Predicting IoT Malware Attacks

Prabhakar Rangarao

Proliferation of IoT Devices Pose Cyber Security Threats

64

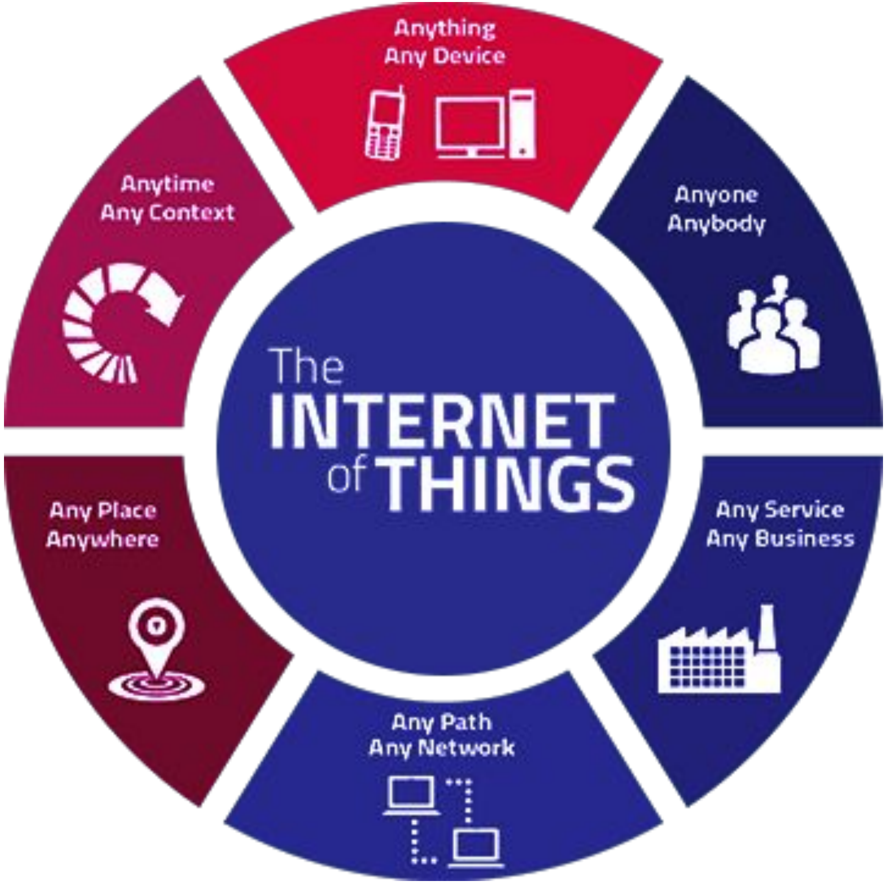
Billion

IoT

Devices

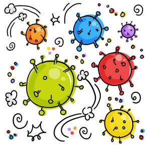
(2025)

(IDC)



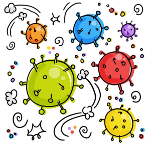
https://dreamztechusa.com/

Leaked Source Codes Makes It Easy For Hackers



MIRAI

- First attack in 2016
- Targeted at IoT devices
- Largest DDos attack (1.2TB/sec)
- Many variations still continue
- Source code in **Github**

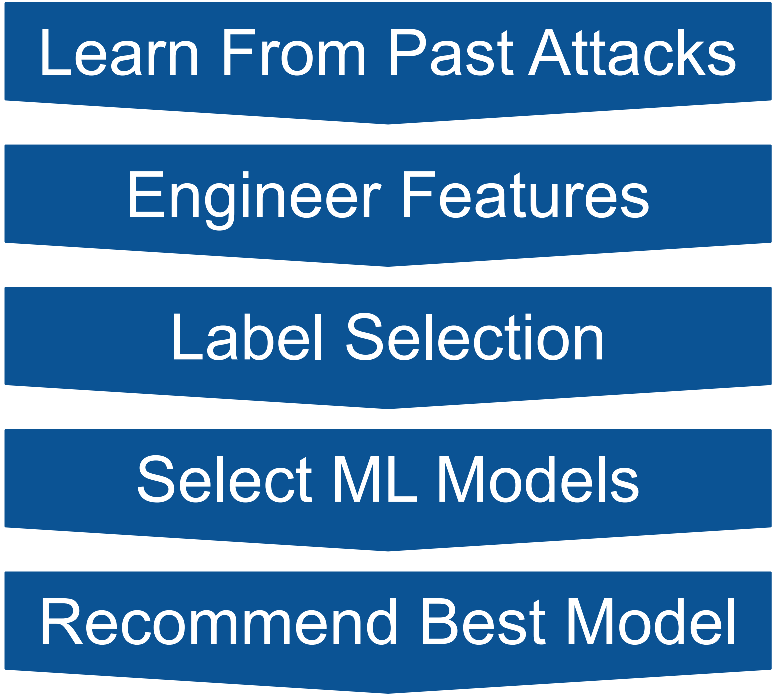


BASHLITE

- First attack in 2014
- 2016 attack mostly on IoT devices
- Variations through l**eaked source code**

Source: DDoS-Capable IoT Malwares: Comparative Analysis and Mirai Investigation (https://www.hindawi.com/journals/scn/2018/7178164/)

Machine Learning As Mitigation Strategy



Learn From Past Attacks

Engineer Features

Label Selection

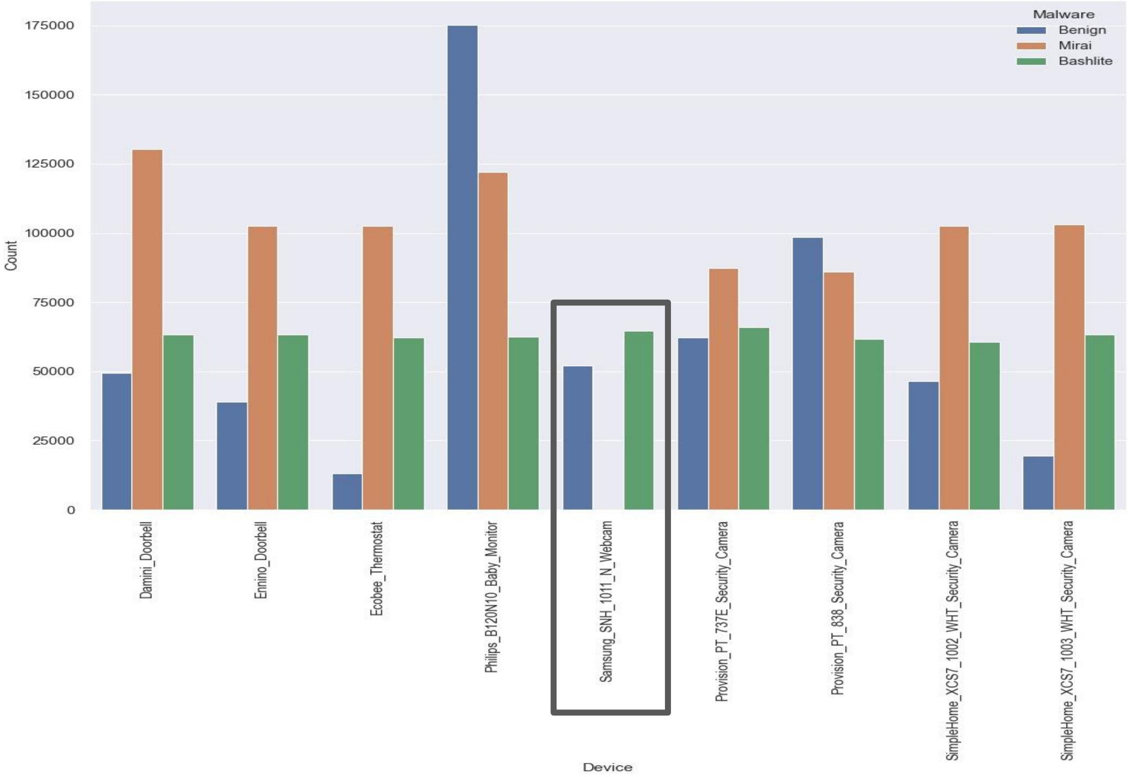
Select ML Models

Recommend Best Model

- Devices infected by Mirai and Bashlite

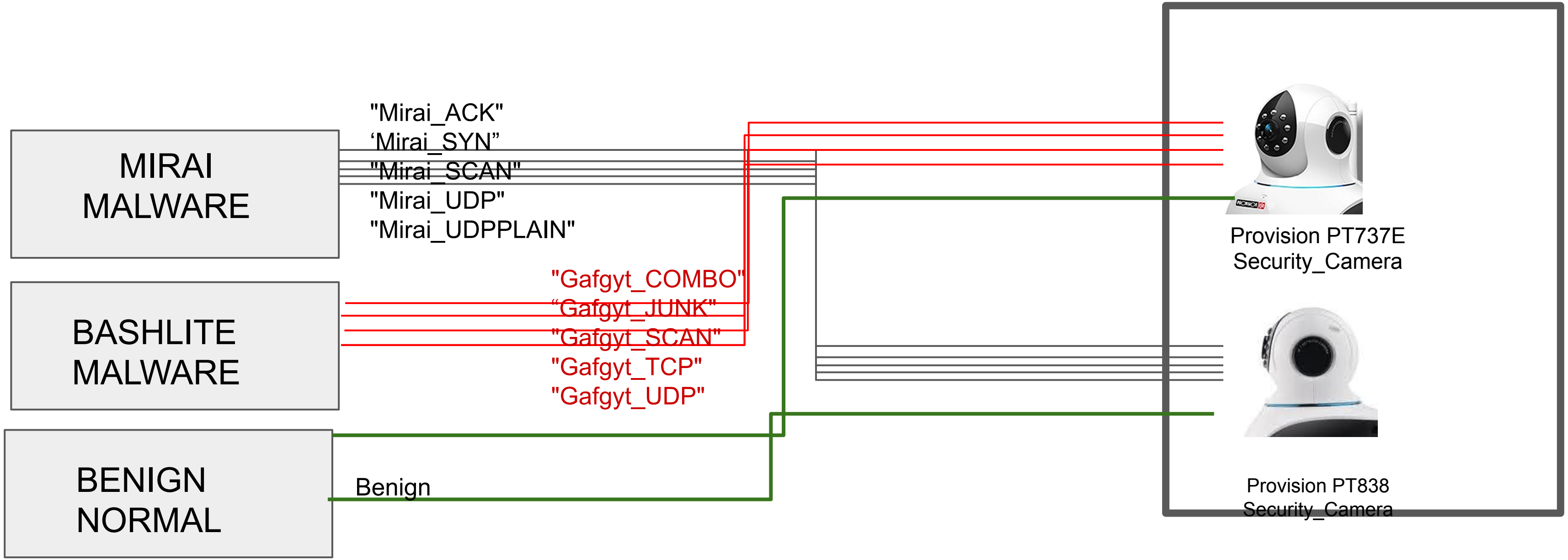
  (Source: UCI Machine Learning Repository)
- Aggregate data and feature selection
- Benign, Mirai, Bashlite(gafgyt)
- Supervised Learning - Classification Models
- Use Recall and Accuracy

Large Datasets with Class Imbalance



7.5GB
Data

DATA UNDERSTANDING AND TRAFFIC TYPES



MIRAI MALWARE

"Mirai_ACK"
‘Mirai_SYN”
"Mirai_SCAN"
"Mirai_UDP"
"Mirai_UDPPLAIN"

BASHLITE MALWARE

"Gafgyt_COMBO"
“Gafgyt_JUNK"
"Gafgyt_SCAN"
"Gafgyt_TCP"
"Gafgyt_UDP"

BENIGN NORMAL

Benign

Provision PT737E
Security_Camera

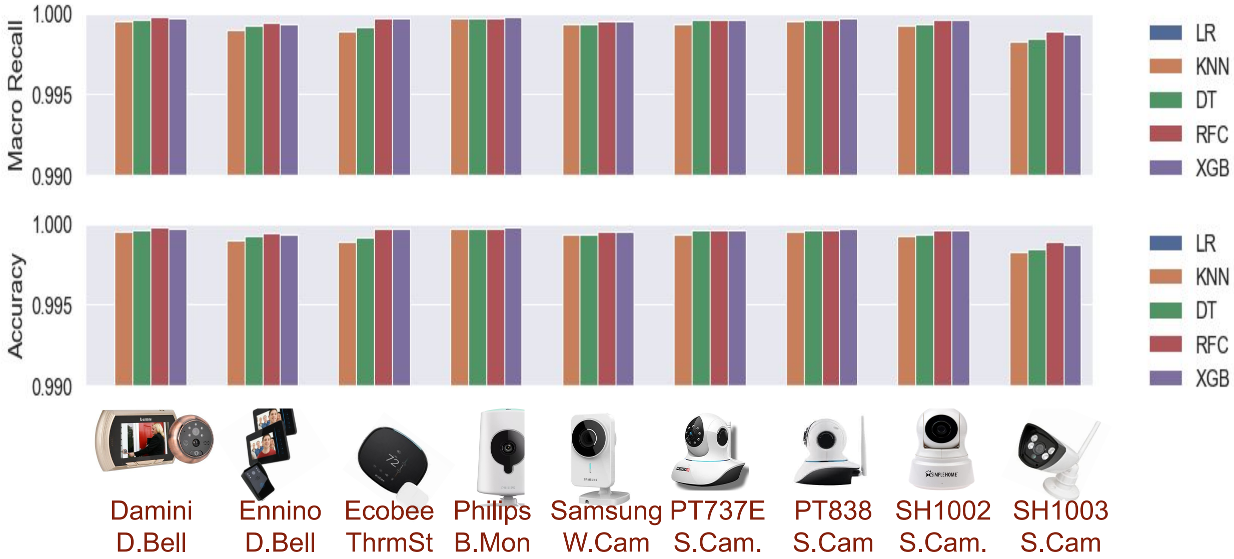
Provision PT838
Security_Camera

7.5M
OBSERVATIONS

115
FEATURES

9
IOT DEVICES

Model Score Comparison

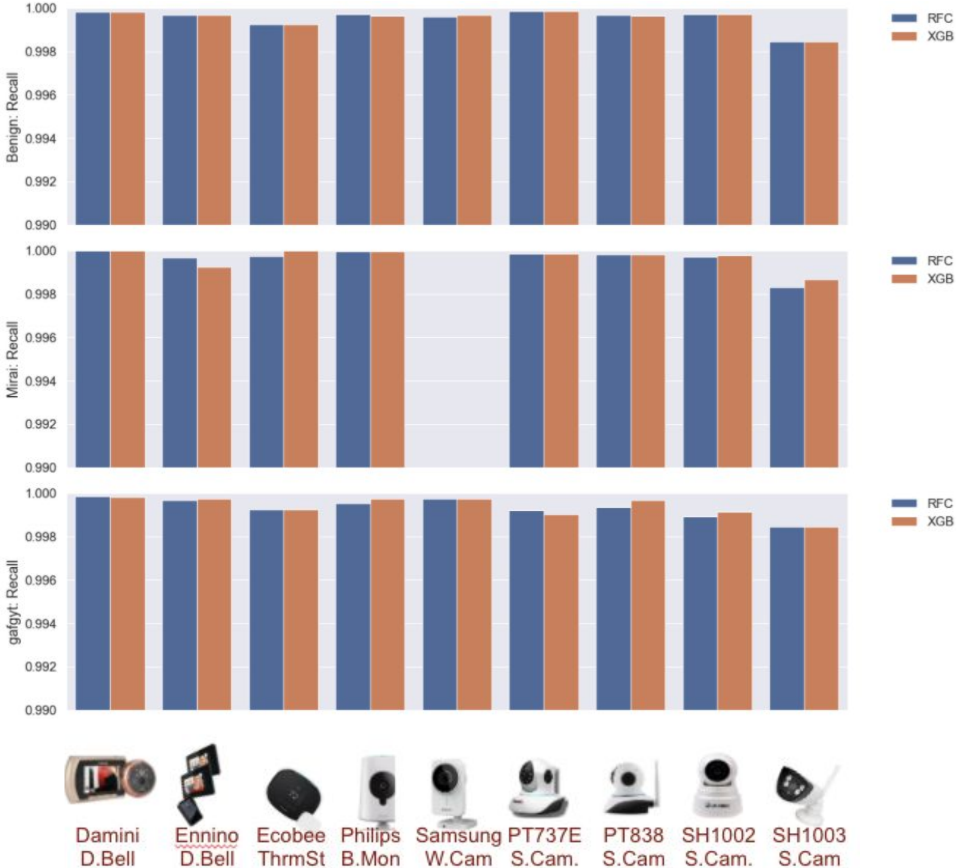


Compared Recall & Accuracy scores from 5 models for the 9 IoT devices using Training, Validation and Test dataset separately.

Random Forest & XGBoost Models

Predicted Malware Attacks Better Than Other Models

Model Comparison: Random Forest VS XGBoost



	Accuracy	Macro-Recall
RFC	0.9994937068921056	0.9994938158669717
XGB	0.9995244975900507	0.999524605556056

HIGHER

Final Model Selection After Hyperparameter Tuning

RANDOM FOREST

XGBOOST

0.99956

Random Forest

0.99953

XGBoost

✓ FASTER

XGBoost Model is the most promising
decision-tree based ensemble model for anomaly detections

Future Research Opportunities ..

- Which of the 10 attacks carried by 2 botnets are more vulnerable?
  - Expand number of classification to 11
- How well the model perform if exposed to 7.5GB data once.
  - Run models in the cloud (AWS) with the entire ~7.5GB of data
- Expose the model to new data infected by different malware attacks.
  - Expand to include additional data sources from more latest malware attacks
- Deployment of the model in live network assets(router, app)

Contact Information

Prabhakar Rangarao is a Data Scientist with background in Statistics, Computer Science and Marketing Strategy.

- Prabhakar.Rangarao@gmail.com
- https://www.linkedin.com/in/prabhakarrangarao/
- https://github.com/ghPRao/
- https://medium.com/p/53cb208cdf0/edit