Confirm
Smart Manufacturing

Confirm
Smart Manufacturing

Edge2Guard: Botnet Attacks Detecting Offline Models for Resource-Constrained IoT Devices

Bharath Sudharsan, Dineshkumar Sundaram, Pankesh Patel, John G. Breslin, Muhammad Intizar Ali

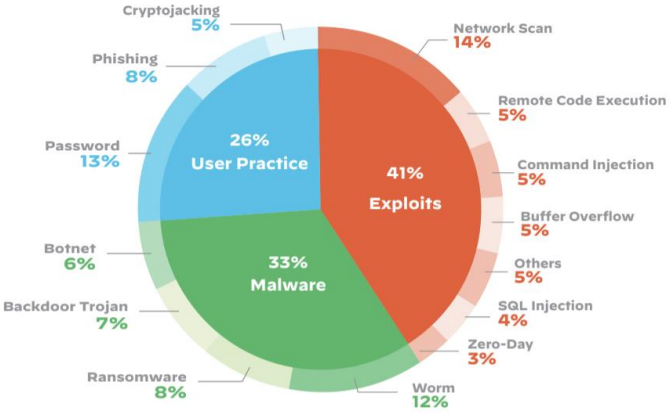
A World Leading SFI Research Centre

Science Foundation Ireland sfi For what's next

University of Limerick Ollscoil Luimnigh · Tyndall National Institute · UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh · NUI Maynooth Ollscoil na hÉireann Má Nuad · Oé Gaillimh NUI Galway · CIT Cork Institute of Technology Institiúid Teicneolaíochta Chorcaí · AIT · LIT

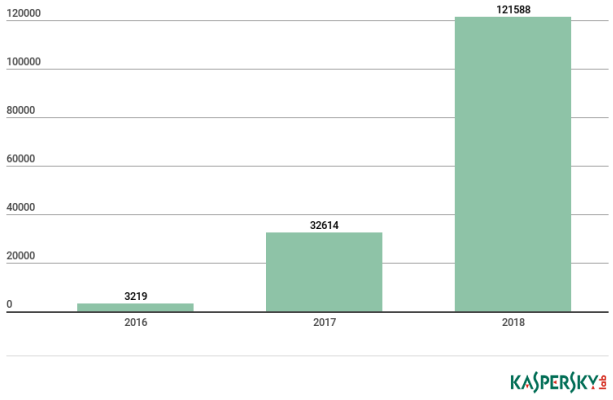
Introduction

Confirm
Smart Manufacturing



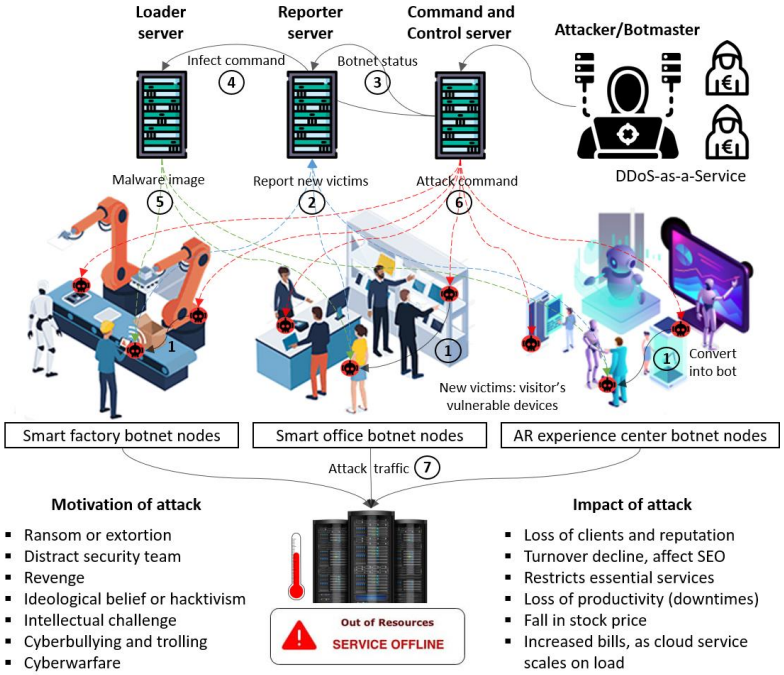
- Over 50 billion devices will be connected to the Internet by 2025, half of which may be vulnerable to multiple cyberattacks
- According to a Symantec report, it takes only 2 minutes to attack an IoT device

- In a Kaspersky Lab threat report, they were able to collect 121,588 malware samples from IoT devices in 2018, ≈4 times more than in 2017



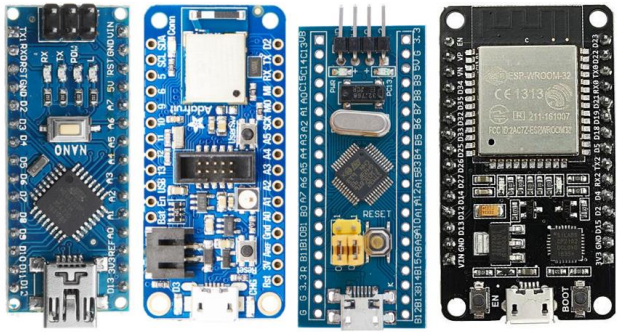
Overview of malware communication

Confirm
Smart Manufacturing



- ✓ Bots starts exploring to find, convert vulnerable devices (1)
- ✓ The new victim's details are sent to Reporter server (2)
- ✓ Investigate the features of the newly acquired devices (3)
- ✓ Botmaster sends infect command (4)
- ✓ Depending on the victim, Hardware binary is flashed (5)
- ✓ Bots listen to instructions from the C&C (6)
- ✓ When commanded, target servers are attacked (7)

Cause of problems



MCU1	MCU2	MCU3	MCU4
ATmega328P	**nRF52840**	**STM32f103c8**	**ESP32**
8 kB SRAM	256 kB SRAM	20 kB SRAM	520 kB SRAM
32 kB Flash	1 MB Flash	128 kB Flash	4 MB Flash
@ 16 MHz	@ 64 MHz	@ 72 MHz	@ 240 MHz

- Despite high threats, the security conditions of IoT devices remain unsatisfactory because

  - ✓ **Feasibility**: to produce lower-power-cost devices MCUs are used as its brains. Challenging to implement attack protection mechanisms on such low-resource hardware

  - ✓ **Cost**: Devices are built with cost as the driving design tenet, they have poor configurations and open design

  - ✓ **Boost sales**: To save memory, simultaneously provide attractive functionalities, manufacturers adopt simplified lightweight versions of protocols in their devices, making them susceptible to various attacks

Problem definition

Confirm
Smart Manufacturing

- To safeguard devices, we cannot expect network-based attack detection mechanisms on all external networks
  - ✓ For example, our smartwatch may connect to dubious free Wi-Fi in public places like shopping malls, when we arrive at the airport, etc.
  - ✓ Then our smartwatch gets attacked by bots or malicious devices in insecure networks
- *Hence, there is a pressing need for a defense mechanism that can execute on memory and power-constrained MCU-based IoT devices, without impairing their lifespan or jeopardizing their functionality*

Edge2Guard overview

Confirm
Smart Manufacturing

- We provideEdge2Guard (E2G) models to alleviate cyber-security issues faced by tiny MCU-based IoT devices
  - ✓ **Real-time detection**: E2G continuously monitor network traffic data to detect malware attacks in real-time
  - ✓ **High accuracy**: We achieved almost 100% accurate malware detection rates using E2G
  - ✓ **Standalone, offline characteristics**: Unlike others, E2G detect attacks without depending on networks (standalone) or any external protection mechanisms (offline)
  - ✓ **Resource-friendly design**: E2G can run on a wide range of IoT devices without imposing computational pressure and without disturbing device routine

DDoS attack defense techniques

Confirm
Smart Manufacturing

- We outline santdard techniques that detect IoT attacks before they occur

  - ✓ **IP traceback:** perform packet filtration closer to the attack source. For e.g., hash-based IP traceback approach to defend against botnet DDoS and reflection-based attacks

  - ✓ **Entropy variations**: to detect slow request or response attacks, users can find the difference in the entropy

  - ✓ **Intrusion detection and prevention systems (IDS/IPS):** IDS can be deployed on any layer, such as cloud for gathering alerts from edge sensors, then correlate and analyze the alerts. IPS can prevent intrusions via packet drops

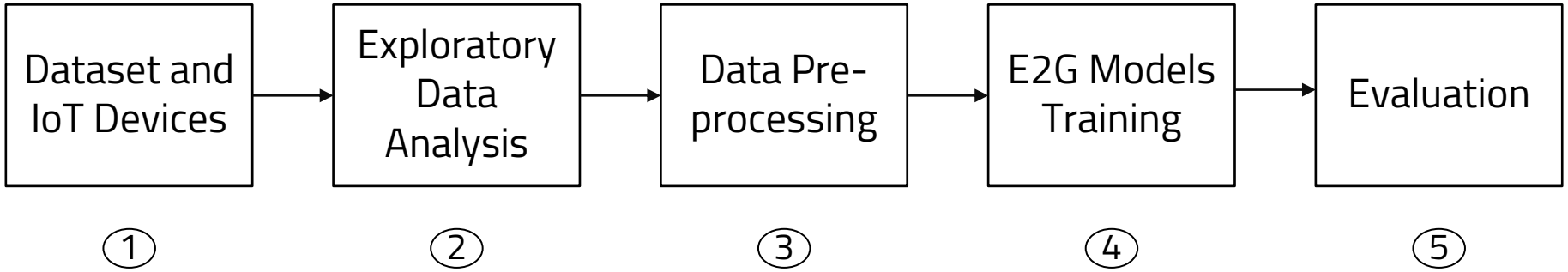
ML techniques to detect bots

Confirm
Smart Manufacturing

- A variety of ML algorithms are available to detect attacks in IoT environment. For examples
  - ✓ OC-SVM detection mechanism for application-layer DDoS attacks
  - ✓ Honeypots detect botnet DDoS attacks by capturing device malware installation attempts
  - ✓ ANN-based method accurately discovered several application-layer DDoS attacks
  - ✓ Autoencoder to detect anomalous network traffic from compromised IoT devices
  - ✓ CNN-based DDoS attack detection and warning system
- Such models are deployed on networking devices. From surveys, many on-device methods were uncited, and offline methods for resource-constrained IoT devices were not mentioned at all

Edge2Guard design

Confirm
Smart Manufacturing



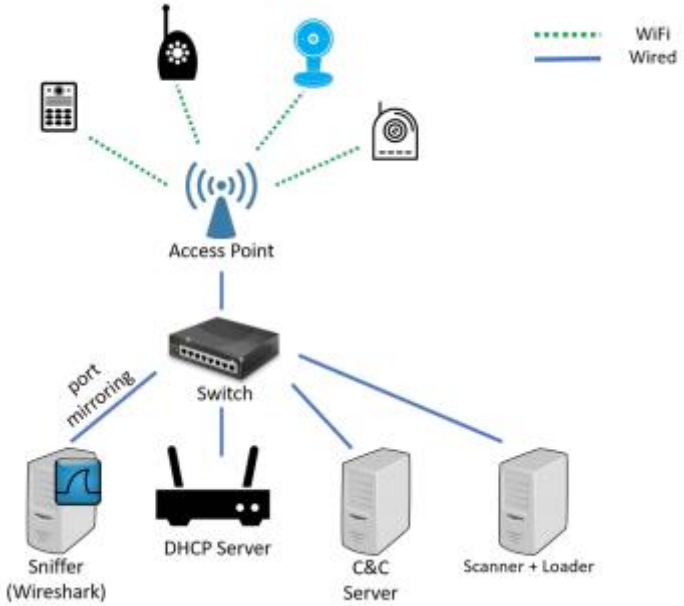
Edge2Guard repository

Confirm
Smart Manufacturing

- We have open-sourced the implementation and E2G models so can be readily used to reproduce our results. Link: https://github.com/bharathsudharsan/Edge2Guard

  - ✓ **Dataset_wrangling.ipynb:** Data dimension, feature info, data profile of each malware, etc.
  - ✓ **Exploratory_data_analysis.ipynb:** Reduce 115 features to 2 and make 2D, 3D scatter plots
  - ✓ **Data_preprocessing_and_E2G_model_training.ipynb:** We pre-process, train multiple models, and evaluate using Accuracy, F1 score, Kappa, and MCC
  - ✓ **Benign/Gafgyt/Mirai_data_profile.html:** Generate profile reports. Contains data *Statistics, Histograms, Common values*, and *Extreme values* to describe data
  - ✓ **E2G_model_training_and_evaluation_results.docx:** Detailed evaluation results of all the types of E2G attack detecting models

Dataset and IoT devices

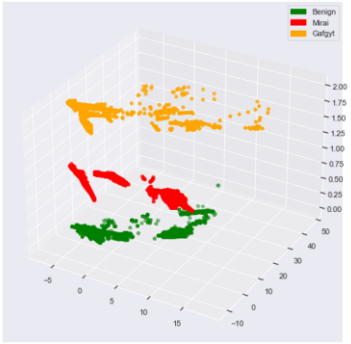


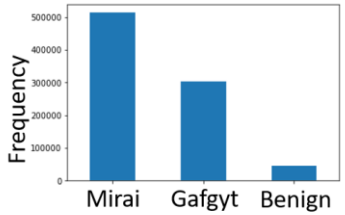
- Dataset: N-BaIoT
  - ✓ Data patterns for normal and attack traffic
  - ✓ Attack traffic collected by infecting devices with authentic botnets from Mirai and Bashlite family
- 9 IoT Devices:
  - ✓ Doorbells: Danmini, Ennio
  - ✓ Thermostats: Ecobee
  - ✓ Baby Monitor: Philips
  - ✓ Web cams: Samsung. Security Cams: Provision, Simple Home

Exploratory Data Analysis (EDA)

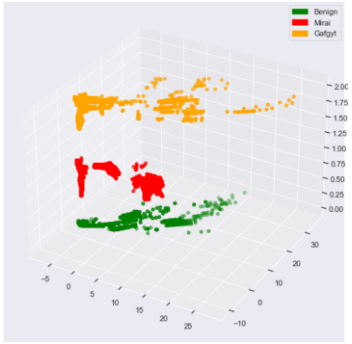
Confirm
Smart Manufacturing

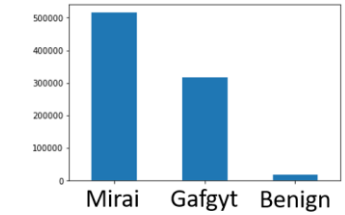
SimpleHome_XCS7_1002
WHT_Security_Camera





SimpleHome_XCS7_1003
WHT_Security_Camera





- Using PCA we reduce 115 features into 2 features
- We visualize features to explore patterns, find trends between the malicious and benign traffic data
  - ✓ Provision PT-737E and PT-838 have similar traffic patterns
  - ✓ Simple Home 1002 and 1003 have similar patterns
- From this, we can infer that both Mirai and Bashlite malware behave the same for devices from the same brand

E2G models

Confirm
Smart Manufacturing

COMPARING PERFORMANCE OF VARIOUS ATTACK DETECTION E2G MODELS FOR DANMINI DOORBELL.

Model	Acc	Recall	Prec	F1	Kappa	MCC
Random Forest	1.0000	0.9999	1.0000	1.0000	0.9999	0.9999
Decision Tree	0.9998	0.9997	0.9998	0.9998	0.9997	0.9997
K Neighbors	0.9980	0.9935	0.9980	0.9980	0.9960	0.9960
Ridge Regr	0.9969	0.9958	0.9969	0.9969	0.9936	0.9936
iForest	0.9700	0.96	1.00	0.98	0.6546	0.6977
OC-SVM	0.9300	0.93	1.00	0.96	0.0453	0.1058
Ada Boost	0.9245	0.9202	0.9340	0.9216	0.8392	0.8522
QDA	0.6834	0.8271	0.8491	0.6724	0.4799	0.5712
Naive Bayes	0.6585	0.3543	0.7312	0.5410	0.0693	0.1829
Linear SVM	0.4204	0.3930	0.4682	0.3959	0.0762	0.1060
LOF	0.1400	0.85	0.09	0.17	0.0182	0.0912
Logistic Regr	0.0486	0.3333	0.0024	0.0045	0.0000	0.0000

- E2G Supervised Learning Models
  - ✓ We did not use Deep Autoencoders, CNNs, ANNs since we target low resource hardware
- E2G One-Class Classification Models
  - ✓ Not feasible to track hundreds of new malware forms and keep updating E2G models
  - ✓ So, we trained OCC models only using benign data

Evaluation

Confirm
Smart Manufacturing

F1 SCORE OF THE TOP-PERFORMING RF AND DT E2G MODELS WHEN TESTING USING IOT DEVICE DATA.

Device	All data with 3 classes		Under sampled data with 3 classes		All data with 11 classes		Under sampled data with 11 classes	
E2G Model	RF	DT	RF	DT	RF	DT	RF	DT
Danmini Doorbell	1.0	1.0	1.0	1.0	1.0	0.86	1.0	0.57
Ecobee Thermostat	1.0	0.99	1.0	1.0	0.99	0.92	0.98	0.77
Ennio Doorbell	1.0	0.99	1.0	1.0	0.99	0.94	0.98	0.98
Philips B120N10 Baby Monitor	1.0	1.0	1.0	1.0	0.99	0.85	0.98	0.87
Provision 737E Cam	1.0	1.0	1.0	1.0	0.99	0.78	0.98	0.85
Provision 838 Cam	1.0	1.0	1.0	1.0	1.0	0.79	1.0	0.87
Samsung SNH 1011 N Webcam	1.0	1.0	1.0	1.0	0.99	0.89	0.99	0.99
Simple Home XCS7 1002 Cam	1.0	0.99	1.0	1.0	1.0	0.91	1.0	0.64
Simple Home XCS7 1003 Cam	1.0	0.99	1.0	1.0	0.99	0.91	0.97	0.84

- Even if a single instance of attack traffic is misclassified as benign, the device will get compromised
- The Random Forest (RF) and Decision Tree (DT) models show top performance
- The RF and DT E2G models accurately classified all 10 types of attacks

Results comparison

Confirm
Smart Manufacturing

- We compare the performance of E2G with papers that cite and use the N-BaIoT dataset
  - ✓ Our RF and DT E2G models outperformed the top models by showing close to 100% detection rates
  - ✓ Offline attack detection capabilities of E2G protect devices even when connected to dubious networks by mistake

Conclusion

Confirm
Smart Manufacturing

- We presented E2G models that can
  - ✓ Comfortably execute within the limited resource of tiny IoT devices
  - ✓ Classify malware attack traffic in real time and offline
  - ✓ Show highest detection performance in comparison to existing approaches

Confirm
Smart Manufacturing

Confirm
Smart Manufacturing

Contact: Bharath Sudharsan
Email: bharath.sudharsan@insight-centre.org

Contact: Ali Intizar
Email: ali.intizar@dcu.ie

www.confirm.ie

Science Foundation Ireland
sfi
For what's next