A Data-Centric Anomaly-Based Detection System for Interactive Machine Learning Setups

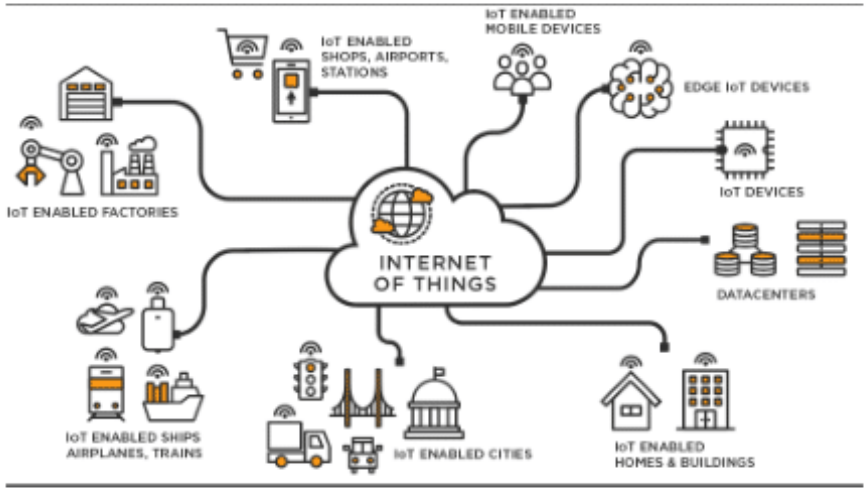
Joseph Bugeja[1] and Jan A. Persson[1]

[1]Internet of Things and People Research Center, Department of Computer Science and Media Technology, Malmö University, Malmö, Sweden

October 27, 2022

The Internet of Things (IoT)

- The IoT has transformed environments, e.g., homes and buildings, connecting them to the Internet
- 29.4 billion connected devices in 2030 (Statista, 2022)
- Global IoT market expected to cross USD 1 trillion landmark by 2024 (Research and Markets, 2022)



Security Challenges of the IoT

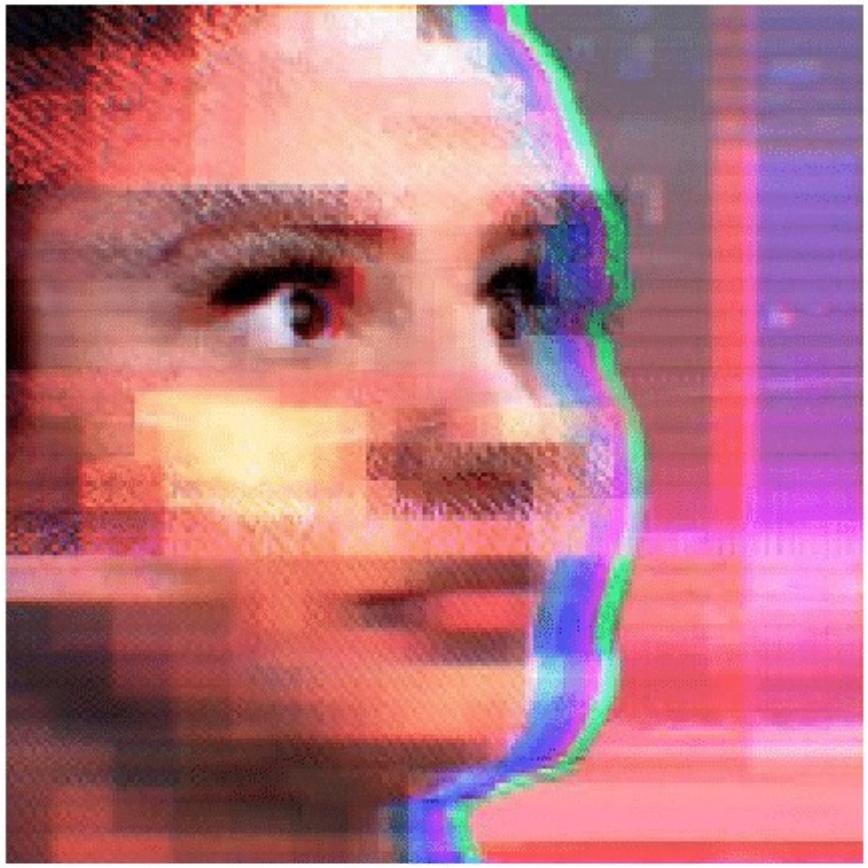
- IoT devices expose organisations to new security threats and a broad attack surface
- Particularly, when machine learning (ML) is used
- It is even more the case when interactive ML is used

Interactive ML Setups

- In interactive ML online learning is done by regular users
- This tends to improve the overall accuracy of the underlying ML models
- Several use-cases, e.g., activity recognition

Poisoning Attacks

- A poisoning (integrity) attack is an attack where adversaries inject fake training data with the aim of corrupting the learned model
- Industry is more concerned about poisoning threats than other adversarial ML threats
- A label-flipping attack exploits classification algorithms by corrupting their training data (e.g., silent $\rightarrow$ convo/meeting)



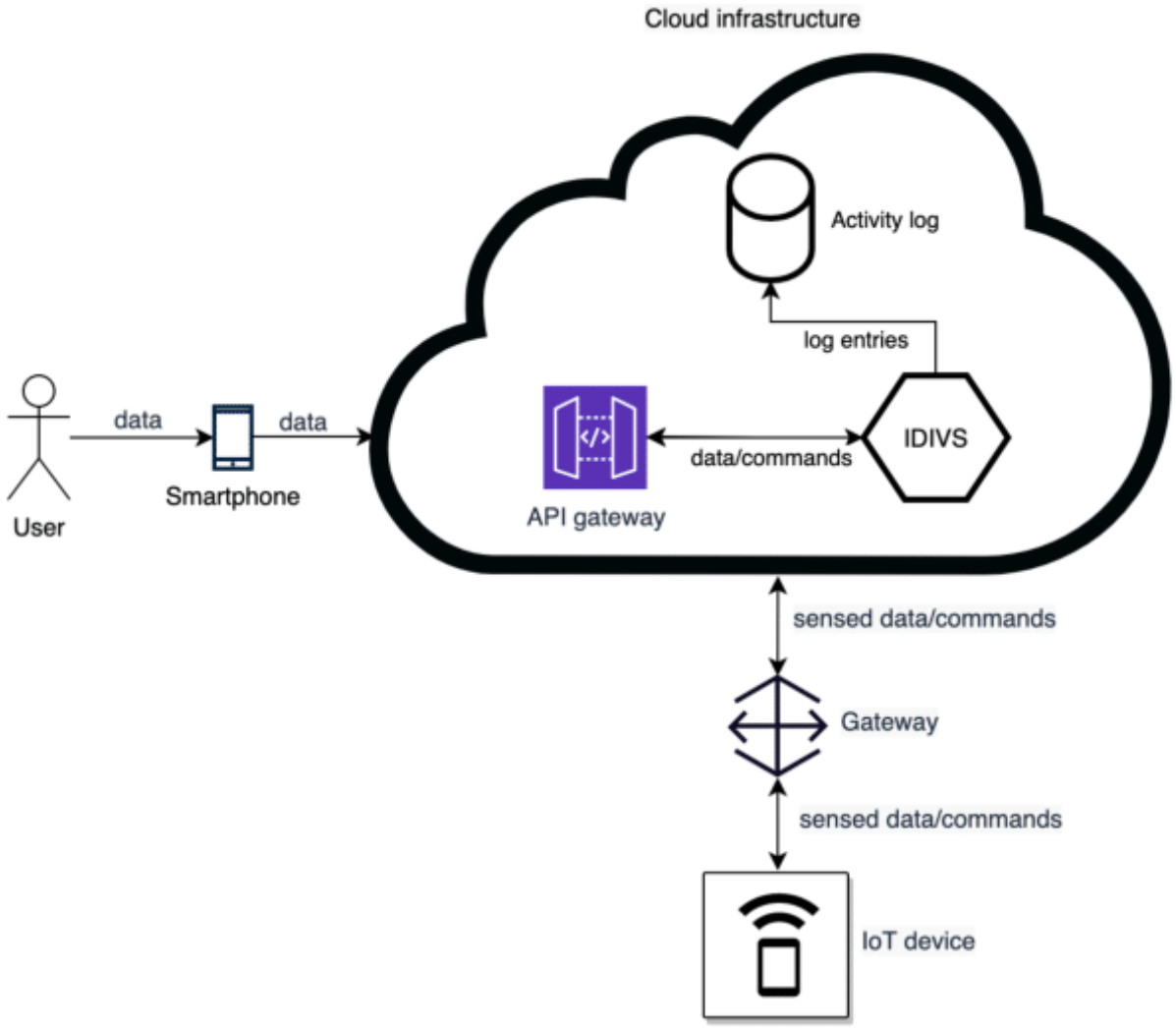
Twitter profile picture of Microsoft's Tay

Related Work and Going Beyond It

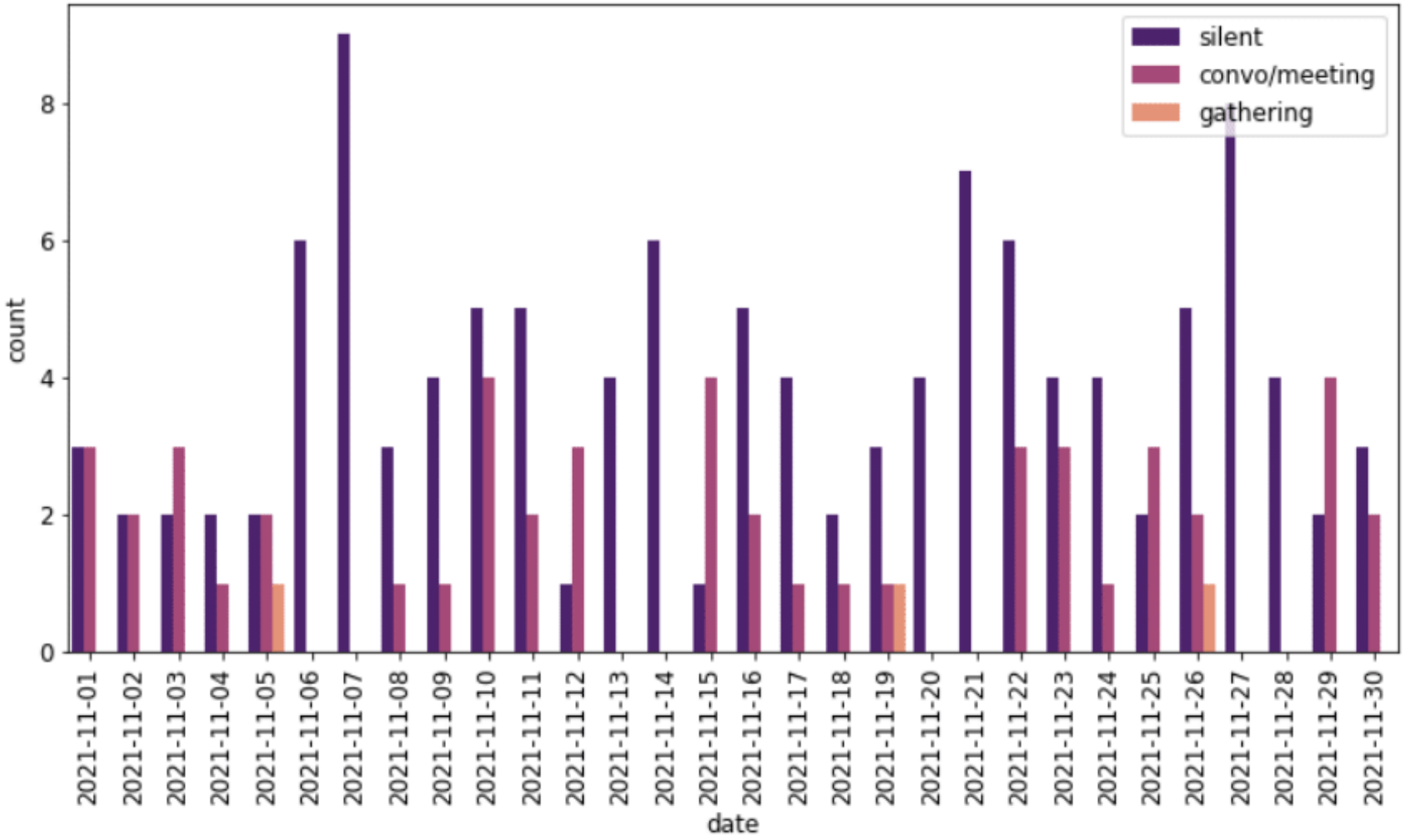
- Leverage the data-centric approach
- Apply anomaly detection on the application layer
- Create our own dataset using actual lab data and synthetic data
- Leverage multiple supervised ML algorithms to find the most accurate approach

Experimental Setup

- Based on a concept known as Dynamic Intelligent Virtual Sensor (DIVS)
- Smart campus setup: smart camera, climate sensmitter, smart lighting, and a smartphone
- Interactive service users furnish feedback via a user feedback process through their smartphone



Type of Activities Occurring in November



Feature Description

A summary of all the dataset features:

Feature	**Description**
timestamp	Date and time the activity occurred
principal	Entity performing the activity
device	Source IoT device used to perform the activity
activity	Type of activity performed
message	Indicates whether the activity is a command or data
attribute	Physical or virtual feature of the environment or system
value	Content of the attribute
anomaly	Indicates whether the activity represents an anomaly

Processed Dataset

Excerpt of data collected during the month of November:

timestamp	principal	device	activity	message	attribute	value	anomaly
1635768000	System	Smart Camera	Building Automation	Data	Presence	1	0
1635771600	User2	Phone	User Feedback	Data	State	convo/meeting	0
1635775200	User1	Phone	User Feedback	Data	State	convo/meeting	0
1635778800	System	Climate Sensmitter	Building Automation	Data	Temperature	37.48483678922438	0
1635782400	System	Smart Camera	Building Automation	Data	Presence	0	0
1635786000	System	Climate Sensmitter	Building Automation	Data	Temperature	39.209483628261744	0
1635789600	User1	Phone	User Feedback	Data	State	convo/meeting	0
1635793200	User2	Phone	User Feedback	Data	State	silent	0
1635796800	User2	Phone	User Feedback	Data	State	silent	0
1635800400	System	Climate Sensmitter	Building Automation	Data	Temperature	36.547992971237925	0
1635804000	System	Smart Camera	Building Automation	Data	Presence	1	0
1635807600	User1	Phone	User Feedback	Data	Count	0	0

Model Development

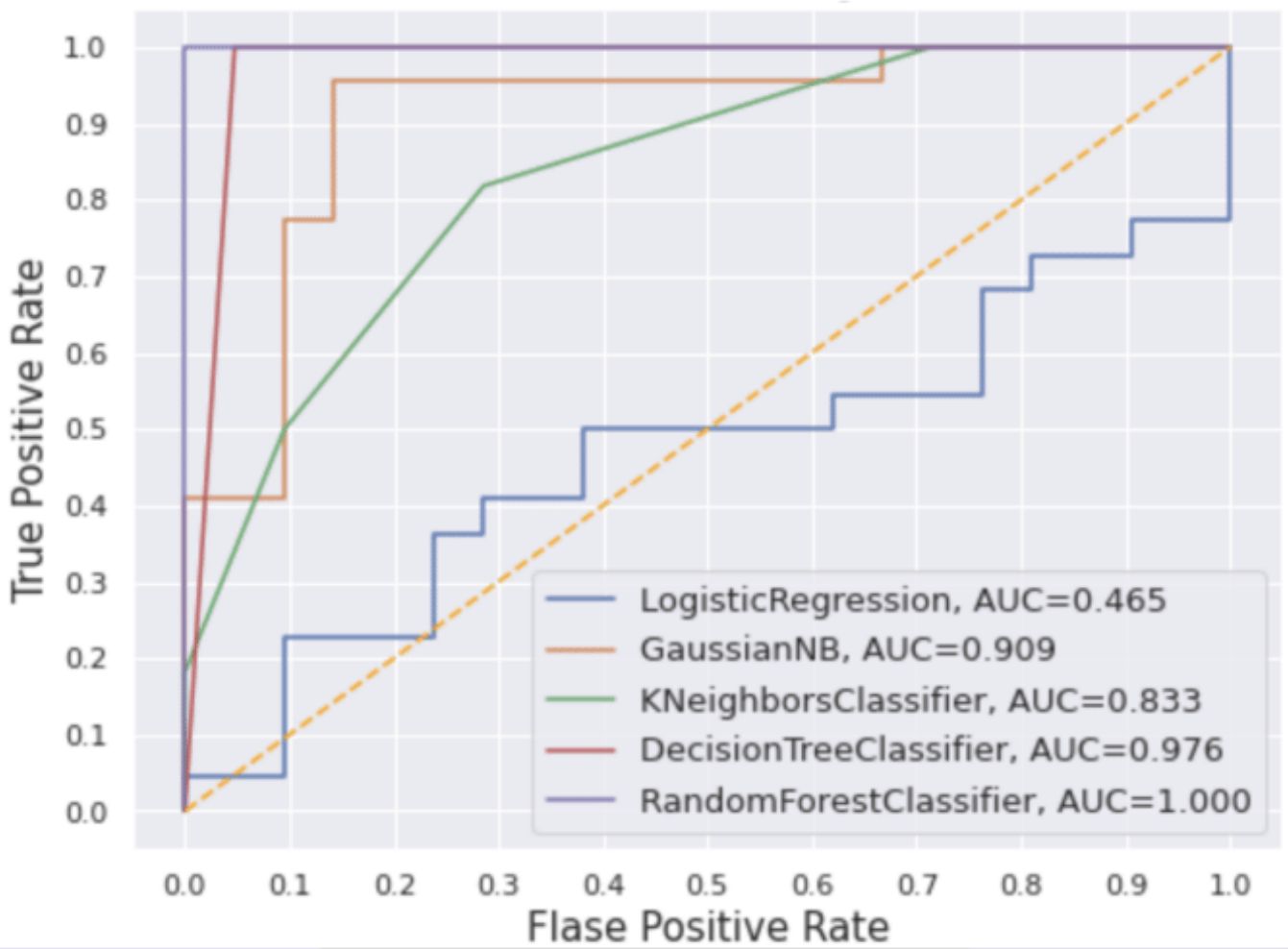
- As part of the training process, a random sample of activities that correspond to the user feedback process was selected
- Then, to simulate a label-flipping attack, those records were poisoned
- In total, 35% of the training dataset was poisoned

Model Development

- Development: Logistic Regression, Gaussian Naive Bayes, K-Nearest Neighbors, Decision Tree, and Random Forest
- Validation: AUC-ROC, accuracy, and F1-Score

Results

The Random Forest was the algorithm with the highest accuracy (0.98), F1-Score (0.98), and AUC-ROC (1),



Some Limitations

- Retraining similar to batch learning
- Potential classifier overfitting
- Dataset size (n=600)

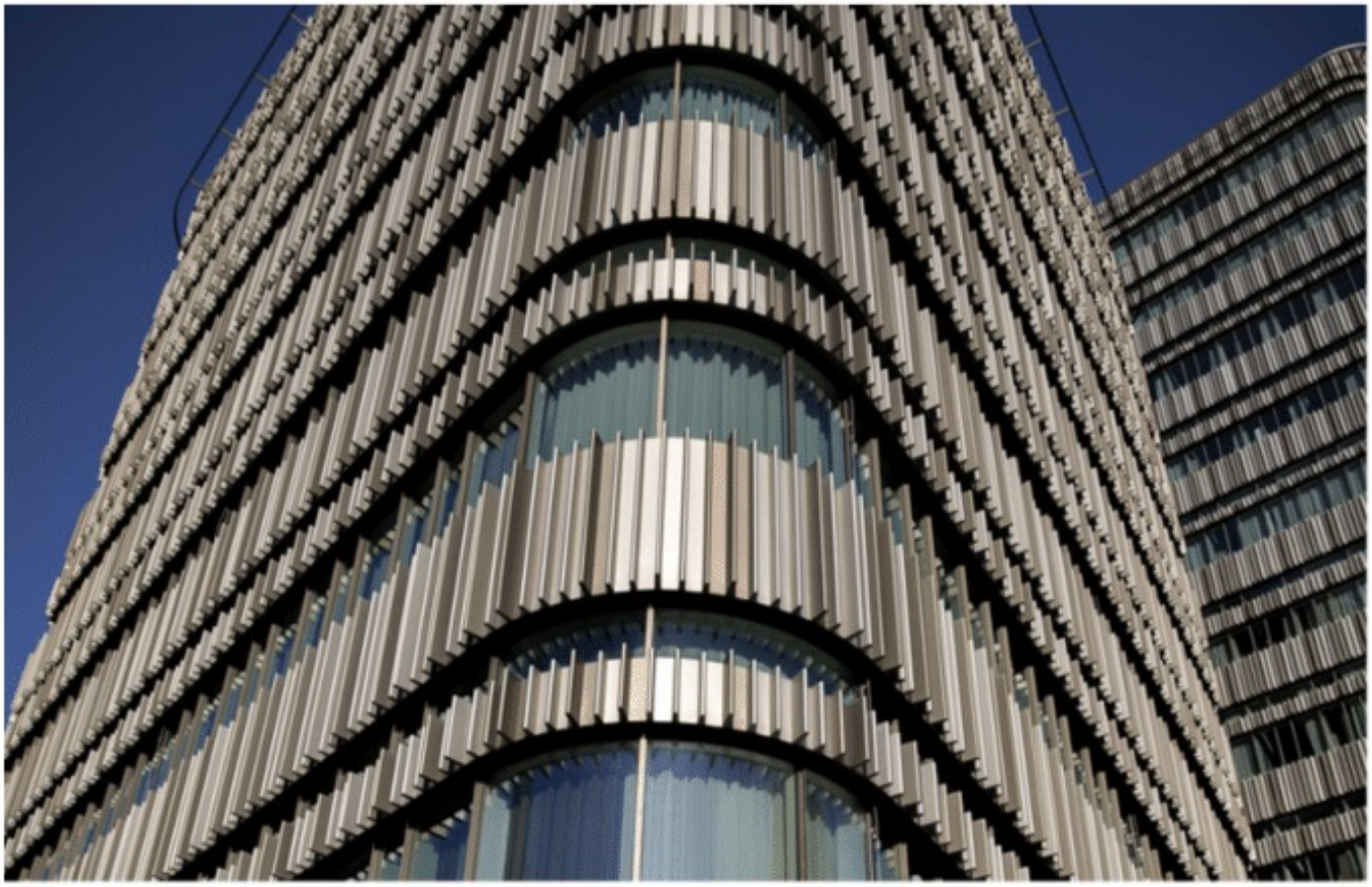
Conclusion

- Attacks were detected using the Random Forest classifier with an accuracy of 98%
- The proposed system can detect new types of cyber attacks without requiring any hard-coded rules
- Easy to extend the system

Future Work

- Leverage Complex Event Processing to help collect and analyze IoT data streams in real time
- Extend the system to be able to detect other attack types, including low-level attacks
- Develop interfaces that can help notify building administrators when the system detects anomalies in real time

Thank You for Your Attention!



Joseph Bugeja
joseph.bugeja@mau.se

Jan A. Persson
jan.a.persson@mau.se

MALMÖ UNIVERSITY
INTERNET OF THINGS AND PEOPLE