Exercise 2: IoT Data Analytics

Sofeem Nasim, Aditya Mehta, Xinyi Yang
sofeem.nasim@oulu.fi, amehta20@student.oulu.fi, xiyang18@student.oulu.fi

21 April 2020

1 Visualisation

1.1 Dataset Description

The dataset for this study is from the Tellus, where various sensors have been installed to monitor parameters such as temperature, CO2, humidity, light, PIR. It includes 331 sensors installed in various areas as the figure 1 shows. We used Open Street Maps to locate various sensors and grouped sensors of a particular area with a specific color for better representation. Upon hover, the location group, latitude, and longitude values of the sensor can be seen easily.

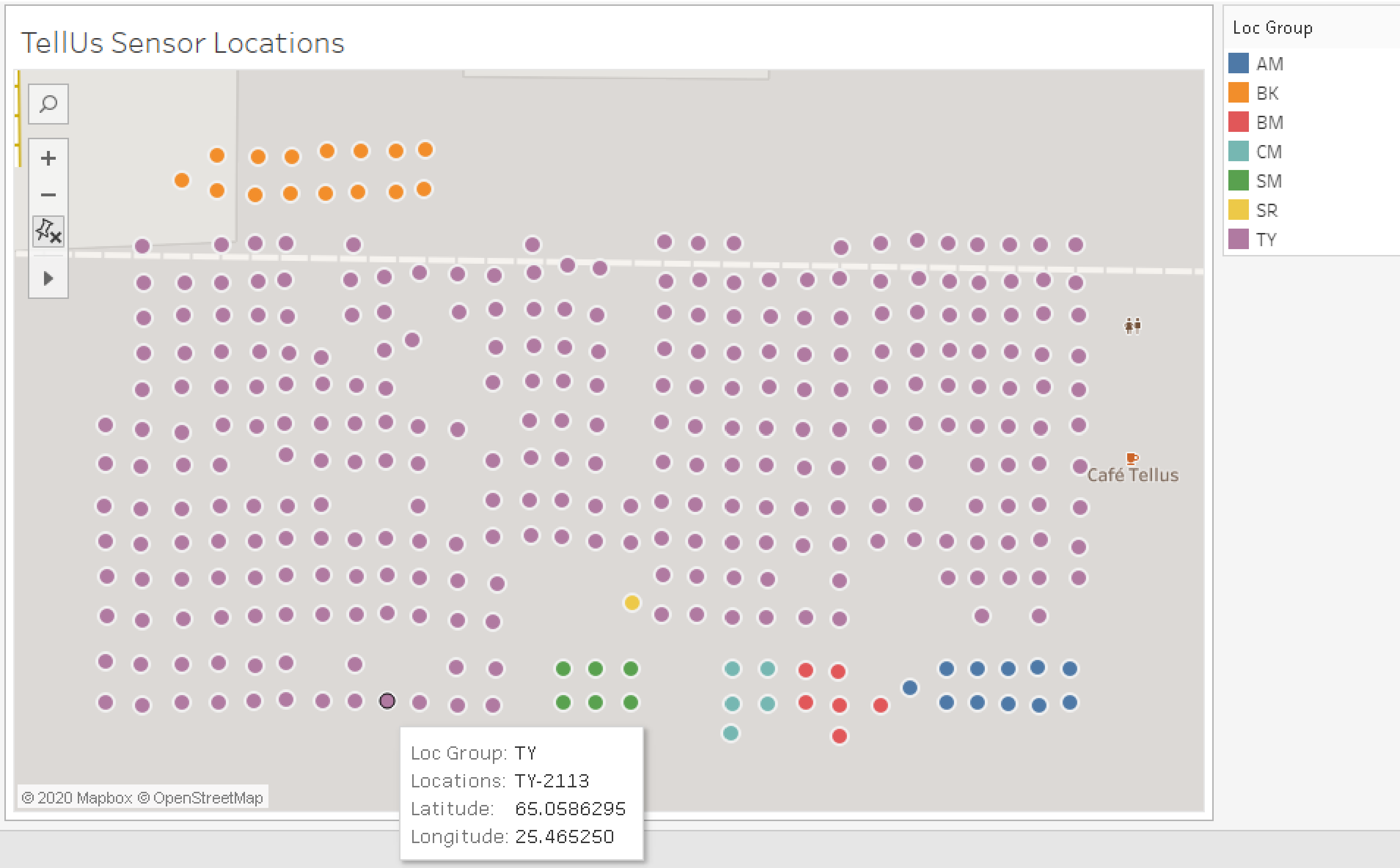

Figure 1: Tellus sensor map

1.2 Data Pre-processing

Before proceeding with the actual analysis, we have to make sure that the dataset has been carefully analyzed for null values, outlier detection, and their removal. As per standard normal curve for usual quantities, an outlier is considered as some data point, which is beyond the three standard deviations from the mean. In this dataset, first, we checked if there are any null or missing values. We found that there are no such records. After that, we identified outliers for various quantities and removed them. In this dataset, outlier examples include temperature values in the range of 625 to 675-degree Celcius, CO2 values in the range of 65000 ppm, which is unaccepted in normal living conditions. The cleaned dataset contains various readings with the timestamp from 2017-06-26 15:12:26 to 2018-11-20 11:30:40 with

9425931 rows. Data cleansing process has been briefly summarized in table 1

We used the sensor's locations file and merged it with a cleaned dataset. We identified that there were many rows in original data, for which no corresponding matching sensor ID was found in the sensor location file, hence we removed such entries.

The PIR data column is the number of motion counts detected by the sensor and it varies from 0 to 255. It represents the number of detected motions between the transmission timestamps, which is approximately 15 minutes per dataset. The PIR increments when the motion is detected and after transmission, it resets to 0. It may stay inactive for some seconds right after detection. To represent motion detected by the sensor, sensor reading can also be considered as binary, i.e. 0 which means no movement and 1 or more means movement. In further processing of the dataset, we created a boolean variable to count the number of activities detected by the PIR sensor.

Table 1: Data PreProcessing Desctiption

Number of records in raw file	9917848
Are there any missing values for all columns?	No
Records after outlier removal for all columns	9509645
Records after merging with sensor locations	9425931
Are there any duplicate records in cleaned file?	No

To facilitate the grouping of sensors based on their location, we added one column loc_group for identification. The cleaned dataset has been described in table 2:

Table 2: Dataset Desctiption

Column Name	DataType
id	object
timestamp	object
co2	float64
humidity	int64
light	int64
temperature	float64
pir	int64
Locations	object
Latitude	float64
Longitude	float64
loc_group	object
pir_status	bool

1.3 Description of animated visualization

The animation used for visualization describe the variation of different location activity occurrence happened during different season. We have only used summer and winter time. The pattern depicted from the reading of PIR sensors showed that the number of activities observed during summer time is less as compared to the activities happening in the winter time.

1.4 Link of animated visualization

The visualization has been created as a video format and available on the following link
https://drive.google.com/open?id=1gDaSUJgnSAhwLtndxnUqVwOxO3edKH9t

1.5 PIR sensor

The PIR sensor is a passive type of sensor, which is capable of detecting infrared radiations that are emitted by any object which comes in front of the sensor. The pyroelectric sensor is the main component of a PIR sensor, and it can be divided into two parts (A & B). When a person crosses the area covered by the PIR sensor, the infrared radiations level in part A will be higher than part B. Thus, the PIR sensor will react to this change and create a higher output voltage. [4]

Passive infrared sensors (PIR sensors) have been used in different applications for different usages, such as detect human activities in a specific area, lighting systems, etc. On the one hand, the PIR sensors are low cost, less intrusive, the immunity from multi-path fading, and work well in a low-light environment[3]. It has been widely used by companies in a very low-cost right now. On the other hand, the PIR sensor not able to detect the number of people around it. Also, most of the PIR-based localization systems require a huge amount of PIR sensor because they only use the binary information[5].

The PIR sensor output can be utilized to estimate person's position by using azimuth change calculations, or to count number of people in the room. PIR sensors generally can be configured using the delay and sensitivity functionalities. The delay can be used to set a time interval delay after the motion detection, while sensitivity can be used to configure the distance to be monitored by the sensor. In the Tellus area, there are 331 sensors and they generate an output signal round the clock with a time interval of 15 minutes from one reading to the other.

2 Seasonal variations

2.1 Motion activities detected by sensors at various locations

As per our data analysis, most of the readings of PIR are zero, i.e. sensor did not track any motion activities. To find out which sensors in a particular area contains more motion detected out as a percentage of all readings by those sensors, we calculated various percentages values. As depicted in figure 2 from the overall dataset combined for all sensors, only 34.17% values are with some motion detected. The aspire meeting room, brisk meeting room, and chill meeting room contains approximately only 15% motion detected from all their readings. The staff meeting room contains 23.28% while the staff room contains 25.96% of such motion activities. The TellUs public area contains 37% of motion activity detected from all readings recorded, which is higher than even the overall average value of 34.17%. However, this can be attributed to the reason that there are 287 sensors installed in the public area, while other areas contain sensors in a range of only 1 to 15. This high ratio can be attributed to the TellUs architecture design and these sensor installation plans as implemented a few years ago.

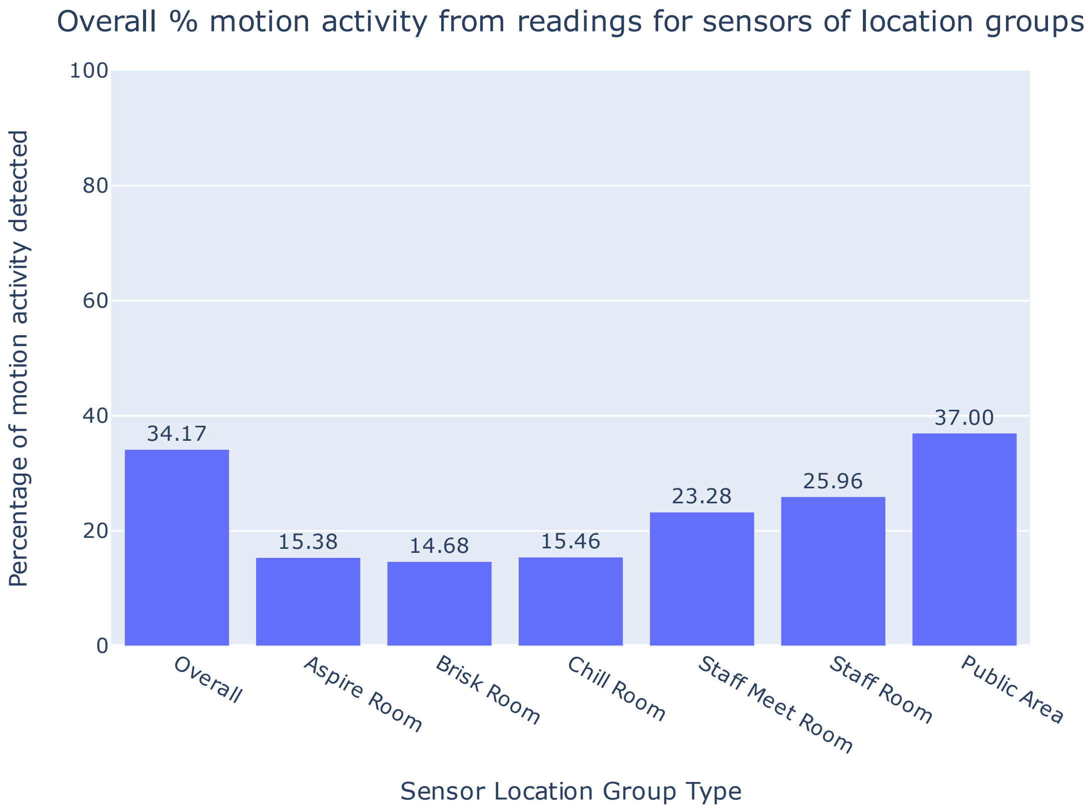

Figure 2: Percentage of motion activity detected as a proportion of all readings

2.2 Analysis of public area sensors for monthly motion activities

To identify monthly variation in public area activities, we analyzed all activities of public area sensors in a particular month and found out the overall trend for monthly motion activities as depicted in figure 3. The graph has been arranged in monotonically increasing order as per the number of motion activities tracked in that specific month. As per our analysis, the month of January from 2018 had the most motion sensing activities (769902 motion activities) in the analyzed time period. This can be related to the new spring semester beginning and orientation sessions for new students, which are typically arranged in the first week of January at the university. However, the December 2017 and March 2018 also has comparably similar activities, which can be attributed to study periods in the respective semester at the university.

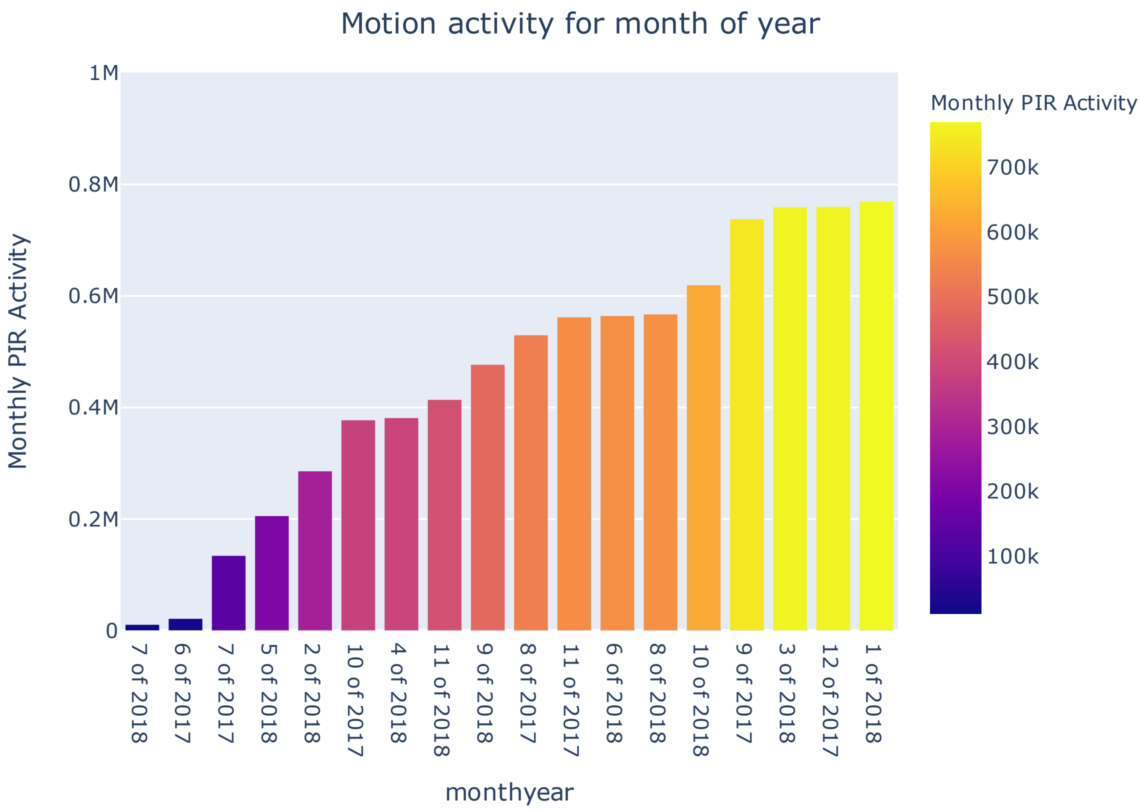

Figure 3: Monthly PIR Motion activity in public area sensors

2.3 Analysis of public area sensors for daily motion activities

As explained in the previous graph, the public area sensors have tracked the most number of motion activity in the given time period. Hence, we focused our analysis on this sensor group. For all sensors of public area, we calculated the overall count of activity, tracked on a daily basis over the time period and plotted as depicted in figure 4. The daily activity graph has been created with color gradients as per the count of motion activity tracked on that day, with various colors as seen in the graph. As the yellow areas in the graph depicts the highest motion activity in the public area. On close analysis, we find that on a daily basis, 26 Nov 2017 has the highest motion activity detected in a day. This can be attributed to some event, for which there might be a public gathering in this area.

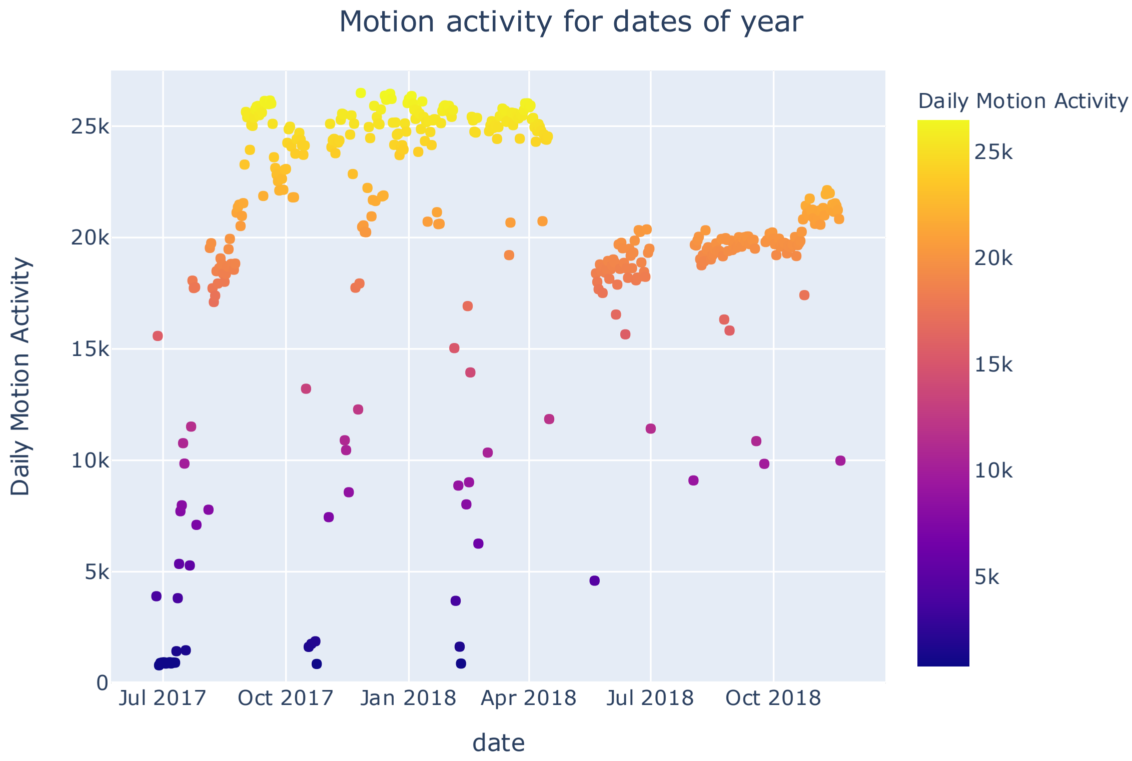

Figure 4: Daily PIR Motion activity in public area sensors

For the seasonal patterns in the daily activities, we plotted daily motion activity for the entire time period as depicted in figure 5. This is an interactive graph. We can easily see that from the time period of Oct 2017 to mid of Apr 2018, the higher number of activities have been tracked as compared with another time period as shown in yellow lines on the graph. Also, it can be understood that from July to October 2018, the daily motion activity has been comparatively lower.

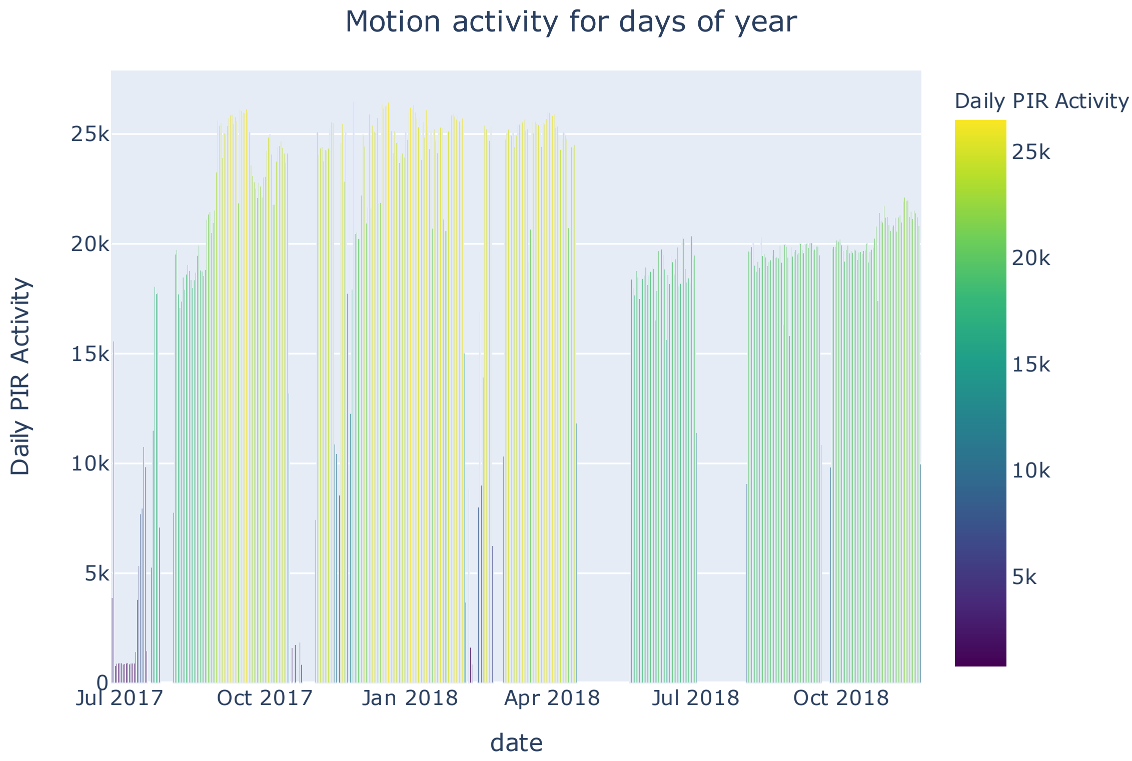

Figure 5: Daily PIR Motion activity in public area sensors

3 Future implications

3.1 What can be learned from seasonal variations in the data?

As section 2 mentioned, the public area has the most number of motion activities. Through tracking these activities, we provide the potential applications which could benefit the usage of Tellus from three perspectives.

3.2 Security issues

PIR sensor has been used widely used in the security system through detecting human activity in the specific area [6]. It means if it has intrusion, the alarm will be raised [1]. For the University of Oulu, the PIR sensor can work together with the IP camera to protect property safety. The IP camera allows user to view and record the real-time video [4]. There are always students stay in the university until the middle of the night, it means PIR sensors work alone not able to provide valuable data. Through working together with the IP camera, it is easy to check the real-time situation in the university.

3.3 Corporate social responsibility

Corporate social responsibility (CSR) is a sustainable model for businesses that affect all aspects of a company's operations. It is a view of the corporation and its role in society that assumes a responsibility among firms to pursue goals additionally to profit maximization and responsibility among a firm's stakeholders to hold the firm accountable for its actions. Nowadays, one of the important stakeholders in the environment.

The University of Oulu attaches great importance to sustainability, such as environmental program, energy efficiency, sustainable development action plan, etc [2]. For the Tellus and the rest of the university. PIR sensors are used together with the other equipment after 7 pm during the weekdays to save electricity and energies. For example, the induction lamp, air-conditioning, etc.

3.4 Cafe management

There has one Cafe located in Tellus to sell coffee, tea, fresh desserts, candy bars, and cold beverages. Sometimes, fresh desserts are unsalable. The Cafe can do a cross-analysis between the data from PIR sensors and their sales records to test if their any correlations between sales and daily human activities in Tellus. Then, if these two datasets have correlations, the Cafe can use this relationship to predict the sales. In this way, the Cafe able to save some costs and wastes.

References

[1] M. Moghavvemi and Lu Chin Seng. Pyroelectric infrared sensor for intruder detection. *Proc. IEEE Region 10 Conf*, 2004.

[2] University of Oulu. Sustainable campuses.

[3] Yordan P Raykov, Emre Ozer, Ganesh Dasika, Alexis Boukouvalas, and Max A Little. Predicting room occupancy with a single passive infrared (pir) sensor through behavior extraction. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1016–1027, 2016.

[4] K. C. Sahoo and U. C. Pati. Iot based intrusion detection system using pir sensor. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pages 1641–1645, 2017.

[5] Tianye Yang, Xuefeng Liu, Shaojie Tang, Jianwei Niu, and Peng Guo. A new pir-based method for real-time tracking, 2019.

[6] Ali Yassin, Youssef Nasser, Mariette Awad, Ahmed Al-Dubai, Ran Liu, Chau Yuen, Ronald Raulefs, and Elias Aboutanios. Recent advances in indoor localization: A survey on theoretical approaches and applications. *IEEE Communications Surveys & Tutorials*, 19(2):1327–1346, 2016.