Passive WiFi Monitoring of the Rhythm of the Campus

Edward Verbree,
Stefan van der Spek

Delft University
of Technology
Faculty of Architecture
Delft, the Netherlands
e.verbree@tudelft.nl,
s.c.vanderspek@tudelft.nl

Eftychia Kalogianni,
Rusnė Šilerytė,
Marco Lam,
Kaixuan Zhou,
Martijn van der Ham

Delft University of
Technology
MSc. Geomatics
Delft, the Netherlands
geomatics@tudelft.nl

Abstract

Within this research-driven project, passive WiFi monitoring of WiFi enabled devices was used to detect users (students, employees, visitors) of buildings at the campus of Delft University of Technology to gain insight into the Rhythm of the Campus: the occupation, duration of stay and moving pattern at and between the different facilities (faculties, library, auditorium). The research was carried out in request of the University's department of Facility Management and Real Estate (FMRE). In total, 20 WiFi monitors were located at the library, the auditorium and five faculties for one week. Simultaneously, ground truth data was manually collected for two days at the library and the auditorium, using an online questionnaire and a counting application. The retrieved data was classified and filtered, which facilitated data analyses and validation. Resulting from statistical analyses a relation between the number of detected mobile devices and the amount of people inside the buildings was developed. This semi-automated method made it possible to get new insights in building use compared to conventional manual counting methods. For visualisation, an online dashboard was developed as a tool to help the client at the decision making process. The research presented in this paper was carried out as a proof of concept. The results presented in this paper give a concise overview of the extensive possibilities of passive WiFi monitoring for facility management.

Keywords: Passive WiFi monitoring, occupation, duration, data visualisation, dashboard.

1 Introduction

Indoor navigation has become very popular in recent years and this technology is leading a new perspective of data storage, analysis and mining as it results in big data. Location-Based Services (LBS) have become an integral part of modern life (3) and the use of WiFi network of mobile devices is increasing rapidly. More and more public places are offering (free) WiFi, which makes people use their location sensitive smartphones outside their homes a lot more than before. This development provides a platform to track people by using WiFi monitoring technology.

Using the primary progress in indoor location sensing systems, this paper aims to provide the reader with a research-driven project description for indoor positioning and estimation of the number of people in and around an area of interest using WiFi data. The case explored here is to identify the use of the buildings and in general, of a campus of a university and their patterns. In particular, the project client is the department of Facility Management and Real Estate of the university and the project was done in terms of a broader perspective of moving the campus to the next century. As a result, a dashboard has been created to provide the client with a tool to make better-informed decisions about the future use of the campus.

This research aims to detect users of buildings at the campus in order to gain insight in the occupation and use of them. Information about the behaviour and movements of people have been studied ever since manually. A 'proof of concept' has been provided to collect this information in an automated way. The focus of this research is on the central library and the auditorium conference centre as they form the heart of the campus since they are used by a large number of people with different backgrounds.

The main question to be answered in the research was:
'What kind of patterns of using the area of the library & auditorium can be recognized by WiFi-monitoring within the campus and how reliable are these results?

2 Related Research

Automated technologies, which enable people counting and pattern tracking, have been used in many scientific researches across the world. Recently, WiFi tracking technology has exploded due its cost effectiveness and ability to track repeated visiting patterns. The collected data is extremely valuable to various users ranging from city planners, business owners and police, interested in crowd control to civilians both as a form of intelligence-gathering or everyday problem relief.

Analysing spatial patterns of urban behaviour has been studied by Van der Spek et al. (4,5) and McArdle et al. (2). The majority of work in this area relies on GPS-devices used by active participants in the process. WiFi monitoring detects the signal that WiFi-enabled devices are sending out all the time in their search for a WiFi access point. As the user is not actively involved in this process – he or she should be informed that this kind of measurements are taken place – we regard this technology as a kind of passive localisation, albeit the WiFi-enabled devices are active.

The information that can be derived from these signals can be useful for crowd control, marketing purposes or real time monitoring of public space use, as discussed by Ogawe (3). Among others, Musa and Eriksson (1) used passive WiFi tracking for getting trajectory estimation for moving devices. The detection method is mainly based on obtaining the beacon frames from the monitor that functions as an access point and make connection between devices and the access point.

3 Data Collection process

During one week data has been collected in the library and the auditorium. As reference locations, monitors were also placed at five of the faculties on the campus. In this chapter the setup of the research, issues encountered regarding the passive WiFi monitoring, locations of the monitors and the collection of the ground truth data is discussed.

3.1 Monitors used

The passive WiFi monitoring detects wireless signatures that smartphones periodically transmit and can be used for detecting, tracking and counting people. The key advantage of this technology is that it works without any modification, as there is no need for installing an app on the smartphone or user interaction. The only requirements are that people should have a mobile device with WiFi functionality and that they have the WiFi function switched on. Then the smartphone sends out a signal, where the MAC address of the device is included, which is unique to the device. The MAC address does not contain any other personal information as names, phone numbers or any other communication data. In order to ensure privacy, MAC addresses are hashed, i.e. anonymised, directly by the monitor.

3.2 Distribution of the monitors

The research covered the area between the library and the auditorium to investigate relations and use in a detailed way. Data collected at the entrances of five faculties was only used as a reference to the data collected at the main area of interest (Figure 1). The coverage of these five faculties was not good enough to do accurate analyses because too few monitors were placed. The data from these faculties was used to cover students and staff working in these buildings and also using the library and the auditorium.

The devices are relatively small and can be placed out of sight. The collected data was sent automatically to a database. Placing of the monitors has been arranged and approved by FMRE and responsible people of the building units. It was coordinated with the legal department of the university due to potential privacy related issues.

The core requirements taken into account for the selection were: the monitor reception must cover the whole floor area of the university's library and catering area of the auditorium; must cover most popular spaces of the different faculties for mobility detection; should exclude areas, which are not of our interest (i.e. the outdoor environment); a power outlet is needed throughout the whole data collection period; the effect of walls and other obstacles needs to be considered.

Figure 1: Distribution of the monitors over the Delft University of Technology campus area.

3.3 Ground truth data collection

The validation data was collected manually during two full days at the library and one day at the auditorium by employing a self-designed people counting application for an android device and an online questionnaire for the survey.

To be able to answer the research question a survey has been set up to collect data. A counting application was used as well during the survey. In total, 559 random respondents were asked to answer the survey upon entering or leaving the library and the number of people inside the library was being registered every 5 minutes by manually registering all entries and exists.

This data is used to mitigate the error of the data collected by the monitors. Furthermore, errors are included to this data as it is a manual work and the data collection period was limited.

4 System Architecture

A short description on how the system work, in general terms, is needed for understanding the Rhythm of the Campus dashboard applications.

The raw data from the monitors contains mobile devices detections per every 10 seconds, however due to its sheer size it is hard to process. Furthermore not all the devices which are present in the area will be detected every 10 seconds. Therefore, data was simplified to detections per 15 and 60

minutes and classified to groups, which contain all the devices in the area of a specific monitor.
Choosing a longer timespan was also possible, however people might not stay more than a few minutes inside the building, the short stay in that case wouldn't be distinguishable from people who have stayed for a longer time. The interval shouldn't be very long, neither too short, in order to ensure that not too many devices will be left outside of the dataset regarding that time span.
Moreover, the data of devices detected by several monitors in the same building during the same sample period of time were aggregated. This was done for the identification and elimination of duplicate entries.
Together with the scanning data, information about the 'health' of the monitors was sent to the server. If a device didn't send or the server didn't receive data from a monitor it is considered to be offline after a certain amount of time has passed from the first detection. The chosen time span is based on the delays in the parsing of the data. In this case an email was sent to the team members notifying them so that unwanted data loss was minimized.

5 Data filtering, analysis and validation

Before analysing the data, it needed to be filtered, as it contained duplicates, static devices and different groups of outliers. In order to work on a "clean" dataset, as many outliers as possible had to be removed from it. Duplicates in this case are caused by devices identified at the same time from multiple monitors. For this project, routers, printers and monitors are considered as static devices. Furthermore, the different groups of outliers can be defined as:

- Devices detected outside of the building,
- Devices forgotten by the user inside the building (during the building closing hours).

Only the data retrieved during the opening hours of each faculty is useful for the analysis. Therefore, data outside the building closing hours was removed from the datasets in order to avoid faulty detections of unidentified static devices or passers-by. For excluding the monitor data regarding the closing hours, a Python script was used. For exploring the mobility between faculties, auditorium and library, the mac addresses were assigned to certain faculties. So that the patterns of how different faculties use library and auditorium can be detected..

Data analysis is an intermediate step between raw data to intelligent data visualisation used for pattern recognition. At the first stage of the project goals and requirements for the analysis have been defined: analysis related to building occupancy; definition of user groups based on their host faculties and identification of building relations based on user mobility between them.
Furthermore, validation was needed to evaluate the reliability of collected data and consequently patterns retrieved from it. Ground truth data from manual people counting and survey at the library and auditorium were used as comparison data for validation.

Measuring occupation of library and auditorium is one part of the analysis requirements, which is needed to define patterns of space busyness. The relation between library and auditorium is also analysed based on the comparison of occupancy patterns between the two buildings.
A crucial part of measuring building occupation was defining a correct function between the number of devices detected by the monitors and an actual number of visitors. This was done by taking into account people counting data both in library and auditorium (ground truth data) and a relevant piece of data taken from the WiFi monitoring devices.

5.1 Distribution by stay duration

One way to group library and auditorium users is by looking into their stay duration patterns. This kind of analysis is needed to identify how often the same people visit library and how much time per day/week/month they stay there.

5 types of user groups have been defined due to the varying purpose of stay:

- Temporal stay (<15min),
- Short stay (15min-1hour),
- Average stay (1 - 4h),
- Long stay (4 – 8h) and
- Extra-long stay (>8h).

Combination of users groups, thus stay duration displayed by faculty also reveals a lot of information about usage patterns since stay duration is highly dependent on the staying purpose. Moreover, comparing stay duration of the same user per weekday/weekend/whole week and aggregating users by the faculties reveals campus-wide long term patterns of facility usage.

5.2 Combination of user groups: stay duration per faculty

In order to speed up queries, two types of tables were used – one containing library data aggregated per every 60min and the other one containing library data aggregated per every 15min. First type of data is queried faster and can be used to find users of average to extra-long stay, while 15min data is used in case a user has been detected in 2 or less time spans of 1 hour. In that case the same operation is preceded to find more precise stay duration. Both user groups can be combined for more explicit analysis.
A number of inadequacies have been identified while comparing data derived from the monitors with a survey conducted at the library:

1. The monitors are able to detect devices, which are outside the buildings, which is especially erroneous in cases where faculties are located close to the main cycling paths.
2. Monitors do not detect all the devices at all times of their presence.
3. Data obtained from the questionnaire can only be considered "ground truth" up until a certain extent since stay duration values are roughly estimated by the respondents and not measured accurately.

6 Visualisation and pattern recognition

An online dashboard was developed to provide a tool, which can help the client to use this data in policy and decision making processes. The dashboard contains various kinds of data visualisation in forms of charts, histograms, dynamic maps and tables, as well as raw data and information about current status of the monitors for the time when monitors are active.

From the dashboard visualisations, different kinds of sample patterns (given that the amount of collected data is only a sample one and does not represent the overall situation) can be identified:

1. Length of stay duration in the library or auditorium (e.g. on weekdays 1/3 of library users tend to stay for 1-4 hours, while on weekends 50% of them stayed for more than 4 hours).
2. Temporal behaviour of different types of users grouped by stay duration (e.g. most of the short-stay users go to the library during lunchtime, however those who are staying for less than 1 hour, are registered mostly during all working day hours. Those who were staying for 4 to 8 hours rather prefer morning hours to evening).
3. Statistics on space busyness in terms of occupation (e.g. during weekdays the library reached its peak of visitors (~750) in the afternoon, while the auditorium reaches its peak during the lunchtime (Figure 2).

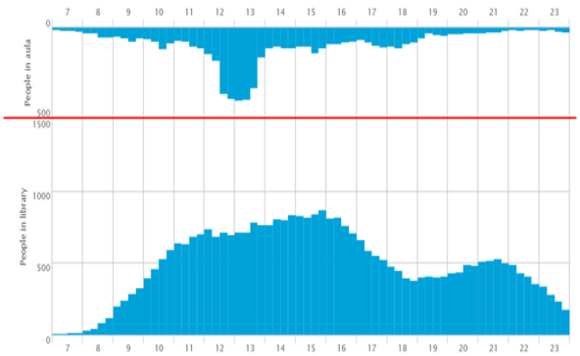

Figure 2: Occupancy histogram of library and auditorium on Wednesday, 24th of September.

7 Conclusions and recommendations

The project investigated what kind of patterns of using the area of the library & auditorium can be recognized by WiFi-monitoring within the campus and how reliable these results are. By use of the dashboard, space usage patterns could be found. It was possible to see how many people are at a certain location and time of the day. As a result it can be seen when people usually arrive and leave buildings.

Furthermore, patterns on how user groups (grouped by stay duration and faculty) use the buildings could be seen; for instance when they usually are in the buildings and which user groups make use of certain buildings more often than others. Migration between the faculties, auditorium and library was also visible up to a certain extend. In general, the statistics for merged data are relatively reasonable and can clearly show the relation between people and devices detected. The formula generated can be used for calculating occupation based on the devices detected in the library. Even it is still not fully reliable for the auditorium, the derived filtering criteria have strongly increased the performance of linear fitting.

To sum up, passive WiFi monitoring is an effective way to identify building occupancy and moving patterns in between the buildings. However, passive WiFi monitoring in current stage of the project is not reliable when identifying movement patterns between faculties and the area of the library the auditorium.

From the limitations of this project several recommendations can be provided for future implementations. The detection of users is not perfect when only monitors are placed only at the entrances of buildings, which is done in the faculties. Placing more of these monitors inside of the buildings to get full coverage will result in better classification of faculty users and easier filtering of people detected while being outside.

The result of the occupancy of the Auditorium wasn't as satisfying as the library, because not enough ground truth data was collected for the auditorium, therefore a larger amount of ground truth data should be collected for every building of interest for a larger period, both indoors and outdoors.

The boundary between indoors and outdoors is fuzzy; this effect can be limited by using directional WiFi antennas. Furthermore, for getting the near real time mobility map, more monitors should be placed in the places where people are most likely to go in order to assure that everyone can be detected continuously. It is important that distribution of monitors and their coverage capabilities should be tested in depth to ensure full coverage of a building.

References

1. Musa A.B.M., Eriksson J. (2012). Tracking Unmodified Smartphones Using WiFi Monitors. SenSys'12, Toronto, Canada.
2. McArdle, G., Demšar, U., van der Spek, S., & McLoone, S. (2014). Classifying pedestrian movement behavior from GPS trajectories using visualization and clustering. Annals of GIS, 20(2), 85-98.
3. Ogawa K., Verbree E., Zlatanova S., Kohtake N.,& Okhami Y. (2011). Towards Seamless Indoor-Outdoor Applications: Developing stakeholder-Oriented Location-Based Services. Geo-spatial Information Science 14(2): 109-118.
4. Van der Spek, S., Van Schaick, J., De Bois, P., & De Haan, R. (2009). Sensing human activity: GPS tracking. Sensors, 9(4), 3033-3055.
5. Van der Spek, S. C., Van Langelaar, C. M., & Kickert, C. C. (2013). Evidence-based design: satellite positioning studies of city centre user groups.Proceedings of the ICE-Urban Design and Planning, 166(4), 206-216.