

Track-id: Activity Determination based on Wi-Fi Monitoring

S.C. van der Spek, E. Verbree, C.W. Quak, IJ.D.G. Groeneveld, R. Sulzer, E. Theocharous, M.S. Tryfona, O.T. Willems and Y. Xu

Delft University of Technology, Geomatics for the Built Environment,
Faculty of Architecture and the Built Environment,
Julianalaan 134, 2628 BL, Delft, The Netherlands

Extended Abstract

The distribution of people in buildings, the occupancy of lecture-, work- and study places and the accessibility of facilities are essential information at university campuses who have to cope with limited and even shrinking budgets and huge, rising real estate costs. Only little insight is gained in both occupancy and movement patterns with traditional counting techniques and user-based questionnaires. Management teams state that rooms and facilities are hardly used, though staff and students complain about overcrowded facilities and limited flexibility. Actual and accurate data on a 24/7 scale with high-granularity is missing.

In general Facility- and Asset Management lacks efficient methods for real-time, comprehensive and high-granularity information of location, capacity and use of tangible and intangible assets. Asset management could benefit from more detailed, more accurate and longitudinal data on assets, providing more insight into efficiency and effectiveness on different levels of scale through time.

Existing technologies could provide a platform delivering those required insights. Navigation- and communication technologies such as GNSS, Wi-Fi, Bluetooth, RFID can be used to 'locate' users, estimate intensities and reveal patterns of movement and patterns of use. For Asset management indoor localisation is essential.

Technology

Wi-Fi is a widespread communication technology used by electronic devices to connect to a Wireless Local Area Network (WLAN) base station or to connect ad-hoc directly between devices. Wi-Fi may be used to obtain



Published in "Proceedings of the 13th International Conference on Location-Based Services", edited by Georg Gartner and Haosheng Huang, LBS 2016, 14-16 November 2016, Vienna, Austria.

internet access, to exchange data, to access an intranet or for sending data to devices like a printer. Today a large range of electronic devices is capable of using Wi-Fi including computers, laptops, smart-phones, tablets, digital cameras, audio players, printers, (video-)game consoles and sensors. Wi-Fi networks are offered in companies, private homes, cities and (semi) public spaces and also at university campuses. Eduroam is a worldwide standard for University Campus WLAN networks.

Wi-Fi cannot only be used as a technology to transfer digital data wirelessly, but also as a tool for Facility and Asset management or as a platform for location-based services (LBS): Wi-Fi Access Points (AP's) can be used as sensors to collect information of connecting devices, delivering dashboards with temporal data on intensity of devices based on the number of unique devices detected and patterns of movement based on detections of the same device at different access points.

Wi-Fi 'user' data can be obtained in two ways: (a) by scanners or (b) by the network.

ad. (a) Wi-Fi scanners register connection attempts from devices within range. Every enabled Wi-Fi device is continuously searching for Wi-Fi access points and therefore broadcasting its unique media access control (MAC) address. No connection between device and scanner is made and no data is exchanged. Scanners can be other Wi-Fi devices as well as (modified) Access Points.

ad. (b) When using the network Access Points the real -established-connections are used. Either the connection start- and endtime is logged, or the system regularly scans for connected devices. At TU Delft a dump of devices connected to Eduroam Access Points is made every five minutes for the whole campus for network management purposes. Personal information such as MAC address and network ID are immediately encrypted (hashed). The research projects described in this abstract use this anonymised data from the Eduroam network for spatio-temporal analysis. Only staff and students from the university connecting to the eduroam Wi-Fi network are incorporated in this research.

In the Geomatics Synthesis Project Wi-Fi is used in a campus-wide experiment to monitor flows and occupation patterns at the TU Delft Campus. Students worked for two months on three parallel projects :

- (1) extracting presence of people at specific places;
- (2) unravelling patterns of movement within buildings and between buildings on the campus; and
- (3) identifying activities and irregular use based on Wi-Fi data.

In all projects, the same dataset is used. All project also had to cover four cross-cutting topics: Privacy, Validity and Accuracy, Representativeness and reflection on the system of Access points (data collection).

Project 3: Track-id: Activity Determination based on Wi-Fi Monitoring

The aim of this project is to recognise the activity of different users in an area through Wi-Fi monitoring. First of all, the estimation of the user's occupation is calculated by the use of a Markov model with the information derived from the Wi-Fi dataset. Their identity is used in order to estimate the activity that a user is probably doing.

The main question of this research is:

To what extent and how reliable is it possible to determine the activities of individuals through the users' characteristics that can be derived from a Wi-Fi network?

This project focuses more on the use of the facilities of the research area during irregular hours. The irregular hours are specified as the hours outside the opening hours of the buildings, not including exceptions like extended opening hours or events. The use of the buildings is examined during irregular hours, to allow efficient real estate management and provide security. The Information Communication Technology (ICT) department of Delft University of Technology (TU Delft) provided a database dump of the Wi-Fi network of TU Delft. These data are analysed, in order to get useful and valuable information about the usage of buildings and the activities in buildings.

Dataset Limitations

The dataset has been provided by a third party source, hence it is not possible to select the tracking technologies and methodologies. In this chapter, the infrastructure that the third party utilised to gather along with the validation, the accuracy and the representativeness of the tracking system is discussed. Finally, an insight about the data protection that is relevant with the data and methodologies that are used, is presented.

Validity and accuracy

Due to the design of the tracking infrastructure, if a user is only passing by an access point and gets scanned, the user gets a minimum connection time of five minutes. Furthermore, the system is designed in a way that a device can only maintain a single connection at a time, thus, if a user moves inside the five-minute period of his last access, he will not be tracked to the new position.

Due to the nature of Wi-Fi tracking, a user that is detected in a certain access point, is not necessarily located around that access point. For example, a user might be passing outside a building and he might be scanned by an access point inside it when actually he is not there. Moreover, false location of scanned users can also happen between floors of a building.

Representativeness

The representativeness of the data collection reflects what categories of users can the implemented Wi-Fi tracking system identify, or what categories of users it cannot.

The first big category is the users that were not connected to the Wi-Fi network. It consists of smaller categories, a) users that use old technology, b) users that are not part of the academic network that has access to the Wi-Fi network, c) users who consciously turned off their Wi-Fi devices, d) users connected to the network through a wired connection.

The second category refers to places that there are no available data in the given database. The missing data accounts for, either buildings that are excluded for security reason, e.g. Nuclear reactor, or for places that are not covered by access points, hence no information is collected.

System of access points

With the implemented system of access points, every device is usually detected in only one access point. Thus, the processes of fingerprinting and trilateration, that would allow the determination of the position of a device, cannot be carried out. Therefore, it is assumed that the detected devices are located to the specific access points that they are detected.

Data protection

According to the Data Protection Directive (DPD) (Directive 95/46/EC) 'Personal data is any information relating to an identified or identifiable natural person'. The data collected from the Wi-Fi network are considered personal data, even if the username and MAC address of each user and device are hashed, hence the data must be treated accordingly. Currently, DPD secures the data privacy of individuals against unlawful use. According to DPD, to process private data, one has to have a valid ground to do so. In the current project, scientific purposes are sufficient reason, yet data must be processed with respect to the Principles mentioned in the DPD and with respect of the Data subject rights.

Research

In order to distinguish the activity of a user, an occupation profile is assigned to each user. According to their occupation profile, their main

building and the type of the building that they are located in, the activity of the specific session is determined. However, the main building of each user is determined as the building where that particular user spent his most time at.

The profile of the users is determined by using a Markov model. A Markov model is a stochastic model that is used to model randomly changing systems, where it is assumed that future states depend only on the current state and not on the events that occurred before it. The stated property is characterised as ‘memorylessness’ or Markov property. Generally, this assumption enables reasoning and computation with the model that would otherwise be intractable. There are different Markov models used in different situations.

According to the Markov model, some training sets are defined considering the different profiles that need to be assigned to the users (student, academic staff, support staff or other). Those training sets are compared with the user’s information that are derived from the dataset and the probability of each user to belong to each occupation profile is determined. Further, the user is assigned with an occupation randomly, based on the different probabilities (Petrushin 2000, Luhr et al. 2003, Mühlenbrock et al. 2004, Stamp 2015). The activities of each individual, during irregular hours, is determined through a deterministic model, which takes into consideration the assigned user occupation, the user main faculty and the type of building the user is located each time.

For determining the activity of each user, his main building, his occupation and his scanned location are taken into consideration. To make an accurate statement, independent of the occupation classification of a user, the total numbers of faculty staff and students are compared to the numbers of the dataset. The guests of the university have to be filtered out since official statistics do not contain guests of the university. After this procedure, the main building of a user is determined, as his main faculty, but only if he spent more than one hour there.

The results of the use of the TU Delft campus are visualised on different spatial levels and on different representations. The spatial levels that are used, are related to the campus, building and floor (“maploc”) levels. Tables and graphs, a dynamic visualisation and a GIS and Web application are created during this project.

Conclusions

The overall accuracy of the determination of the number of users per building is 94%. Regarding the determination of the users’ occupation using the Markov model the accuracy of the process is 50%. Considering the above results, the pre-process and the analyses conducted to detect distinct users

in a complex of buildings, are regarded of good quality and can be used further. On the other hand, the determination of user occupation is not accurate enough and further research is required. Additionally, the identification of specific events and exceptions on the opening hour of buildings can be identified by detecting irregularities of user connections. Finally, it is clear that through Wi-Fi tracking it is possible to extract information that will allow efficient real estate management and provide security solutions.

Acknowledgement

This project is part of the Geomatics Synthesis Project (GEO1101) 'Rhythm of the Campus' which was carried out in Spring 2016. The data is collected by ICT of TU Delft and made available for the project. In a later stage – if no obstructions are encountered – the data will be made open-source.

The course was initiated and coordinated by Stefan van der Spek, director Geomatics for the Built Environment and Edward Verbree, section GIS-technology. Edward Verbree, Martijn Meijers and Wilko Quak mentored the three project teams. Bart Valks and Iljoesja Berdowski represented Facility Management of TU Delft. Alexandra den Heijer and Ruud Binnenkamp (Management in the Built Environment) assisted to the overall scope of the project.

References

- DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 1995, Official Journal L 281 , 23/11/1995 P. 0031 - 0050
- Mautz R (2012) Indoor Positioning Technologies. ETH Zurich, Department of Civil, Environmental and Geomatic Engineering, Institute of Geodesy and Photogrammetry
- Mühlenbrock M, Brdiczka O, Snowdon D, Meunier JL (2004) Learning to Detect User Activity and Availability from a Variety of Sensor Data, Xerox Research Centre Europe, Meylan, France
- Petrushin V A (2000) Hidden Markov Models: Fundamentals and Applications Part 1: Markov Chains and Mixture Models
- Petrushin V A (2000) Hidden Markov Models: Fundamentals and Applications Part 2: Discrete and Continuous Hidden Markov Models
- Stamp M (2015) A Revealing Introduction to Hidden Markov Models