

# Botnet Detection on IoT Devices

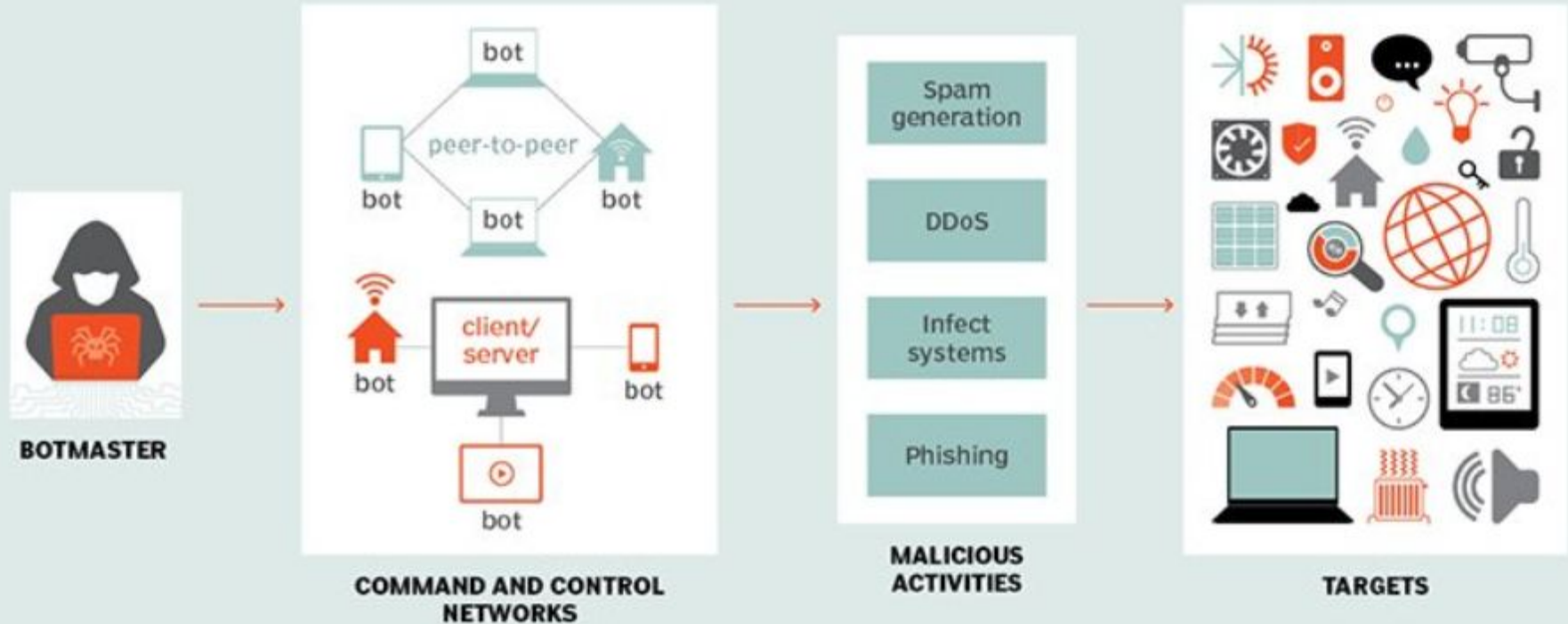
Dineshkumar Sundaram  
Data Science Capstone project - Springboard

# The Problem

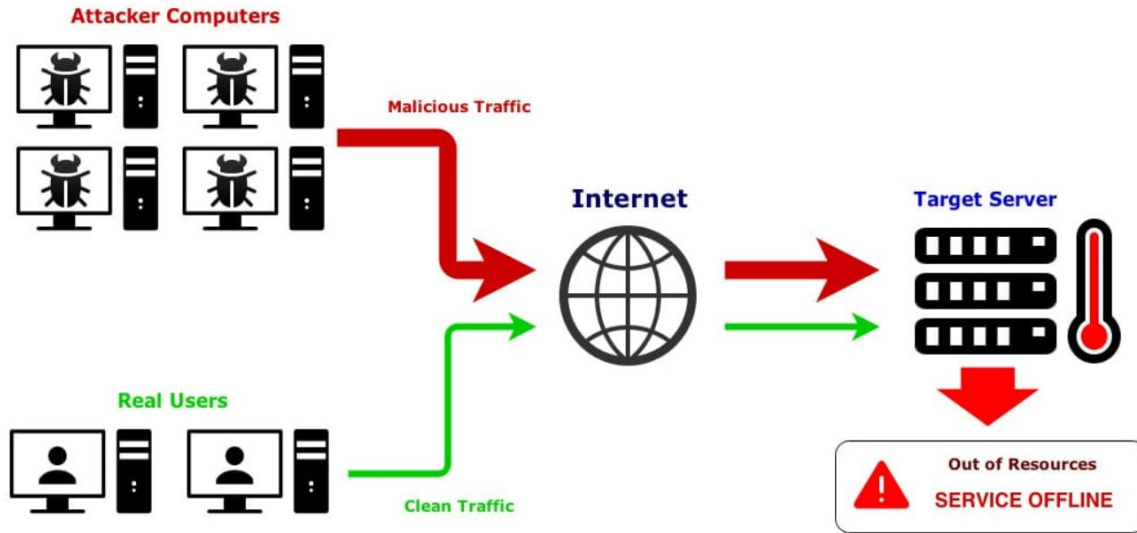
- Past **2 years DDoS attacks** has risen by **20%** & the scale and severity of their impact have risen by nearly **200%**.
- Sharp rise in protocol DDoS attacks.
- Increasing number of IoT devices are increasing the risk of DDoS attacks.
- 5G will fuel botnet-driven DDOS attacks in upcoming years.

1. <https://cybersecurityventures.com/the-15-top-ddos-statistics-you-should-know-in-2020/>
2. <https://www.indusface.com/blog/ddos-attack-trends/>

# What is Botnet?



# What is DDoS?



## Distributed Denial-of-service

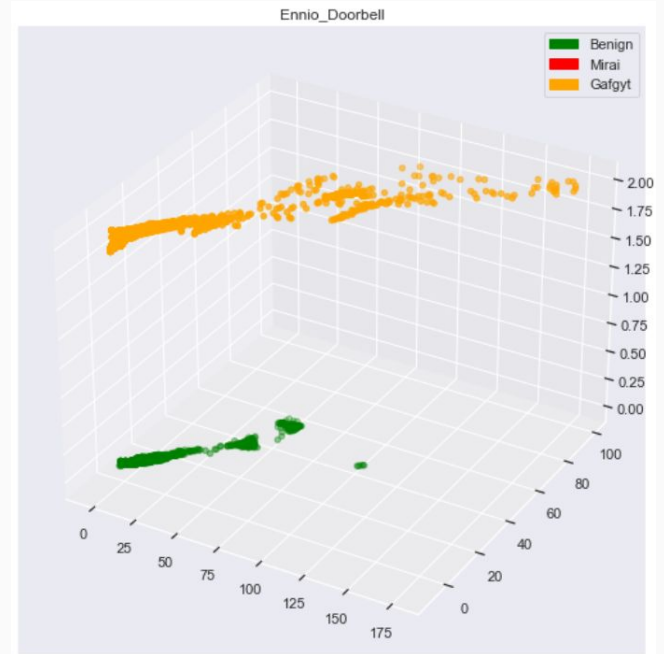
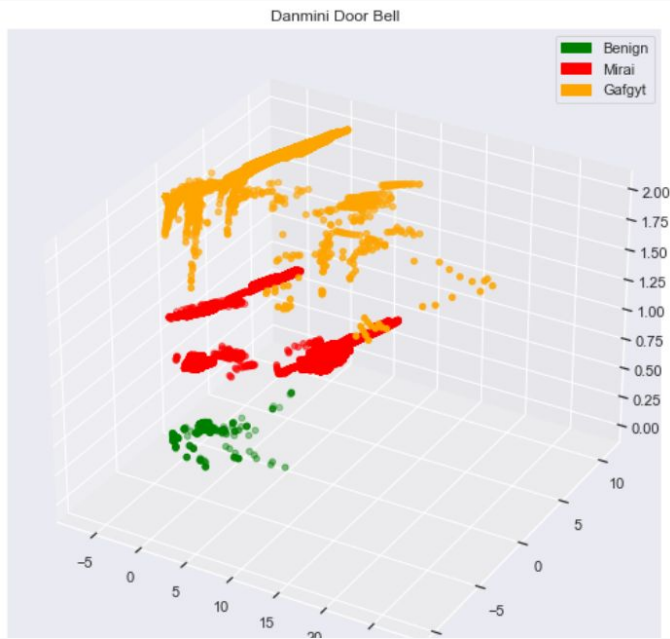
1. <https://www.cloudflare.com/en-gb/learning/ddos/what-is-a-ddos-attack/>

# Who might care?

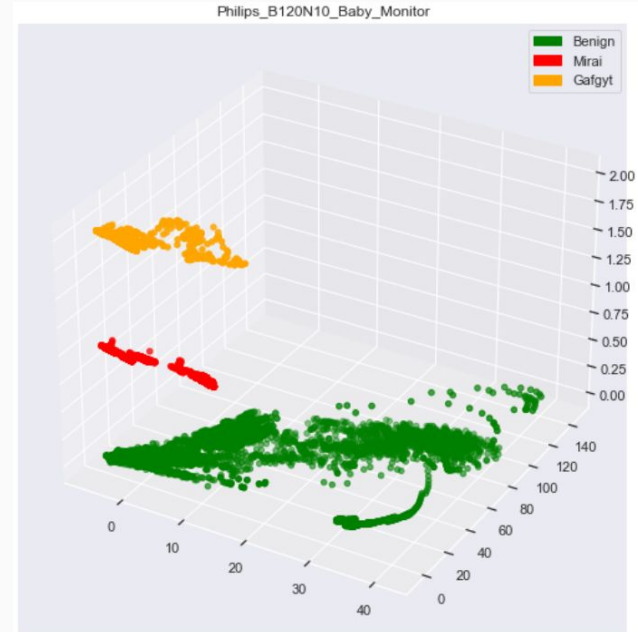
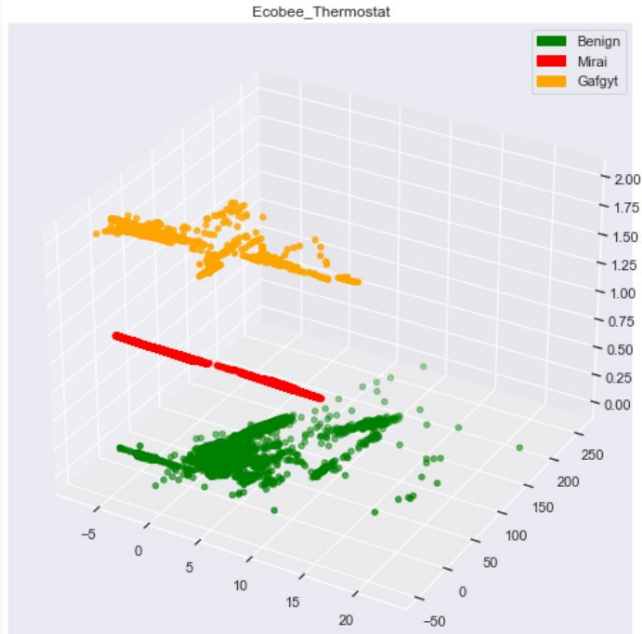
- Cyber security firms.
- Large financial / corporate enterprises.
- IoT Device Manufacturers.
- Anyone who uses internet !

- 2 type of Malware - Mirai , Bashlite
- 5 Categories and 9 IoT Devices - Baby monitor, Webcam, Security Camera, Doorbell, Thermostat
- Set of 23 features with 100ms, 500ms, 1.5sec, 10sec, 1 min time interval.
- Summary statistics - network snapshot of the device.

# Data - Doorbell

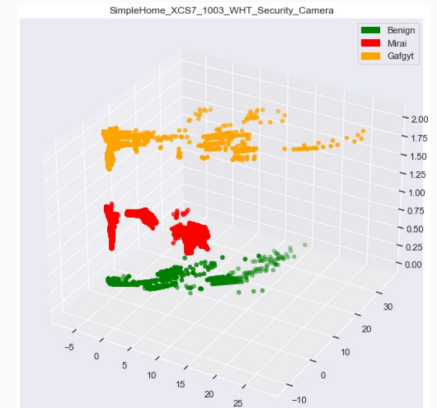
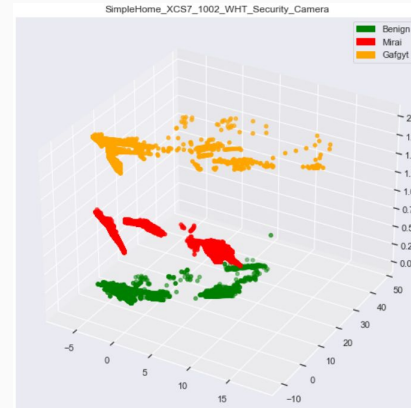
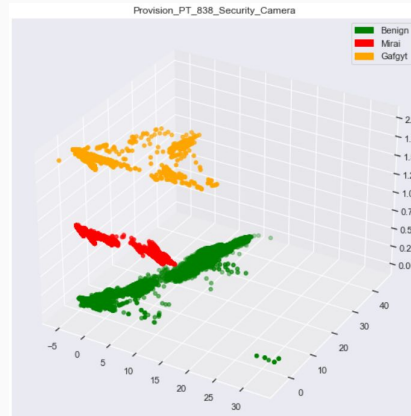
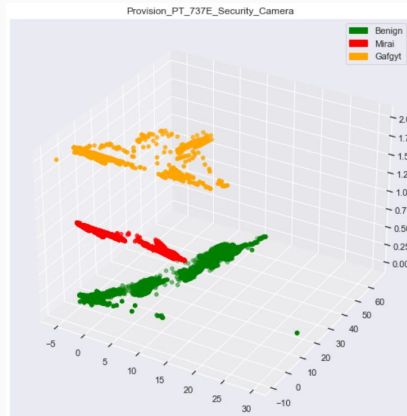


# Data - Baby monitor & Thermostat





# Data - Security cam



- Supervised Learning
- 3 class & 11 class
- Highly imbalanced data
- Scikit learn and imblearn

## Model comparisons - Danmini Door bell

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
0	Random Forest Classifier	1.0000	0.0	0.9999	1.0000	1.0000	0.9999	0.9999	2.8086
1	Decision Tree Classifier	0.9998	0.0	0.9997	0.9998	0.9998	0.9997	0.9997	21.2064
2	K Neighbors Classifier	0.9980	0.0	0.9935	0.9980	0.9980	0.9960	0.9960	25.6996
3	Ridge Classifier	0.9969	0.0	0.9958	0.9969	0.9969	0.9936	0.9936	1.2116
4	Ada Boost Classifier	0.9245	0.0	0.9202	0.9340	0.9216	0.8392	0.8522	144.0179
5	Quadratic Discriminant Analysis	0.6834	0.0	0.8271	0.8491	0.6724	0.4799	0.5712	5.3659
6	Naive Bayes	0.6585	0.0	0.3543	0.7312	0.5410	0.0693	0.1829	0.8091
7	SVM - Linear Kernel	0.4204	0.0	0.3930	0.4682	0.3959	0.0762	0.1060	6.0382
8	Logistic Regression	0.0486	0.0	0.3333	0.0024	0.0045	0.0000	0.0000	4.2906

## Model performance

# Assumptions, Limitations

- Model - Individual devices
- New Device - Train model again
- Model training - only network traffic data
- Deployment - Need optimization
- Data - Only current version of malware

## Improve the model in future

- Develop Generic model.
- Trained with both network traffic and device action data.
- Convert model into device firmware and deploy into edge device.
- Malware constantly evolve, need to update the model when new vulnerability found on the internet

# Conclusions

- Random forest model performed well compare to other supervised learning model.
- All 115 features used to train the model since malware can attack the device on different time interval
- 70%-30% Splitting the test data gave F1 score of 1.0
- Constant monitoring of the malware, the model can be improved in the future.

# Thank you!

Dineshkumar Sundaram

<https://github.com/dineshh912>