due to the way the model is developed. The model gets more accurate with the more data that is given to it. In the case of the Echo, it had good accuracy since it had a large dataset.

Future Work

- ❑ "Your results are only as good as your data"
- ❑ Expanding malicious dataset to detect additional forms of traffic
- ❑ More Effective Man in the Middle detection
- ❑ Refine model parameters
- ❑ Live processing model