USING MULTIPLE SOURCES OF DATA FOR MONITORING FACILITY USAGE ON CAMPUS

P. WEPULANON [a], W. H. K. LAM [b] and A. SUMALEE [c]
[a,b,c] *Department of Civil and Environmental Engineering,*
The Hong Kong Polytechnic University, China
Email: w.piyanit@gmail.com

ABSTRACT

This paper aims to investigate the possibility of developing facility usage monitoring system on campus, using multiple sources of information. The data sources can be classified as primary and secondary data. The primary data is significant as it can be processed to derive activity-mobility information of the students, whereas the secondary data is valuable for the in-depth analysis of student behavior. In this paper, the Wi-Fi communication data broadcasted by students' portable electronic devices is considered as a primary data. To derive student mobility information and monitor facility usage, data processing algorithms are introduced. The system performance is evaluated using a case study on The Hong Kong Polytechnic University. The results show that the status of facility usage on the campus can be identified with acceptable accuracy. Moreover, the integration of multiple data sources can enhance the deliverable results of student behavior analysis.

Keywords: facility usage monitoring, Wi-Fi signature data, multiple data sources, activity-mobility analysis, smart application

1. INTRODUCTION

Facility usage monitoring is an essential process to improve the performance of facility management, development, and planning. In the city scale, an essential facility is the station of public transport. The facility usage information is significant so as to understand the passenger demands and evaluate the accessibility of public transport stations. In the same way, facility usage monitoring in the campus of university is a significant process to evaluate the actual student demands on campus facilities. The difficulty of facility development is how to provide and manage the facilities based on the actual needs of students within the limited campus areas. To derive facility usage information, researchers have been focused on analyzing pedestrian mobility information and estimating the pedestrian demands over time of the day.

The traditional methods to obtain pedestrian mobility information are interview surveys. Conducting surveys is costly and time-consuming. Survey methods provide a small proportion of mobility data collected from a group of total population. In addition, bias in mobility analysis can be occurred due to reporting errors caused by the interviewees. As a result, the mobility information cannot be used for monitoring facility usage in the long run, i.e. over months and/or years.

In the most recent years, the more promising data collection methods have been proposed to obtain mobility information of individual pedestrians. Smartphone devices are the major tool for collecting such information since the devices are integrated with various sensing technologies. Moreover, it can be assumed that people usually carry their smartphones.

To derive the activity-mobility information, pedestrian locations can be identified using the built-in GPS (Global Positioning System) of smartphones (Murakami and Wagner, 1999; Du and Aultman-Hall, 2007; Iqbal et al., 2013). The quality of activity-mobility information has been tremendously improved. However, the positioning system may involve significant errors in particular conditions such as urban areas, and severe weather conditions. Hence, GPS-based data collection methods are limited for indoor environments.

Considerable attention has been paid to the methodologies for deriving mobility information for indoor environments. Pedestrians' location can be estimated using smartphones' built-in sensors e.g. accelerometer and Wi-Fi function (Chon and Cha, 2011; Kang et al., 2012; Kim et al., 2012). To estimate pedestrian locations, the pedestrians have to install a smartphone application which must be always active during the localization process. Battery consumption issues should be concerned for the long-term collection of pedestrian m data.

In the case of a university campus, the variety of students' mobility poses more challenges in activity-mobility data collection. Apart from the mobility in horizontal dimensions, students usually perform their activities in buildings. Therefore, the location data in vertical dimensions is also necessary to be collected for student's mobility analysis. The methodologies to derive mobility information using Wi-Fi signatures on campus have been developed (Yoon et al., 2006; Danalet et al., 2014).

To this end, the main objective of this study is to introduce the available information from multiple data sources on campus, and to propose an automatic facility monitoring system which can provide the facility usage information. In this study, Wi-Fi data is the primary information for deriving mobility information. The proposed system is evaluated by a case study on The Hong Kong Polytechnic University (PolyU) campus. Furthermore, the possibilities of using multiple data sources to enhance the deliverable results of the system are discussed. The results of facility usage analysis can be used for sustainable campus development with taking account the actual usage demand by facility types.

This paper is organized as follows. Section 2 describes the general data sources in a university campus. Section 3 proposes the necessary data processing steps to extract facility usage information from a primary data source, follow by experimental results from a case study in Section 4. Finally, this paper is concluded and the further studies are discussed in Section 5.

2. ON-CAMPUS DATA

This section introduces the types of data sources in a university campus which can be considered as potential inputs for smart campus development. Figure 1 demonstrates a conceptual diagram of developing a smart campus using the available information. The on-campus data sources can be classified into 2 types: primary data sources and secondary data sources. The available information can be effectively processed to reveal the mobility information on the campus including student activity-mobility, and facility usage. Finally, the efficiency of campus development can be improved in the sustainable ways based on the actual student demands.

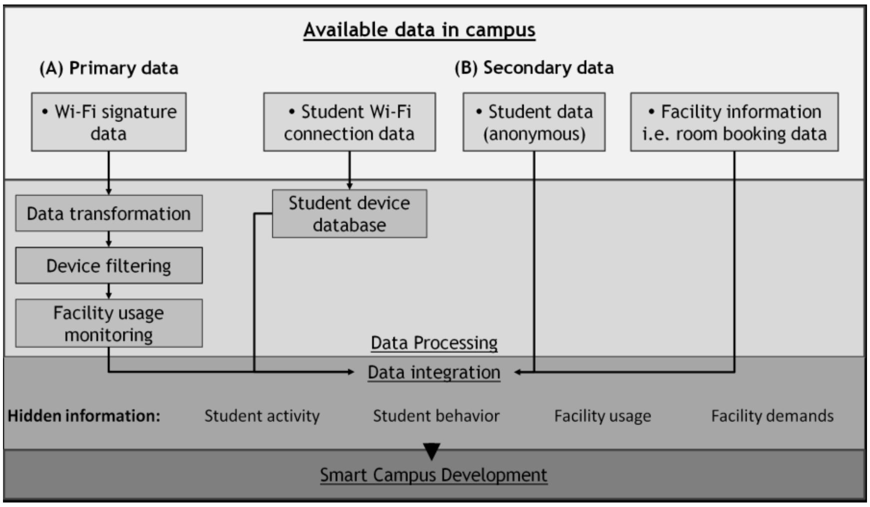

Figure 1. The conceptual diagram of using on-campus data for smart campus development

2.1 Primary data

The information, which has the potential to provide spatial-temporal mobility information of individual students on the campus, is identified as primary data. In this study, the information derived from Wi-Fi communication systems is considered as primary data. The student mobility information can be discovered by analyzing the sequences of Wi-Fi communication data broadcasted by students' portable devices.

A Wi-Fi enabled device such as a smartphone can be considered as an opportunistic sensor for tracking spatial-temporal mobility of the user, since the device usually broadcasts Wi-Fi communication messages when the Wi-Fi function is enabled. The broadcasted messages can be detected by Wi-Fi scanner devices. The scanner will continuously listen to the broadcasted messages within the scanner's detection area. A device will be detected by a scanner when the scanner can capture a message broadcasted by the device. The necessary information of the detected devices then will be recorded including (i) the Media Access Control (MAC) address which is the identification number of each Wi-Fi device, (ii) detection timestamp, and (iii) Received Signal Strength Indicator (RSSI) which can be used to estimate the distance between the detected device and the scanner. It should be noted that the recorded information cannot be used to reveal any personal information of the device's owner.

With the capability of the Wi-Fi scanner, students' devices can be detected without the necessity of campus Wi-Fi network connection. In the case of PolyU campus, the mobility data provided by the scanner is more reliable than the data from PolyU Wi-Fi system. This is because the PolyU Wi-Fi system can detect a device with the lower sampling frequency. In particular, a device will be detected only when it made a successful authentication for accessing PolyU Wi-Fi networks. Apart from the scanning capability, the Wi-Fi scanner can also be operated as an access point at the same time. Therefore, it is possible that the device will be operated as an access point in the near future.

2.2 Secondary data

The secondary data cannot be used for deriving student mobility information. However, the data could benefit the in-depth analysis of the student mobility-activity information obtained from the primary data. For example, the historical records of PolyU Wi-Fi connections can be considered as a secondary data source. In order to access to PolyU Wi-Fi services, authentication process is required. Students generally use their student ID as the username to access the services. Therefore, the username can be matched with the MAC address of student's devices and increase the significant of the mobility information of student's devices provided by Wi-Fi scanners.

It should be highlighted that privacy issues have been aware. The Information Technology Service (ITS) department always encrypts the device MAC address and username information before granting any information usage. As a result, PolyU Wi-Fi connection data cannot be used to identify private information of students (e.g. student name, address, and mobile number).

3. FACILITY USAGE MONITORING SYSTEM

This section introduces the necessary data processing steps for developing facility usage monitoring system using the primary data from Wi-Fi scanners. In particular, the algorithms for classroom occupancy monitoring system are introduced as an example of facility usage monitoring system. The data processing steps to obtain classroom occupancy status are described in the following sub-sections.

3.1 Data transformation

Supposed that a set of Wi-Fi scanners $S(S=\{s_i\}, i=1,2,...,n)$ are installed on the campus, the raw data derived from a scanner s_i can be denoted by $D_{s_i}\,(D_{s_i}=\{\vec{d}^{\,j}_{s_i}\}, s_i \in S, j=1,2,...)$ where $\vec{d}^{\,j}_{s_i}$ is a vector representing the information of the j^{th} Wi-Fi data detected by the scanner s_i. The information consists of timestamp, MAC address, and RSSI ($\vec{d}^{\,j}_{s_i}=\begin{bmatrix} t^j_{s_i} & mac^j_{s_i} & rssi^j_{s_i} \end{bmatrix}^T$).

To derive the spatial-temporal mobility of each device, the raw data needs to be transformed to the more proper data structure. For each scanner, the raw data D_{s_i} can be sorted by MAC addresses and its detection timestamp. Activity time duration of a device at the scanner's location can be obtained, by considering the continuity of detection timestamps of the device. The data transformation algorithm can be illustrated as the pseudo-code below.

Algorithm1: Data transformation

```
for each MAC address
    - set lastObservationTime = time of the first record
    for first to last record
        if record's timestamp – lastObservationTime < 1 hour then
            - update device information (exit time, average RSSI, detection frequency)
            - set lastObservationTime = timestamp of the current record
        else
            - record a new data structure
        end if
    next record
next MAC address
```

As a result, the raw data set D_{s_i} can be transformed to the new data structure $DV_{s_i}\,(DV_{s_i}=\{\overrightarrow{dv}^{\,j}_{s_i}\}, s_i \in S, j=1,2,...)$ where $\overrightarrow{dv}^{\,j}_{s_i}$ is a vector representing the information of a detected device during a time period including MAC address, detection date, first detection time, last detection time, average RSSI, and detection frequency ($\overrightarrow{dv}^{\,j}_{s_i}=[mac^j_{s_i} \quad dt^j_{s_i} \quad t\_in^j_{s_i} \quad t\_out^j_{s_i} \quad rssi^j_{s_i} \quad freq^j_{s_i}]^T$).

3.2 MAC address filtering

Since the Wi-Fi scanner is able to capture the Wi-Fi enabled devices within the detection area, noise data is generally included in the raw data sets. To develop a facility usage monitoring system, MAC address filtering process is essential to retain only the potential devices carried by the facility's users. In the case of classroom occupancy monitoring system, three types of devices should be filtered out.

First, the MAC addresses which are usually detected by a scanner for considerable time periods over a day can be identified as background MAC addresses. Basically, the background MAC addresses belong to stationary devices nearby the scanner such as access points and the Wi-Fi devices in offices. Musa and Eriksson (2012) described a simple heuristic filtering method by considering the total presence time duration and the time since the last observation of each MAC address.

The second type of noise data is from the passerby's devices. A device will be filtered out when the activity time duration is lower than a duration threshold $TH_{Duration}$. The threshold value can be pre-defined by implementing data training process.

$$t\_out.dv^j_{s_i} - t\_in.dv^j_{s_i} \leq TH_{Duration} \qquad (1)$$

Finally, RSSI filtering is required as the detection area of Wi-Fi scanner may exceed the facility area. RSSI filtering aims to filter the device which is located outside the area of interest. For example, a device should be filtered out, if it was carried by a student who is occupying other classrooms nearby. Supposed that a Wi-Fi scanner is available for each classroom, the location of a device at particular time duration can be assumed to be the same location as the Wi-Fi scanner s_i which provides the maximum RSSI value. Given a device MAC address mac, the maximum RSSI max_rssi of the device at particular time duration can be identified:

$$max\_rssi(mac) = \max \sum_{i=1}^{N} rssi_{s_i}(mac) \; ; s_i \in S \tag{2}$$

where N is the total number of Wi-Fi scanners. Moreover, the minimum RSSI value TH_{RSSI} should be pre-defined to avoid significant localization errors.

$$max\_rssi(mac) \geq TH_{RSSI} \tag{3}$$

The remaining devices after implementing MAC address filtering can be considered as the potential devices carried by the facility's users, and can be used to determine the status of facility usage.

4. A CASE STUDY ON POLYU CAMPUS

An experiment was conducted on PolyU campus during the weekdays from 14 March 2016 to 1 April 2016. A classroom area on the 4/F of Block Z building is selected to evaluate the performance of the facility usage monitoring system. Figure 2 demonstrates the plan of the classroom area. There are 9 classrooms and 10 Wi-Fi scanners installed in the classroom area.

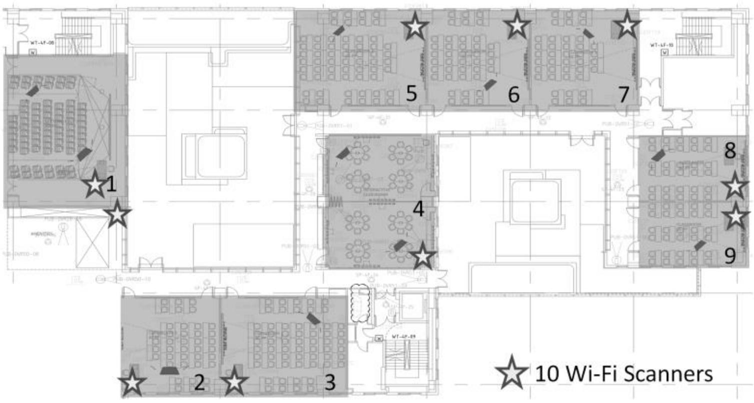

Figure 2. The classroom plan with Wi-Fi scanner locations

In this study, the Wi-Fi data provided by the scanners is processed to identify the classroom occupancy status: available or occupied. With the availability of Wi-Fi scanner data over time-of-day and day-of-week, the classroom occupancy information can be identified regardless of monitoring time duration.

4.1 Room occupancy monitoring accuracy

The raw data derived from Wi-Fi scanners was processed by implementing data transformation and MAC address filtering algorithms. The duration threshold $TH_{Duration}$ and RSSI threshold TH_{RSSI} can be determined by performing data training process using 20% of the available data sets.

To evaluate the system performance, the number of people occupying the classrooms was manually observed for once an hour during the experiment. Based on 246 observations, the proposed system performs 86.2% accuracy with 7.7% of false negative and 6.1% of false positive errors. Figure 3 shows the number of detected devices in the 9 classrooms during 8:00-22:00 of 21st March 2016 (Monday).

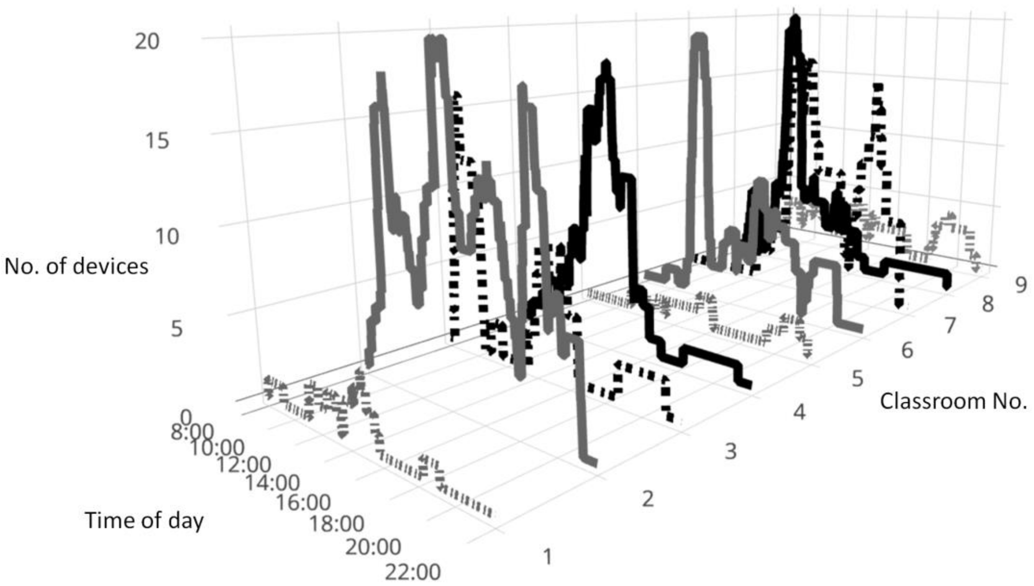

Figure 3. The plot of detected devices during a day

First, it can be observed that the number of detected devices is varied from 0 to 20 devices approximately. In fact, a classroom can be occupied by several purposes including lecture, tutorial, and personal use. Students may randomly occupy the classroom without any booking. According to observation data, the classrooms were usually occupied for personal use. For example, a group of students may prefer to occupy a classroom to perform group activities. It is also found that a classroom was frequently occupied by a student since the student wants to use the personal computer in the room. Second, the classrooms had different usage demands. For instance, the classroom No. 1 was rarely occupied during the day while the classroom No.2 was occupied for most of the time. The results of classroom occupancy usage can be used to evaluate the efficiency of classroom facility. The considerable demands of some classrooms can be distributed to the available classrooms. The efficiency of facility management then can be improved.

Figure 4 shows a plot of room occupancy status of the classroom number 8 during 5 weekdays. The classroom size is 30 seats. Based on observation data, the maximum number of room occupants is 23. It can be seen that the classroom usage can be monitored by considering the number of detected devices in the classroom. As can be seen, the number of room occupants is varied from 1 to 20 which correspond to the observed data.

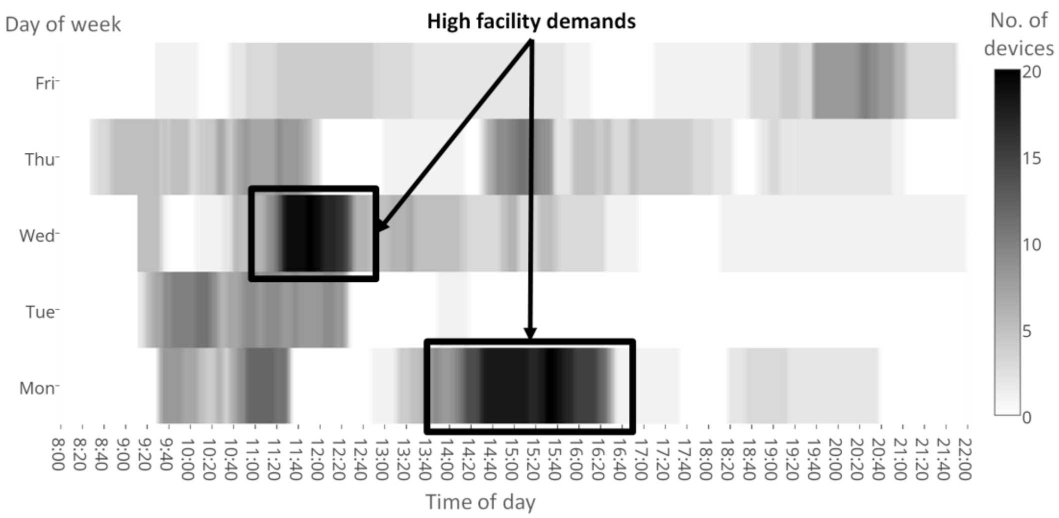

Figure 4. The number of detected devices in a classroom during 5 weekdays

4.2 Repeated behavior

A student may visit the same location at different time periods. The re-visit activity can be considered as repeated behavior of the student when the re-visit is occurred in the similar time period. Table 1 shows the examples of devices which were detected in the same weekday (Thursday) in the classroom No. 8.

Table 1. Examples of re-visit and repeated behavior

Device No.	MAC address	Date	Time in	Time out
1	38:AA:3C:93:99:50	17 March 2016	**10:51:09**	**11:29:53**
1	38:AA:3C:93:99:50	31 March 2016	**10:39:38**	**11:33:24**
2	38:D4:0B:D0:FD:EA	17 March 2016	**10:40:05**	**11:32:08**
2	38:D4:0B:D0:FD:EA	31 March 2016	**10:32:55**	**11:32:44**
3	3C:15:C2:37:AC:86	17 March 2016	**10:35:03**	**11:31:09**
3	3C:15:C2:37:AC:86	31 March 2016	**10:33:16**	**11:30:06**
4	D0:22:BE:29:12:84	17 March 2016	16:28:02	17:32:32
4	D0:22:BE:29:12:84	31 March 2016	12:47:08	15:12:33

It can be assumed that device No. 1, 2, and 3 belong to the students who attended a class during 10:30-11:30. According to classroom booking information, the classroom was reserved for a lecture (CSE20359) during 10:30-11:30. The device no. 4 could be an example to illustrate the classroom occupancy for personal use since the two occupancy time periods as well as the occupancy duration are totally different.

4.3 Integration of multiple data sources

Secondary data sources can be utilized to enhance the significance in deliverable results of the system. First, PolyU Wi-Fi connection records can be used for refining the number of room occupants. The number of detected devices cannot represent the precise number of room occupants. The results can be underestimated when the users disable Wi-Fi function on their devices. On the other hand, it can be overestimated since a student may carry multiple portable devices such as smartphones and tablets. The availability of username and MAC address from PolyU Wi-Fi connection data has the potential to compromise such overestimation. Second, the room occupancy status can be compared with room booking information to analyze the performance of room booking system. For example, unofficial occupancy can be identified when students occupied a classroom without any reservation; in the same way as false reservations when a classroom was reserved but actually unoccupied. Finally, study behavior can be analyzed using anonymous student information. For instance, the variable affected study performance can be identified i.e. class attendance, punctuality, study program, etc.

5. CONCLUSION

This paper introduces the concept of facility usage monitoring system using multiple sources of data in a university campus. With the availability of the primary data source derived from Wi-Fi scanners, student activity-mobility information can be obtained. The necessary data processing steps for developing facility usage monitoring system based on Wi-Fi scanner data are proposed. The system performance is evaluated using a case study on PolyU campus. The results show that the system can monitor the classroom occupancy status with the acceptable accuracy. The possibility of integrating secondary data sources is also introduced to enhance the deliverable results of the system. The proposed framework can be applied to other facilities in urban areas e.g. subway stations and shopping malls.

In the future research, the algorithms for location optimization problems can be developed to reduce the necessity of Wi-Fi scanner installation since installing a Wi-Fi scanner for each facility may not be feasible. Consequently, the filtering algorithms should be developed by considering the instability of RSSI values when the number of Wi-Fi scanners is optimized. Moreover, the algorithms to estimate the number of facility users can be improved.

6. ACKNOWLEDGEMENT

The authors would like to thank the Information Technology Services Office for the provision of PolyU Wi-Fi connection data, as well as technical suggestions for the long-term development. We also would like to thank the Campus Development Office for the valuable discussions on smart campus development. The work described in this paper was jointly supported by research grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. PolyU 152057/15E) and the Research Committee of The Hong Kong Polytechnic University (Project No. 1-ZVFJ).

7. REFERENCES

Murakami, E. and Wagner, D.P. (1999) Can using global positioning system (GPS) improve trip reporting?. Transportation Research Part C: Emerging Technologies, 7(2), pp. 149-165.

Du, J. and Aultman-Hall, L. (2007) Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues. Transportation Research Part A: Policy and Practice, 41(3), pp. 220-232.

Iqbal, S., Siddique, A. B., Islam, M., and Choudhury, C. F. (2013) Predicting tour patterns derived from ubiquitous data sources. Presented at the Transportation Research Board 2013 Annual Meeting.

Chon, J. and Cha, H. (2011) Lifemap: A smartphone-based context provider for location-based services. IEEE Pervasive Computing, (2), pp. 58-67.

Kang, W., Nam, S., Han, Y., and Lee, S. (2012) Improved heading estimation for smartphone-based indoor positioning systems. *In Proceedings of IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC)*, pp. 2449-2453.

Kim, Y., Shin, H., and Cha, H. (2012) Smartphone-based wi-fi pedestrian-tracking system tolerating the rss variance problem. *In Proceedings of 2012 IEEE International Conference Pervasive Computing and Communications*. pp. 11-19.

Yoon, J., Noble, B.D., Liu, M., and Kim, M. (2006) Building realistic mobility models from coarse-grained traces. *In Proceedings of the 4th international conference on Mobile systems, applications and services*, pp. 177-190.

Danalet, A., Farooq, B., and Bierlaire, M. (2014) A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. Transportation Research Part C: Emerging Technologies, 44, pp. 146-170.

Abedi, N., Bhaskar, A., and Chung, E. (2013) Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement, *In Proceedings of the 36th Australasian Transport Research Forum*.

Musa A. and Eriksson, J. (2012) Tracking Unmodified Smartphones Using Wi-Fi Monitors, *In Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pp.281-294.