Anomaly Detection for Privacy Preserving Time Series Building IoT Data

Akbar Fadiansyah – Supervised by Dr Chehara Pathmabandu

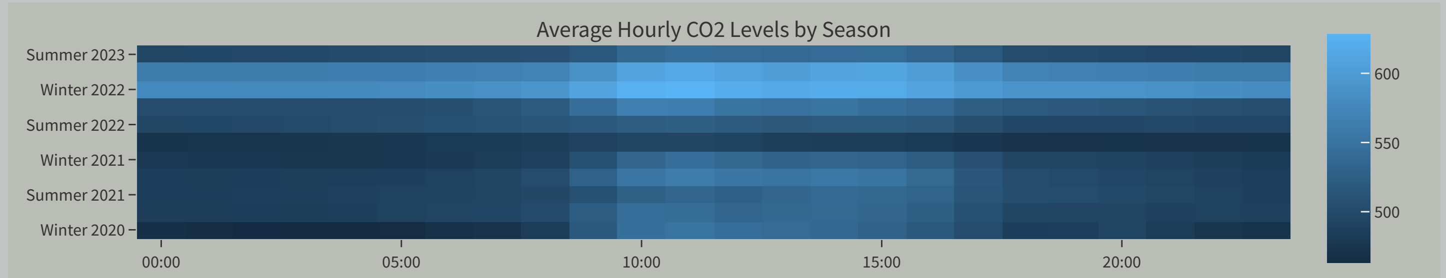
CSIRO – Data61

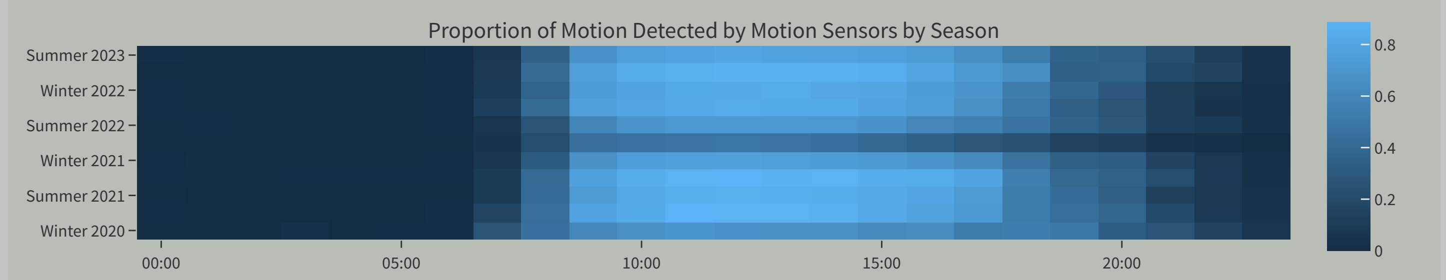
Introduction

- Anomalous or unusual readings from IoT sensors may present privacy threats to the building and its occupants
- These data can be used to infer activities in a building and use it for malicious purposes, such as finding times and points to break into buildings. [1], [2]
- We developed an anomaly detection system for time series building IoT data, and proposed a system to preserve privacy without impacting its functionality.

Dataset

- We trained CO_2 and motion sensor data from air handling units (AHUs) and rooms of CSIRO sites from April to December 2022
- Building and equipment ontology (information about its properties) are from Data Clearing House (DCH), and raw observation data are from Senaps
- The training set period is chosen based on data availability and to avoid biases from COVID-19 pandemic
- CO_2 and motion sensor data are chosen as both have strong correlation with building occupancy (see graph below)
- The training set is unlabelled, meaning there is no annotation to indicate which data points are anomalous.





Model

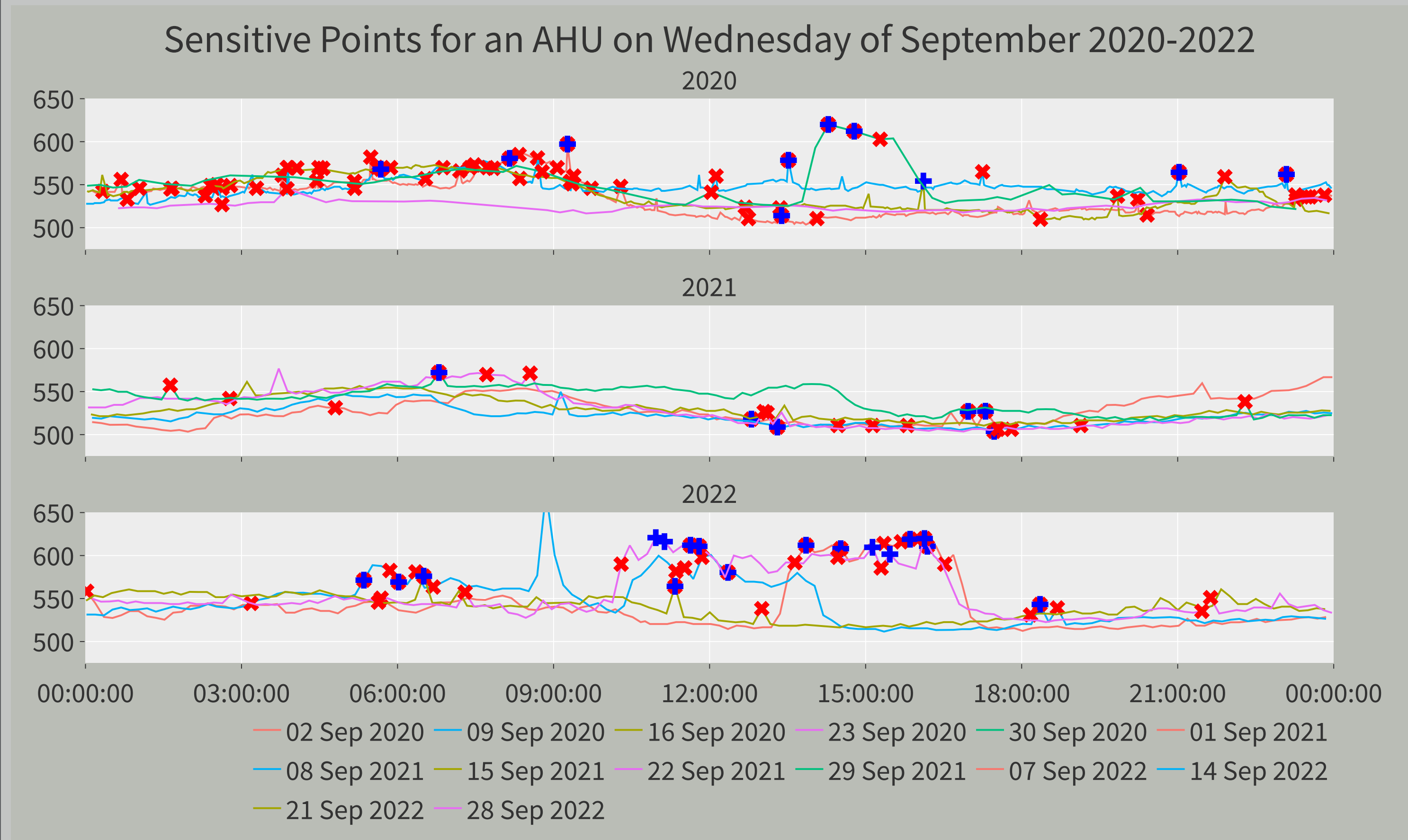
Two types of anomaly detector algorithms are used:

- Local Outlier Factor (LOF) – for short term/local anomaly detection (deviations from neighbouring values)
- k-Nearest Neighbours (KNN) – for long term/global anomaly detection (extended deviations from 'normal' observations)
- Detectors are selected based on its performance on the respective anomaly type, according to ADBench [3]
- There is no single detector that performs the best on all types of anomalies [3], [4]
- We used PyOD which is a Python anomaly detection library to train and test models, using its default parameters for each detector [5]
- Each AHU and room have its own model, trained on sensors associated to that room.

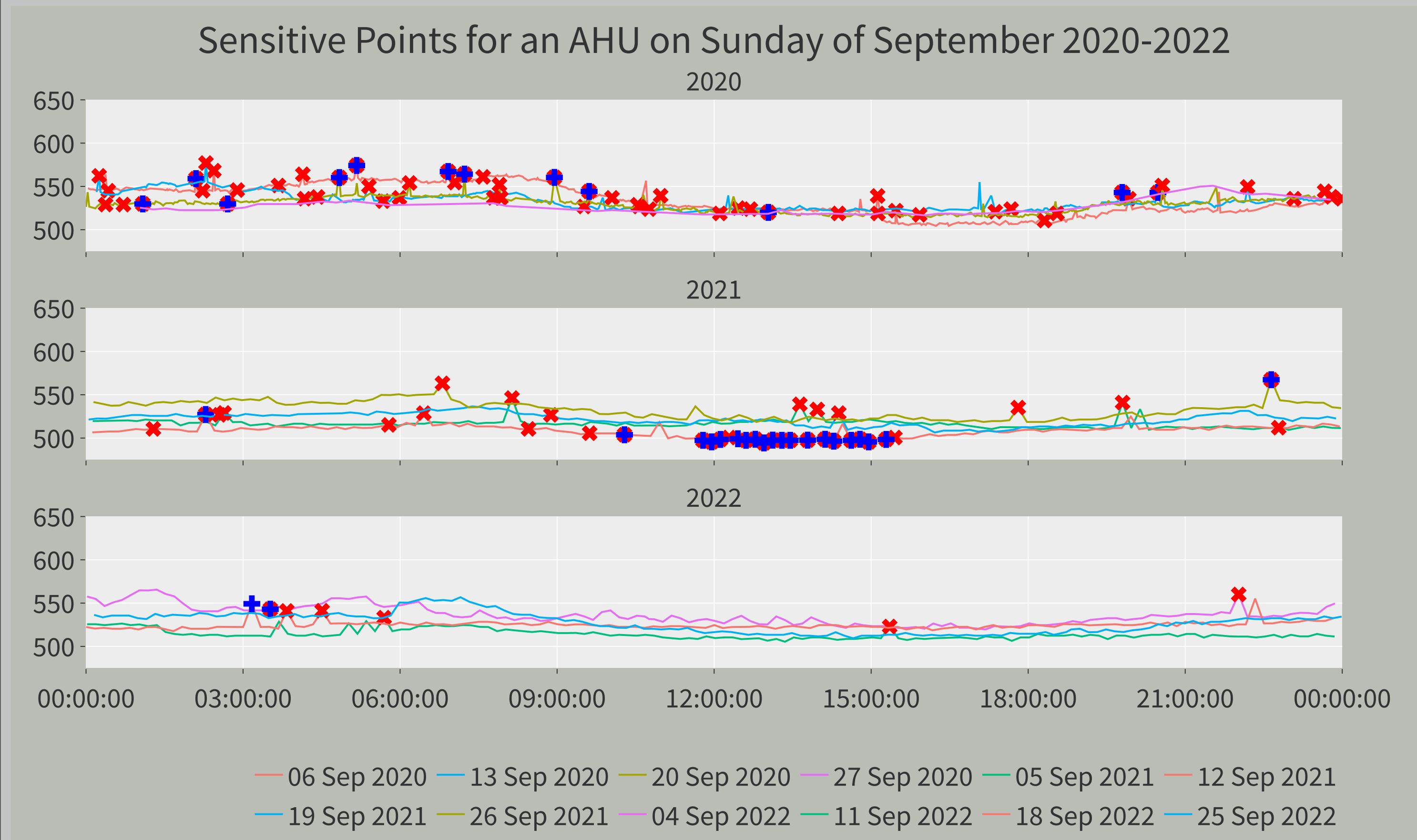
Results

- The graph below shows anomalies detected from the training set as well as test results from 2020 and 2021 for a CO_2 sensor in an AHU
- Plots a grouped by day of the week (Wednesday and Sunday shown here), and each plot is further divided to each year
- September is selected due to the lockdowns in both 2020 and 2021, to see if there are significant patterns and differences in detections
- Red cross (×) indicate short term anomalies and blue plus (+) indicate long term anomalies.

(a) Wednesday



(b) Sunday

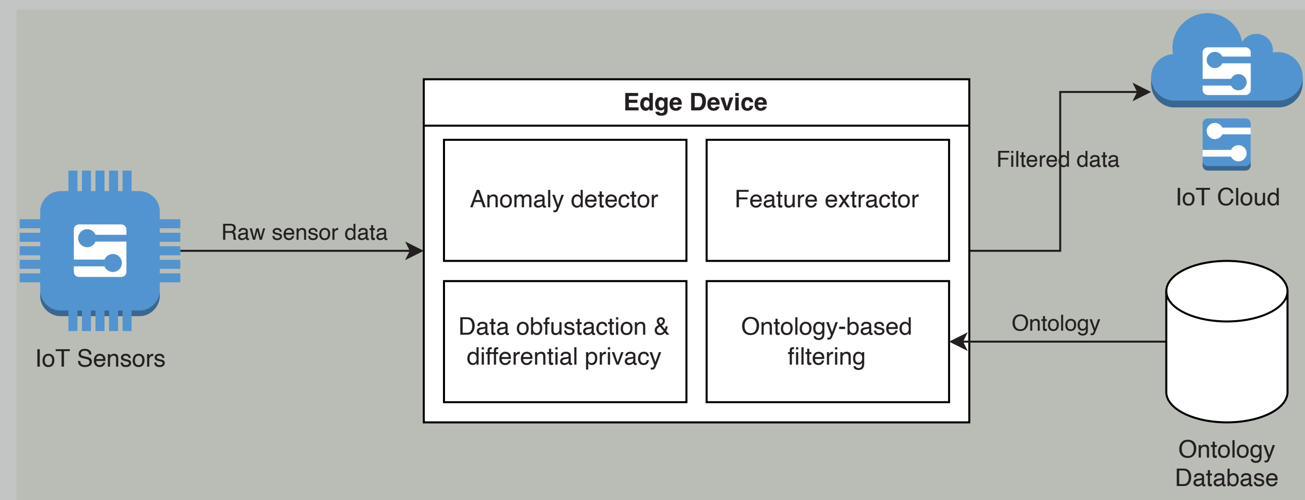


Discussion

- In general, there is a lack of significant rises in weekdays of 2020 and 2021 during business hours, especially in 2021
- This can be largely attributed to the lockdowns during that time period
- There are also no rises during office hours on Sunday for all years, as expected
- Interestingly, there is an inconsistent pattern in the rises for Wednesdays of September 2022, suggesting that hybrid work is taking place
- The long term detector's accuracy is impacted by our post-pandemic training set.

Proposed privacy preserving IoT system

- Based on local differential privacy model, where the data curator (e.g. cloud) is considered untrusted [6]
- An edge device sits between the sensors and cloud, applying techniques such as differential privacy to filter out sensitive data before being sent
- The edge device can be a low cost, low powered computer, such as Raspberry Pi.



Conclusion

We developed anomaly detection system for time series IoT data, and demonstrated its current capabilities over various time periods. We also proposed a privacy preserving IoT system based on edge computing.

Special thanks

- Chehara Pathmabandu
- Edric Matwiejew
- Mahathir Almashor
- John McCulloch
- Pawsey Supercomputing Centre
- PyOD

For the mentoring, computing resources, dataset, software and everything else in between.

More informaton

References, methodologies and more.
Scan me!