Understanding occupancy and user behaviour through Wi-Fi based indoor positioning

Article

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Understanding occupancy and user behaviour through Wi-Fi based indoor positioning

Abstract

A 30-day monitoring campaign was conducted in a university library building to investigate the usefulness of a novel Wi-Fi based indoor location system for revealing indoor occupancy patterns and related user behaviour. The system has demonstrated its effectiveness in providing occupancy information with a relatively high degree of granularity and accuracy in this study. The occupancy results revealed that the 24-hour opening policy for the library during the term time was not necessary. On the other hand, the 8-hour library-opening duration during the summer vacation could be extended to include the early evening hours to benefit user productivity. Four occupancy patterns were identified based on cluster analysis. Most users were found to belong to the short-occupancy one-time visitor type, while a minority were the long-occupancy users. The cross-correlations between various occupancy parameters were investigated. For example, the pattern of user arrival times at the library was found to be significantly correlated with their study durations. Further data analysis showed that the majority of long-occupancy users tended not to have frequent breaks with some taking no break for 4 hours. This could have implications for their health and wellbeing as well as their productivity.

Key words: Occupancy, building user behaviour, Wi-Fi based indoor positioning, data mining.

1. Introduction

Building occupancy is a critically important piece of information for building design, operation, maintenance, management as well as for research into building energy efficiency, indoor air or environmental quality (IAQ or IEQ), health and wellbeing in buildings, and effectiveness of building utilization. For example, dwellings with daytime occupancy (e.g. elderly or infirm) could suffer twice as much overheating exposure in summer, compared with the same dwelling occupied by people who work or study during the day and return during the evening (Hallett et al., 2013; Porritt, Cropper, Shao & Goodier, 2012). Exposure to indoor environmental conditions or hazards e.g. excessive CO_2, radon, or formaldehyde is substantially related to the exposure duration as well as the level or intensity of hazards (Gilbert, 2005 and COMEAP, 2004). Duration of human exposure to indoor air pollutants is also a key parameter to relevant exposure models (Rosenbaum et al., 2015).

In addition, there has been increasing realisation (Primo, 2015) that modern organisations and companies, especially those engaged in knowledge-based activities and enterprises, will benefit from having access to a variety of types of work spaces catering for diverse types of work and occupancy, beyond the basic room types, e.g. those for formal meetings or solitary work. One practice-based study has outlined 10 different workspace types that meet the requirements of different scenarios of individual work, informal discussions and group collaborations (Primo, 2015). Information on

2

occupancy patterns will be valuable for identifying and optimising such workspaces to best meet users' needs.

Detailed occupancy data including occupant numbers and arrival /departure times are key input parameters for building dynamic thermal simulation and occupancy pattern prediction (Mckenna, Krawczynski & Thomson, 2015; Dar, Georges, Sartori, & Novakovic, 2015; Gul & Patidar, 2015; Ortega, Han, Whittacker, & Bowring, 2015), where the occupancy data quality, e.g. accuracy, granularity and depth of insights into occupancy patterns, impacts significantly on the quality of the modelling output. Occupancy information is also important for building simulation tools for IEQ and IAQ assessment (Duarte, Wymelenberg & Rieger, 2013), and significant deviation of prediction to ground truth can occur when fixed occupancy profile assumptions are used during simulation to represent highly variable or stochastic occupancy scenarios (Chang & Hong, 2013). In addition, the impact of occupancy on energy demand reduction has been widely recognised (Wang & Shao, 2017) and detailed occupancy monitoring has brought significant advancement to the quantitative study of building space usage (Spataru, Gillott, & Hall, 2010).

The presented paper identifies and quantifies occupancy patterns, centring on a university library, where a 30-day consistent monitoring campaign was conducted to investigate the potential of the advanced Accuware Wi-Fi based Indoor Triangulation System (ITS). The results were analysed to reveal occupancy patterns that are important to building operation and user

wellbeing and productivity. Based on the collected data, authors have identified and quantified the occupancy duration patterns, associated user behaviour, e.g. the regularity of users taking breaks between consecutive long-occupancy periods, and the popularity of a space or space utilisation by members of the library.

2. Occupancy detection technologies

Occupancy data collection traditionally relies on surveys and questionnaires, e.g. during the POE of newly completed buildings. The time use survey from the Office of National Statistics (ONS) is widely considered as an important resource, particularly for understanding occupancy within domestic households (Widén, Molin, & Ellegård, 2012). However, potential inaccuracy in these occupancy data could be significant, as the data collection methods are based on memorised estimation. Sensor based occupancy investigation is gaining popularity, given that a wide range of technologies have become available in recent years and decades, as outlined in the following. The most frequently used sensor types for monitoring building occupancy include Passive Infra-Red (PIR) motion detectors and CO_2 detectors, which are widely available commercially, cost-effective and non-obtrusive. However, false negative outputs are generally expected from PIRs when monitoring stationary occupants. Improvements of this technology including optimized time-delay have been proven to be effective in dealing with immobile building users, but the PIR is generally still incapable of occupant counting without appropriate advanced algorithms (Guan, Li, Guo & Wang, 2014). Occupancy detection based on

CO$_2$ sensors relies on indoor CO$_2$ concentration measurement, so the accuracy is influenced by a plethora of factors including occupants' activity type, human metabolic rate, ventilation rate, air leakage through the building envelope, and the opening status of doors and windows (Gunay, Fuller, O'Brien & Beausoleil-Morrison, 2016). Improvement in measures to address these aspects through data mining and machine learning only work sufficiently well when the above data are available. A further development in recent years is the application of probabilistic tools e.g. Artificial Neural Network and Decision Tree Model to generate occupancy profiles, based on the data obtained from a multitude of sensors e.g. temperature, humidity, sound, CO$_2$ and PIRs. These sensors complement each other to help improve occupancy detection accuracy (Yang & Becerik-Gerber, 2014; Yang, Li, Becerik-Gerber & Orosz, 2014; Dong & Andrews, 2009; Han, Gao & Fan, 2012).

Moreover, wearable sensors like SenseCam (Gauthier, & Shipworth, 2015) have been used in research to assess the real-time occupancy achieving a reasonable level of accuracy. A distinct advantage of this approach is in providing specific occupancy information of time and location, but privacy issues can arise when it is applied on a large scale, although this could be minimised in the future with the development of privacy protection technologies. Occupancy information can be also inferred based on electricity consumption data, e.g. through existing smart meters and plugs, but naturally the algorithms for estimating occupancy involve varying degrees of approximation or assumptions (Akbar, Nati, Carrez & Moessner,

2015; Kim & Srebric, 2015; Albert & Rajagopal, 2013). In contrast, the Ultra-wideband (UWB) indoor positioning technology has an excellent level of accuracy but suffers from exceptionally high costs (Spataru & Gillott, 2011). A lower-cost alternative is the indoor occupancy detection based on analysis of visible and infrared videos of indoor spaces. Although the accuracy is often good - up to 90-95% - but there are significant privacy concerns when it is used to track occupants. Furthermore, the method based on video image analysis could not detect occupants behind obstacles like desks or workstation partitions (Liu, 2015). Recent development in low-power Bluetooth technology offers lower-cost indoor location detection with room-level accuracy. However, Bluetooth functions are often switched off in personal devices, thus rendering the technology unusable for many situations (Subhan, Hasbullah, Rozyyev & Bakhsh, 2011).

With the popularization of Wi-Fi network and Wi-Fi enabled smart devices, Wi-Fi signals or access points have been utilized to detect occupancy. Martani et al. (2012) inferred the occupancy level through measuring the number of Wi-Fi connections registered with specific Wi-Fi access points. Device-free Location (DfL) such as E-eyes (Wang et al., 2014) has been shown to be able to detect occupant activity such as walking. This technology is based on equipment with embedded Wi-Fi functionality e.g. smart televisions, refrigerators, thermostats and has demonstrated good accuracy in an unobtrusive and low-cost way, but detecting multiple occupants is still under exploration. Advances in Wi-Fi based indoor positioning technology has enabled real-time sensing of occupancy levels

6

and positions. The technology benefits from wide availability of Wi-Fi devices including mobile phones, the simplicity, relatively lower costs of the sensors and system set-up (Accuware, 2016). This technique is scalable, applicable to buildings of any size as well as for multi-storey buildings. The existing Wi-Fi network infrastructure in a building could serve as nodes of the system although some adjustments, e.g., adjusting the locations of access points would be needed. The privacy protection afforded by the Wi-Fi based occupancy detection has been significantly enhanced through approaches such as the Media Access Control (MAC) address truncation method described in Section 3, making it more practically suitable for a range of applications, including fixed-period occupancy study in public buildings. The Wi-Fi indoor location technology has thus been adopted for this research.

3. Experiment

The indoor occupancy study was carried out over a period of 30 consecutive days in the library of University of Reading (UOR), UK. The Knowledge Exchange Room as shown in Figure 1 was selected as the study area. The library was chosen because it is right at the heart of the UOR campus featuring a range of academic and information technology services and a very popular café. Thus it provides a highly dynamic range of visitor volumes at different times of the day, ideal for testing the usefulness of this innovative occupancy monitoring technique. The Knowledge Exchange Room is quite close to the library main entrance and the café and served by the password-secured campus Wi-Fi network. Again, these features will help to ensure a

dynamic flow of visitors. The public nature of the space also makes it more suitable for the study than a more private office space for issues relating to privacy, which will be discussed below. Table 1 shows the opening hours of the library during test period. To simplify data analysis, the room boundaries are approximated by a series of straight lines L1 - L8 in Figure 2, resulting in a small (1.7%) reduction of the floor area. An indoor positioning system comprising Wi-Fi sensors and an Accuware location server was used to determine the location of Wi-Fi enabled devices, including smartphones, laptops and tablets. The spacing between sensors should be within 12 m for greater accuracy (Accuware, 2016). In this experiment, 6 sensors (green dots in Figure 2) were deployed with spatial separations of 5 - 10 m, further ensuring the accuracy.

Insert Figure 1 here

Insert Table 1 here

Insert Figure 2 here

Privacy is often an issue of concern in the domain of occupancy detection (Demir, Cunche & Lauradoux, 2014). In this work, several approaches have been adopted to address this issue. Firstly, authors did not collect any direct information about the identities of the library users. The sensor data yields only latitudinal and longitudinal values of Wi-Fi devices, their MAC addresses and the test time. Secondly, the access to the raw data is protected through the user ID and secure password for only authorized persons on the official website. Moreover, the MAC address truncation

8

method was adopted in the data processing, where only the second half (the last 6 digits) of each of the 12-digit MAC address were kept in the data processing and analysis. This approach makes it about a billion times more difficult to identify positively an MAC address found in the test. In addition, this truncation method can be changed to select any six digits in an MAC address or indeed other number of digits (e.g., 4, 5, or 7 digits), adding another layer of protection for the privacy of the library users.

4. Calculation

The data collected by the Wi-Fi detection nodes is processed by a triangulation algorithm (Accuware, 2016) to determine the locations of detected Wi-Fi devices. Two basic questions then need to be dealt with. The first question is whether a detected device is within the study area, and this is addressed by using the function of 'IN = inpolygon (X,Y,xv,yv)' in Matlab. Secondly, it is necessary to determine the duration when a detected device is inside the studied space. A Visual Basic programme was developed by the authors for this purpose (Wang, 2016). A more detailed occupancy-pattern analysis was carried out using open source software R.

The data for each MAC address in any one test day is sorted using algorithms and presented as a vector consisting of the first occupancy duration (OT_1) of a day, the first absence duration (AT_1) and then the following OTs and ATs until the last detected status of the Wi-Fi device in that day, shown as vector (1). The total number of times a person is absent from the room is given by equation (2). Note that the last detected status of

a Wi-Fi device in one test day must be inside the room, so the last instance in the vector (1) is OT_n.

$$D_M = (OT_1, AT_1, OT_2, AT_2, \ldots, OT_{n-1}, AT_{n-1}, OT_n) \qquad (1)$$

$$RT_M = n - 1 \qquad (2)$$

where M refers to the order number of devices; n refers to the status number for one device; D_M refers to the duration vector; OT_n refers to the duration of the n th period of occupancy; AT_n refers to the duration of the n th period of absence from the room; and RT_M refers to the total absence times for one device in a test day. Note that different vectors of D_M can belong to the same device but on different test dates.

Furthermore, cluster analysis technique is applied to the whole dataset to identify occupancy patterns using the open source software R. Cluster analysis is one data mining method to group items with higher similarity into the same cluster. The clustering work involves choosing a method to measure the similarity, followed by an appropriate algorithm and a performance evaluator for cluster models. The similarity can be measured by Euclidean distance, Chebyshev Distance, Manhattan distance etc. Euclidean distance, which is commonly adopted due to its common application and simplicity, is used for this study and calculated based on Equation (3) below:

$$d\left(D_i - D_j\right) = d\left(D_j - D_i\right) = \sqrt{(OT_{i1} - OT_{j1})^2 + (AT_{i1} - AT_{j1})^2 + \cdots + (OT_{in} - OT_{jn})^2} \quad (3)$$

where i and j refer to two different vectors; $d\left(D_i - D_j\right)$ refers to the distance between these two vectors; OT_{in} and OT_{jn} refer to the n th occupancy and absence duration in the i th and j th vector; same for the AT_{in} and AT_{jn}.

The algorithm of K-means, a method of vector quantization, is chosen due to its simple and effective nature and it works by partitioning recorded vectors with nearer distances into a certain number (K) of clusters (Hartigan, & Wong, 1979). K is decided according to individual requirements, and in the context of this work, it is appropriate to choose K from 2 to 8. Silhouette Value is to evaluate the performance of the number of clusters by measuring the similarity of vectors to their belonging cluster in comparison to other clusters. That is, it is used to examine how appropriately data has been clustered and then the best cluster number of K is chosen. This value ranges from -1 to 1, and a higher value indicates a better clustering performance (Rousseeuw, 1987).

Before presenting the results of occupancy patterns, it should be reported that the indoor location system demonstrates accuracy levels consistently up to 0.5-meter in a series of measurements. Additionally, to investigate the correlation between the occupancy number (ON) and the Wi-Fi device number (DN), a series of sampling tests on 14 separate days with a 1-hour test period on each day were carried out, as detailed in Table 2.

Insert Table 2 here

In every sampling session, authors sampled multiple times and calculated the DN/ON ratio for each sample. Specifically, the ON was counted

manually by authors who acted as an ordinary occupant without disturbing others, while the DN was estimated based on data from the ITS system and algorithms described above. Then, the average of DN/ON ratio for all samplings is calculated according to equation (4).

$$\left(\frac{DN}{ON}\right) = \sum_{1}^{n} \frac{(DN)_i}{(ON)_i} \tag{4}$$

where $\left(\frac{DN}{ON}\right)$ is the average DN/ON ratio; $(DN)_i$ is device number in i th sample; $(ON)_i$ is occupant number in i th sample; n is the overall sample number; i is between 1 and n.

This value of this ratio reveals the extent of differences between DN and ON. The usefulness of the data generated from the tests does not strictly depend on DN / ON ratio of 1. So long as one is aware of the disparity between DN and ON and thus be able to account for its effect, the occupancy information from the tests would still be meaningful in many situations. In this investigation, the average DN / ON ratio is calculated as 1.16, with information about the range of values shown in Figure 3. This level of disparity between DN and ON values is considered acceptable for extracting useful information reported below.

Insert Figure 3 here

5. Results and discussions

5.1 The occupancy patterns on representative days

Insert Figure 4 here

Figure 4 shows the number of occupants throughout a 24-hour opening period on 6th June, a summer term day. The 'Knowledge Exchange Room' was continuously occupied on this day except for a few hours in the early morning. Figure 4 also reveals three peaks of room use in terms of occupancy number. The primary peak was from approximately 3 pm to 10 pm. A moderate peak was found between 9 am and 12 am, while a mini peak at about 1am. This primary peak is line with authors' daily observation that students tended to come to the library during the period from the mid-afternoon to the late evening, while the mini peak around 1 am is a reason for the University to adopt a 24-hour library opening policy during term time. However, as the occupancy pattern shows, the occupancy dropped to practically zero from 3 am to 9 am, indicating that an 18-hour opening period would probably be sufficient, while saving substantial energy consumption and staff time/cost. Similar patterns were found in other summer term days.

Insert Figure 5 here

Figure 5 shows the number of occupants throughout an 8.5-hour opening period on 15th June, a summer vacation day. Compared to Figure 4, the occupancy number was distributed more uniformly. It is noticeable that the occupancy number did not grow very significantly until the late morning. One important observation is that a large number of library users were still in the room even a few minutes before the library was closed at 5pm. This occupancy pattern suggests that these occupants still wanted to continue study but had to leave because of library closure. Thus, it would have been

13

beneficial to user productivity if the vacation library-opening hours had been extended to include the early evening period or the early morning opening period had been shifted to early evenings. In addition to the overall occupancy numbers reported here, later sections will show that the occupancy durations of users arriving at different times of the day are also different, probably indicating different types of study activities.

5.2 Identifying room use /occupancy patterns

Insert Figure 6 here

Figure 6 shows the Silhouette Values when the K-means method is applied to the occupancy vectors. It can be seen when the K = 4, the Silhouette Value is the highest, meaning the best clustering performance is achieved when all the vectors are classified as 4 clusters. Presenting the centroids (namely the mean values) of each partitioned clusters, the conventional method, may mask the distinct characteristics of the found patterns, given the defined vectors may have different dimensions. E.g., averaging the vector of D1 that has 2 recorded OTs with 1 absence period and D11 that has 5 recorded OTs with 4 absence periods makes little sense. In this study, the distribution of OT and AT durations is used to characterise OT and AT patterns for each common vector type.

Four distinctive occupancy patterns are identified from the entire dataset, illustrated in Figure 7. In Figure 7, the person silhouette refers to a general room user. Take Pattern A for example, the first person silhouette shows he/she walks into the room, and the second person silhouette means he/she

14

leaves the room. For Pattern B, the wider separation between the two person silhouettes indicates a longer occupancy than in Pattern A. Pattern C indicates a succession of short occupancy periods by the same person, and D a succession of long occupancy periods.

Insert Figure 7 here

Description of Pattern A-D is as follows:

Pattern A can be named as '**observers**', indicating those who showed up in the room only once in a day, but staying for a quite short period.

Pattern B can be named as '**intensive learners**', indicating those who stayed inside the room for a long period but only once in a day.

Pattern C can be named as '**inspectors**', indicating those who returned once or several times, staying for a short period followed by a long period of absence. Members of security staff may exhibit such a pattern.

Pattern D can be named as '**normal learners**', indicating those who needed to go outside for a break to rest or other reasons before another long period of occupancy.

5.3 Occupancy durations and arrival time

Insert Table 3 here

The distribution of OT and AT for each of 4 patterns is summarised in Table 3. These patterns are based on OT and AT durations of Wi-Fi devices recorded within one day as described in Section 4. They are different from the patterns of repeat or once only visiting during the whole experiment

15

period discussed later. For Pattern A, the most popular length of time a user stayed in the room was less than a minute. In contrast, for Pattern B, the main occupancy duration was between 30 and 60 minutes. Similar situations could be found for Pattern C where the peak occupancy duration was less than 1 minute and for Pattern D where the main duration was between 30 and 60 minutes.

Insert Figure 8 here

Figure 8 shows the daily distribution of 4 different patterns in terms of device numbers during the test period. All tested dates are named according to the corresponding day and week number, e.g. Fri 1 for Friday in the 1st week. The first impression is the total number of devices dropped markedly from the term to the vacation (since Mon 4 in Figure 8), partially because the library opened 24 hours during term time but 8.5 hours in the vacation. Another feature is that Pattern A, 'observer', was the dominating type across the whole period, given the associated device number was significantly larger than those of other 3 patterns. Combining the occupancy duration information about Pattern A, it is clear that more people went to the room and only stayed for less than a minute. It is also interesting to find that the average daily amount of 'observers' fluctuated around 55 in the term, with 2-3 'observers' showing up in the room in every library opening hour, while for the vacation period, this figure was approximately 20, but the hourly average of 2-3 'observers' remained roughly same.

As indicated above, the total amount of 'observers' and 'inspectors', the visitors with short-occupancy duration, was far greater than that of 'intensive learners' and 'normal learners', the visitors with long-occupancy duration. While this is permitted, this occupancy pattern is not expected, as the Knowledge Exchange Room is designed for informal small-group meetings and bilateral discussions that usually would last much longer. This occupancy information could help the improved space design in future. E.g., this Knowledge Exchange Room could be moved away from the busy ground floor or additional non-meeting seating areas could be created next to spaces similar to the Knowledge Exchange Room to cater for many building users who have apparently been seeking a different type of space. It is envisaged if data had also been collected from other rooms and spaces from the same library, a more comprehensive overview of visitors' choice of different library areas could be available to facilitate library space optimisation and management, leading to greater productivity. Occupancy patterns and insights gained will be valuable for identifying and fine tuning such work spaces to meet best users' needs.

Insert Figure 9 here

As discussed above, Pattern A and C can be grouped as 'short-occupancy visitors', while Pattern B and D are 'long-occupancy visitors'. Figure 9 shows the relative frequency distribution of arrival time for these two type users, revealing remarkably different features in these two periods. During the term time, short-occupancy visits arrival peaked around the late morning and

lunchtime hours, compared to long-occupancy visits that peaked around 2-4 pm in the afternoon. This reports an important observation about time use pattern of students. In the vacation, the short-occupancy visits showed a more uniform distribution of arrival time, while the peak in long-occupancy visits was more pronounced in general. Additionally, there were several 'midnight oil' type users who used the room beyond midnight among both groups.

Insert Figure 10 here

Insert Figure 11 here

Further analysis addresses Pattern B and D users and short-occupancy types of 'Pattern A and C' are removed from these analyses. Figure 10 shows the distribution of total occupancy durations for each test day and Figure 11 shows the distribution of duration of individual occupancy periods for each test day. E.g., if a Wi-Fi device has two occupancy periods in one day, then the total of these two period durations is presented in Figure 10 and these two durations of occupancy periods are presented in Figure 11. As can be seen during term time, there was a greater number of long-occupancy visitors than in vacation. These visitors also stayed longer during term time: on three days during the term time, there were visitors who stayed for over 8 hours in total, but none did so during vacation; on almost every day during term time, there were a significant amount of visitors who stayed for 4-8 hours, but during the vacation, not so many such visitors were found daily. However, there was a broad similarity between the term time and the

vacation in terms of the general trends of the distributions of total occupancy durations. Given curtailed opening time in vacation reduced returning possibility (the library opened 24 hours and 8.5 hours in two periods respectively), the total occupancy durations of 1-2 hours and 2-3 hours seemed to be the most popular in all test days and in comparison, 0.5-1 hours and 1-2 hours were the most popular durations that visitors would spend in one sitting. The similarity in the use patterns reflects the large postgraduate and researcher populations in the university who continued to work as normal during the vacation.

5.4 Pattern of users taking breaks

Insert Figure 12 here

Figure 12 presents the number of times individual users were away from the room in relation to the total amount of time they spent in the library room respectively in a day. It is clear the number of dots decreased significantly with the increasing occurrence of absence and the maximum number of times a user was absent from the room was 6 times. Generally, as the total occupancy duration increased, the number of times a user was away from the room also increased, although there was a wide range of variations in the total occupancy duration for any given number of occurrences of absence, reflecting the diverse behavioural patterns of individual users. From a health and wellbeing perspective, users especially those who work with computers require regular breaks. While there is no legal requirement on frequency of breaks, the UK Health and Safety Executive suggests it is

19

advisable to take short frequent breaks of 5-10 minutes rather than longer infrequent breaks e.g. 20 minutes after 2 hours (Health and Safety Executive, 2016). The occupancy and absence patterns in Figure 12 show that as some users took frequent breaks, the majority did not, with examples of users taking no break for a 4-hour period, 1 break in 5 hours, 2 breaks in 6 hours etc. This occupancy pattern information is thus valuable in assisting the monitoring and improvement of health and wellbeing of building users, which in turn would affect their productivity. Seen from another angle, the occupancy information in Figure 12 could help to study and understand student behaviours and habits, e.g. their 'study stamina'. The bottom left part of this figure stands for 'light study with occasional absences', the bottom right for 'light study with frequent absences', the top left for 'extensive study with occasional absences' and the top right for 'extensive study with frequent absences'.

5.5 One-time and repeat visitors

Insert Table 4 here

Data analysis also reveals information about one-time visitors and repeat visitors. In total, 1666 duration vectors (D_M) belonging to 1282 different MAC addresses were generated from the raw data gathered in this experiment. It is worthwhile to summarise again that one vector contains occupancy and absence durations for one detected device in one monitoring day. E.g., a device with 2 different vectors means it showed up in the room on 2 separate monitoring days. Data analysis showed there were 1091 one-time visitors

20

and 191 repeat visitors during the test period, as shown in Table 4. As the data for both one-time and repeat visitors is presented by vector D_M that has been clustered in 4 patterns and the probability distribution of OT and AT for each pattern has been detailed, the analysis for one-time and repeat visitor can be linked to occupancy patterns identified above. An overwhelming majority (85.7%) of one-time visitors are the 'observer' type, meaning most of one-time visitors had a quite short-occupancy duration in a single day, while 88.5 % of repeat visitors were 'true room users (Pattern B and D)' and 'inspectors (Pattern C)'. The visiting frequency ranged from 1 to 9. Most (68.9%) of repeat visitors used the room for 2-3 different days and they were all 'true room user' types, while the visitors using the room for 6-9 days were all 'inspectors'. After consultation with the library management personnel, these 'inspectors' with higher frequencies possibly were library security and cleaning staff with busy duty shifts. This may indicate that except for the library staff, few visitors used this room routinely or more than once a week on average.

A significant and somewhat surprising outcome is the lack of repeat visitors. At one level, this could be a good sign, as no one seemed to dominate the use of the Knowledge Exchange Room, which is a space designated for informal discussions and meetings. On the other hand, it remains to be seen in a future larger-scale study whether the lack of repeat visitors is a common feature for other library study spaces. As a primary destination for private and group study on campus, attracting repeat visitors could be a sign of successful library design. Besides, a further study of the data shows that no

device showed a pattern of repeat visit after midnight, i.e., users visited after midnight either just once or not at all. These findings offer valuable insights into diverse types of users and their use pattern of the library room.

6. Conclusions

The occupancy patterns have been identified and studied statistically based on data from Wi-Fi based indoor location system and data mining techniques. The system in this study has demonstrated its effectiveness in providing occupancy information with a relatively high degree of granularity and accuracy in this study. The occupancy patterns indicate that the 24-hour opening policy for the library during term time was unnecessary and an 18-hour opening period would probably be sufficient for users' needs, while potentially reducing energy consumption and staff time/cost substantially. On the other hand, it could benefit user productivity if library-opening hours during the summer vacation were extended to include the early evening hours.

Four occupancy patterns were identified as 'observers', 'inspectors', 'intensive learners' and 'normal learners'. Most visitors were found to be 'observers' and 'inspectors', the short-occupancy visitors, while a minority were 'intensive learners' and 'normal learners', i.e. the long-occupancy 'true room users', for whom the Knowledge Exchange Room was established. This points to a need for the space to be redesigned or relocated to help improve the appropriate room utilisation. The numbers of visitors of all types dropped significantly from the term time to the vacation period. Individual

and total occupancy durations were found to peak at 0.5-2 hours and 1-3 hours respectively for long-occupancy visitors. The occupancy duration was significantly correlated with the arrival time. During the term time, short-occupancy visits peaked around late morning and lunchtime hours, compared to long-occupancy visits which peaked around 2 - 4 pm in the afternoon. This could have significant implications of optimal opening hours targeting the 'true room users'.

Further data analysis showed that the majority of 'true room users' tended not to have frequent breaks with some taking no break for 4 hours and some having 1 break in about 5 hours. This could have health and wellbeing implications for users as well as for their effectiveness and productivity. Furthermore, it is shown that the vast majority of users were one-time visitors, most of whom were the 'observers' type of users, while a minority were repeat visitors. Even among repeat visitors, most of them used the room less than 1 day a week on average, further highlighting the somewhat surprise absence of regular users.

Future investigations could include larger-scale monitoring campaigns involving a greater number and more types of spaces. The quantitative investigation could be combined with some direct engagement with users to bring additional understanding of context and underlying reasons of specific user behaviours. Occupancy duration information could be also important for the assessment of indoor air quality (IAQ) and user exposure levels, where the exposure duration is a key parameter and could be provided by

the Wi-Fi based indoor location technology. Thus the future work may include IAQ and wellbeing monitoring.

References

Accuware (2016). https://www.accuware.com/support/wi-fi-location-monitor introduction/. Accessed on December 2016.

Akbar, A., Nati, M., Carrez, F., & Moessner, K. (2015). Contextual occupancy detection for smart office by pattern recognition of electricity consumption data. *IEEE International Conference on Communications*. IEEE.

Albert, A., & Rajagopal, R. (2013). Smart meter driven segmentation: what your consumption says about you. *IEEE Transactions on Power Systems, 28*(4), 4019-4030.

Chang, W. K., & Hong, T. (2013). Statistical analysis and modeling of occupancy patterns in open-plan offices using measured lighting-switch data. *Building Simulation, 6*(1), 23-32.

COMEAP (2004). Guidance on the Medical Effects of Air Pollutants; Department of Health, London, U.K. Retrieved form http://webarchive.nationalarchives.gov.uk. Accessed on December, 2016.

Dar, U. I., Georges, L., Sartori, I., & Novakovic, V. (2015). Influence of occupant's behavior on heating needs and energy system performance: a case of well-insulated detached houses in cold climates. *Building Simulation, 8*(5), 499-513.

Demir, L., Cunche, M., & Lauradoux, C. (2014). Analysing the privacy policies of Wi-Fi trackers. *Workshop on Physical Analytics*, 2014:39-44.

Dong, B., & Andrews, B. (2009). Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. *Proceedings of Building Simulation*.

Duarte, C., Wymelenberg, K. V. D., & Rieger, C. (2013). Revealing occupancy patterns in an office building through the use of occupancy sensor data. *Energy & Buildings, 67*(4), 587-595.

Gauthier, S., & Shipworth, D. (2015). Behavioural responses to cold thermal discomfort. *Building Research & Information, 43*(3), 355-370.

Gilbert, N. (2005). Proposed residential indoor air quality guidelines for formaldehyde. Retrieved from http://vre2.upei.ca/govdocs/fedora/repository/govdocs%3A510/PDF/PDF. Accessed on December, 2016.

Guan, Q., Li, C., Guo, X., & Wang, G. (2014). Compressive classification of human motion using pyroelectric infrared sensors ☆. *Pattern Recognition Letters, 49*, 231-237.

Gul, M. S., & Patidar, S. (2015). Understanding the energy consumption and occupancy of a multi-purpose academic building. *Energy & Buildings, 87*, 155-165.

Gunay, H. B., Fuller, A., O'Brien, W., & Beausoleil-Morrison, I. (2016). Detecting occupants' presence in office spaces: a case study. eSim 2016 Conference, Hamilton DOI: 10.13140/RG.2.2.18486.09280.

Hallett, S., Shao, L., Jones, K., Pryce, G., Fowler, H., Blenkinsop, S. (2013). Community Resilience to Extreme Weather – CREW project final report. Retrieved from http://www.arcc-network.org.uk/wp-content/pdfs/CREW_Final_Report.pdf. Accessed on December, 2016.

Han, Z., Gao, R. X., & Fan, Z. (2012). Occupancy and indoor environment quality sensing for smart buildings. *Instrumentation and Measurement Technology Conference* (Vol.8443, pp.882-887). IEEE.

Hartigan, J. A., & Wong, M. A. (1979). A k-means clustering algorithm. *Applied Statistics, 28*(1), 100-108.

Health and Safety Executive (2016). http://www.hse.gov.uk/msd/faq-dse.htm. Accessed on March, 2017.

Joakim Widén, Andreas Molin, & Kajsa Ellegård. (2012). Models of domestic occupancy, activities and energy use based on time-use data: deterministic and stochastic approaches with application to various building-related simulations. *Journal of Building Performance Simulation,5*(1), 27-44.

Kim, Y. S., & Srebric, J. (2015). Improvement of building energy simulation accuracy with occupancy schedules derived from hourly building electricity consumption. *Ashrae Transactions, 121*(1), 353-360.

Liu, Y. Y. (2015). Research on library lighting intelligent control based on infrared image processing techniques. *Optik - International Journal for Light and Electron Optics, 126*(18), 1559-1561.

Martani, C., Lee, D., Robinson, P., Britter, R., & Ratti, C. (2012). Enernet: studying the dynamic relationship between building occupancy and energy consumption. *Energy & Buildings, 47*(47), 584-591.

Mckenna, E., Krawczynski, M., & Thomson, M. (2015). Four-state domestic building occupancy model for energy demand simulations. *Energy & Buildings, 96*(8), 30-39.

Ortega, J. L. G., Han, L., Whittacker, N., & Bowring, N. (2015). A machine-learning based approach to model user occupancy and activity patterns for energy saving in buildings. *Science and Information Conference* (pp.474-482). IEEE.

Porritt, S. M., Cropper, P. C., Shao, L., & Goodier, C. I. (2012). Ranking of interventions to reduce dwelling overheating during heat waves. *Energy & Buildings, 55*(55), 16-27.

Primo, O. (2015). Workplace Evolved. Worktech 15 Conference. New York. 14 May 2015.

Rosenbaum, R. K., Meijer, A., Demou, E., Hellweg, S., Jolliet, O., & Lam, N. L., et al. (2015). Indoor air pollutant exposure for life cycle assessment: regional health impact factors for households. *Environmental Science & Technology, 49*(21), 12823-31.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational & Applied Mathematics, 20*(20), 53-65.

Spataru, C., & Gillott, M. (2011). The use of intelligent systems for monitoring energy use and occupancy in existing homes. *Intelligent Buildings International, 3*(1), 24-31.

Spataru, C., Gillott, M., & Hall, M. R. (2010). Domestic energy and occupancy: a novel post-occupancy evaluation study. *International Journal of Low-Carbon Technologies, 5*(3), 148-157.

Subhan, F., Hasbullah, H., Rozyyev, A., & Bakhsh, S. T. (2011). Indoor positioning in Bluetooth networks using fingerprinting and lateration approach. *International Conference on Information Science and Applications* (pp.1-9). IEEE.

Wang, Y. (2016). Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning and data mining (unpublished master's thesis). University of Reading, UK.

Wang, Y., & Shao, L. (2017). Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning. *Building and Environment,* 114:106-117.

Wang, Y., Liu, J., Chen, Y., Gruteser, M., Yang, J., & Liu, H. (2014). E-eyes: device-free location-oriented activity identification using fine-grained WiFi

signatures. *International Conference on Mobile Computing and NETWORKING* (pp.617-628). ACM.

Yang, Z., & Becerik-Gerber, B. (2014). Modelling personalized occupancy profiles for representing long term patterns by using ambient context. *Building & Environment, 78*(8), 23-35.

Yang, Z., Li, N., Becerik-Gerber, B., & Orosz, M. (2014). A systematic approach to occupancy modelling in ambient sensor-rich buildings. *Simulation, 90*(8), 960-977.

List of captions for all figures

Figure 1. The location of Knowledge Exchange Room.

Figure 2. (a) The measurement sensor (left) and (b) The nodes placement plan (right).
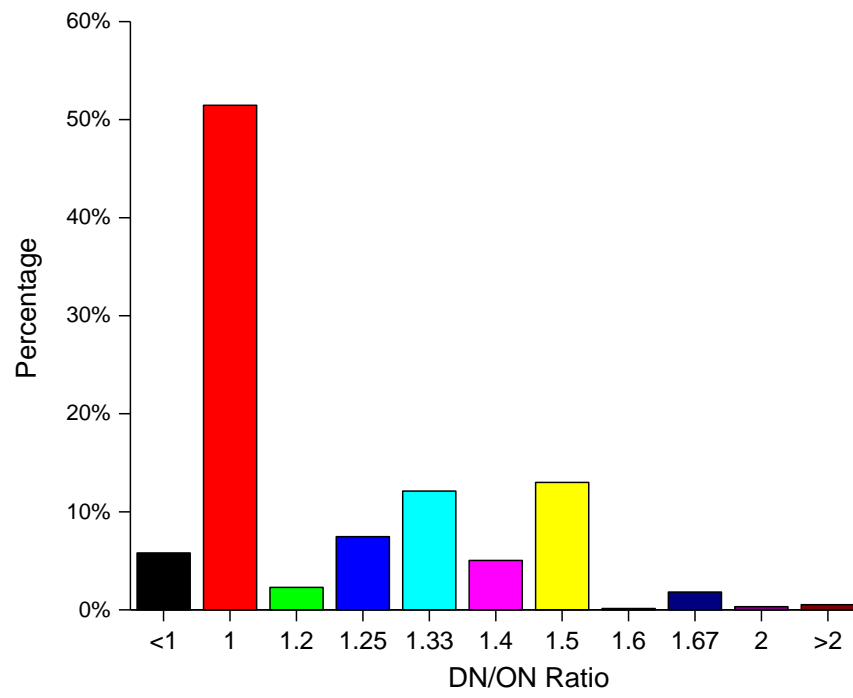
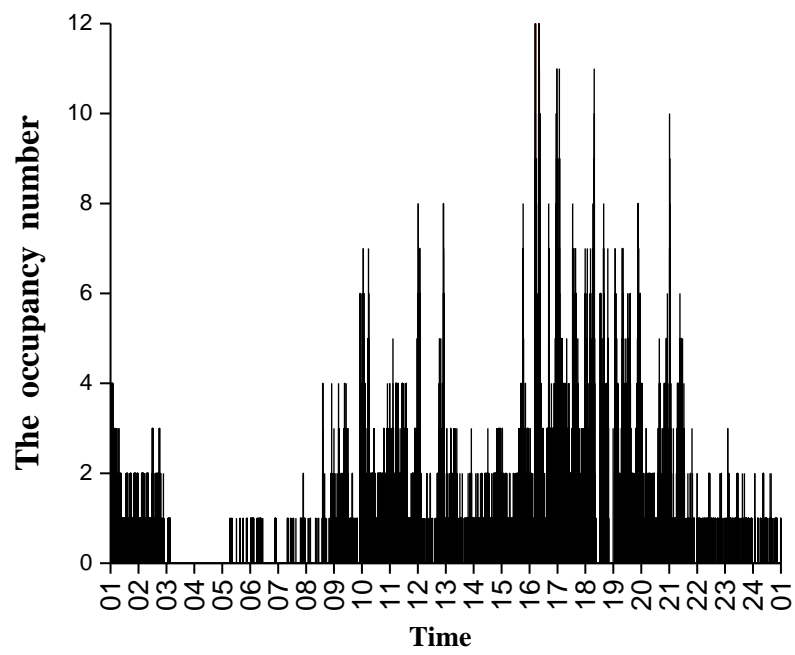Figure 3. The frequency of individual DN/ON ratio.

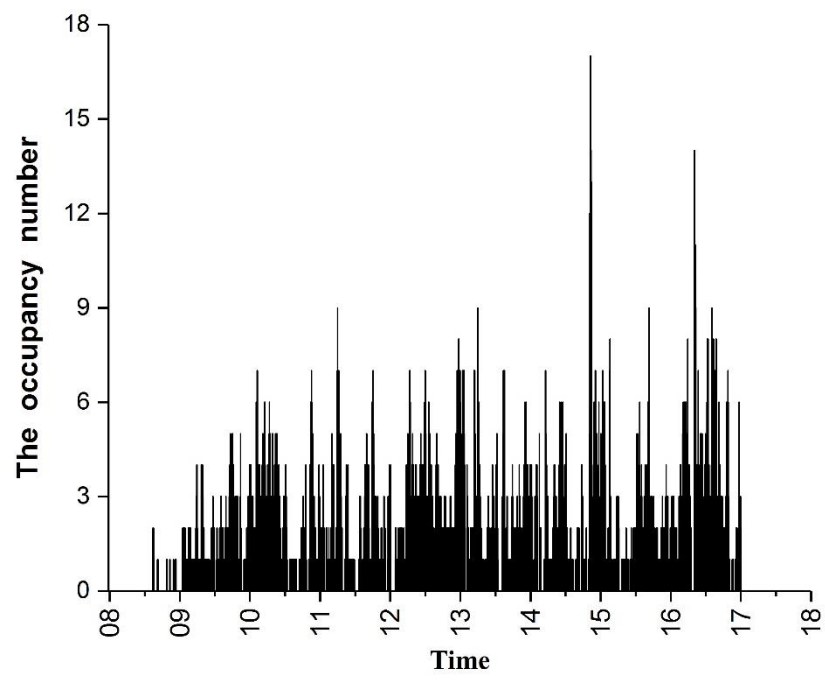Figure 4. The occupancy pattern on 6th June.
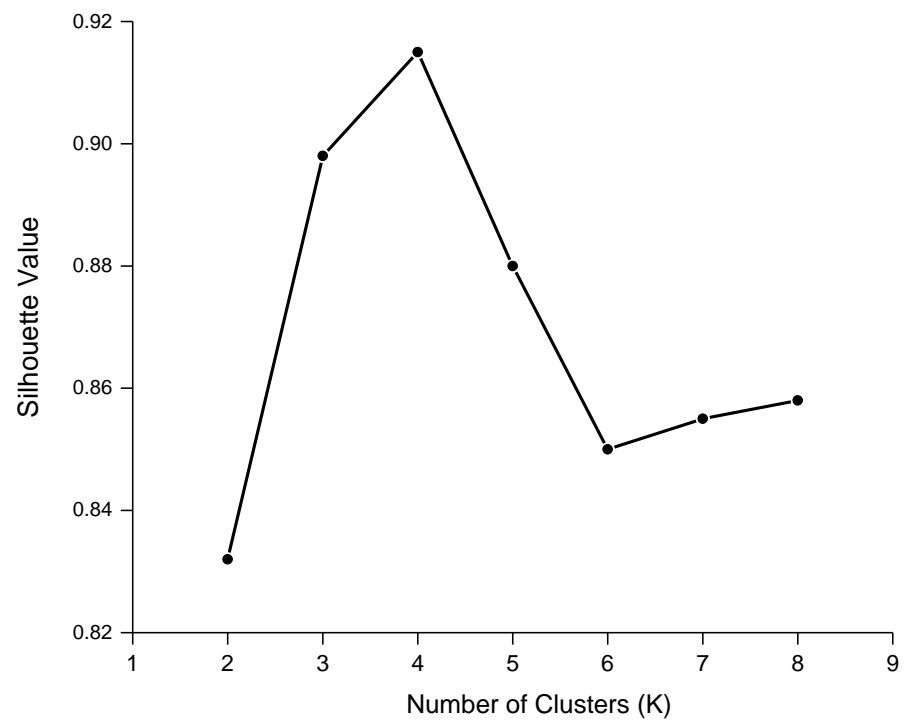
Figure 5. The occupancy pattern on 15th June.

Figure 6. The relation between Silhouette Value and cluster number.

Figure 7. Four different occupancy patterns.

Figure 8. Distribution of different occupancy patterns for each test day.

Figure 9. Arrival time for short/long - occupancy visitors.

Figure 10. Distribution of total occupancy durations for each test day.

Figure 11. Distribution of durations of individual occupancy periods in each test day.

Figure 12. The number of times absent from the room in relation to users' total occupancy duration in a day.

Figure 1. The location of Knowledge Exchange Room

Figure 2 (a) The measurement sensor (left) and (b) The nodes placement plan (right)

Figure 3. The frequency of individual DN/ON ratio

Figure 4. The occupancy pattern on 6th June

Figure 5. The occupancy pattern on 15th June

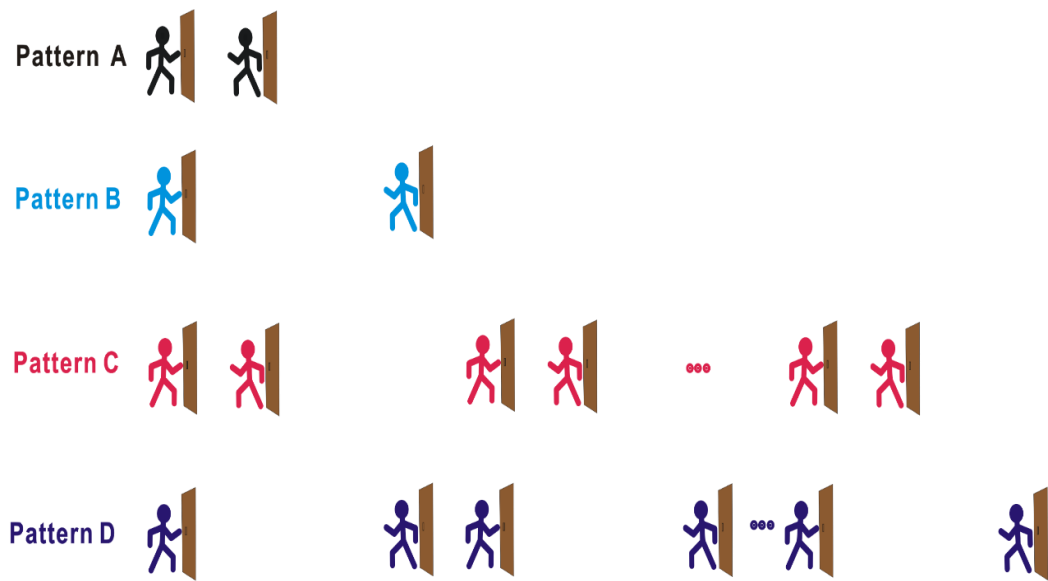Figure 6. The relation between Silhouette Value and cluster number

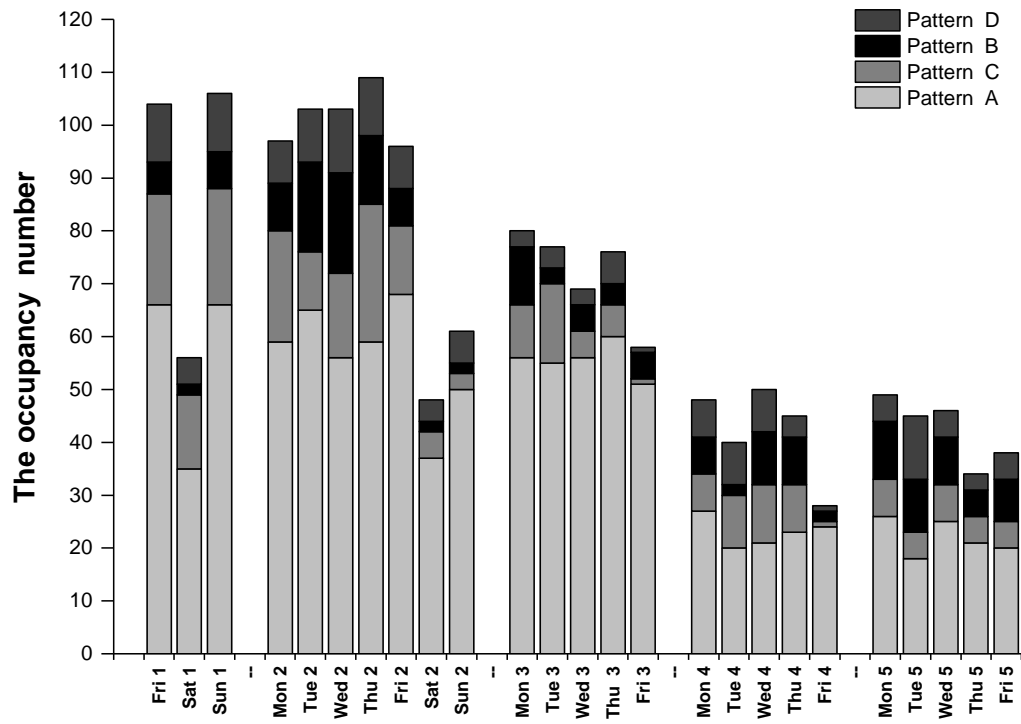Figure 7. Four different occupancy patterns

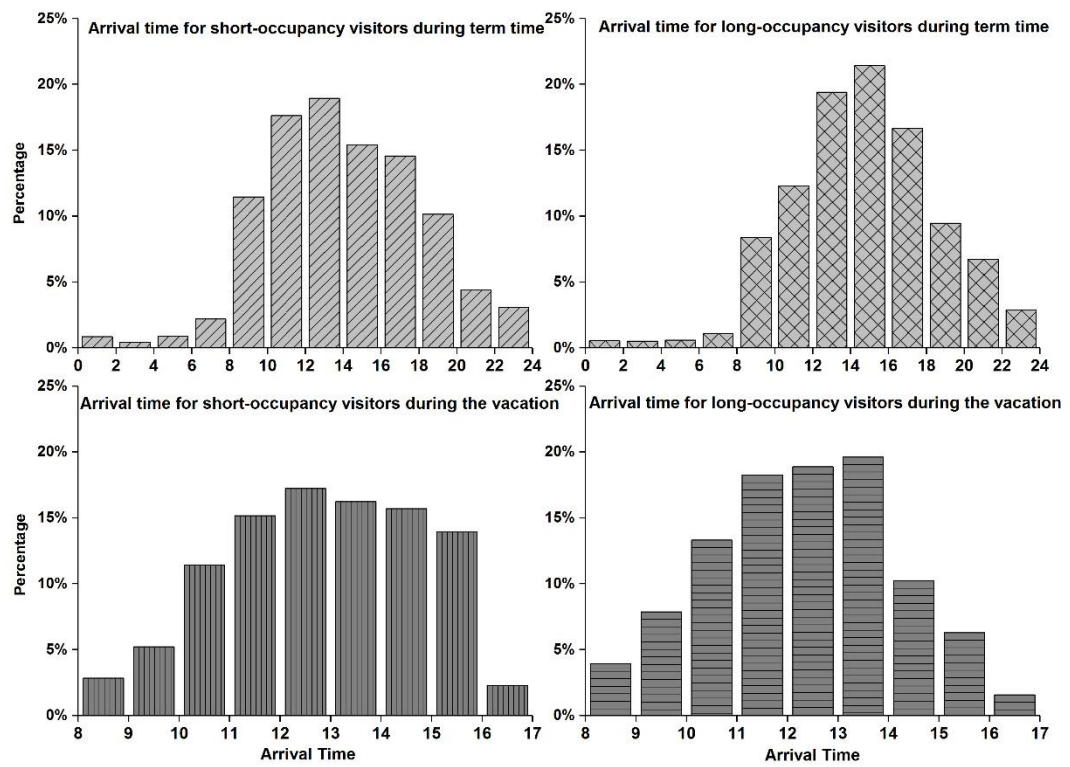Figure 8. Distribution of different occupancy patterns for each test day

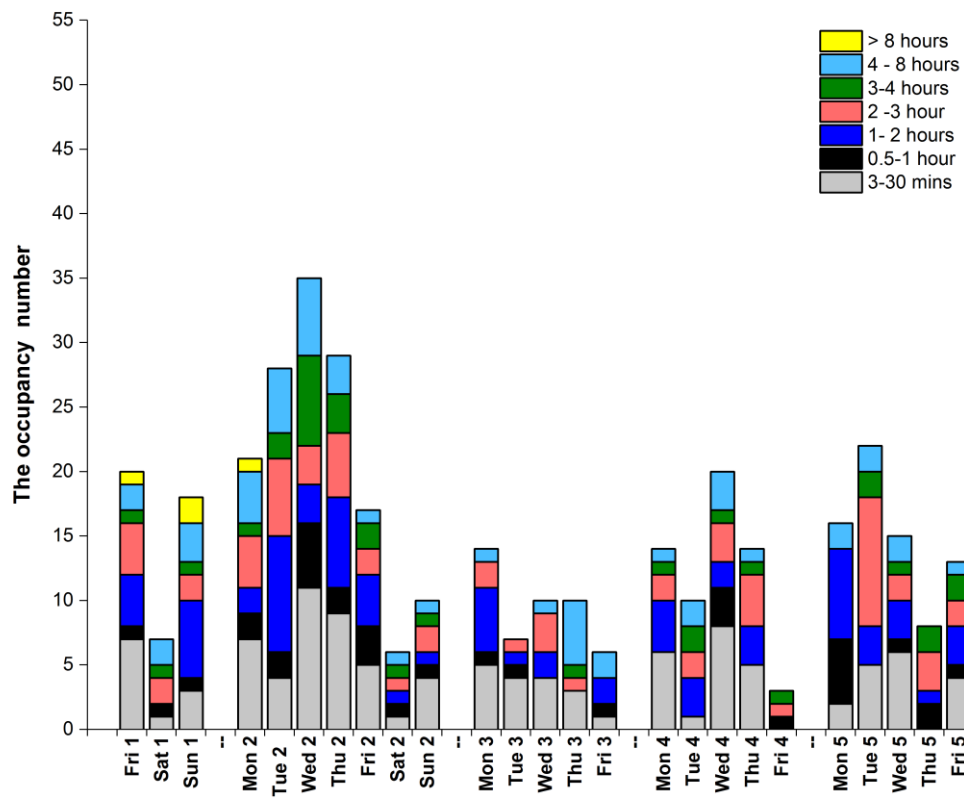Figure 9. Arrival time for short/long - occupancy visitors

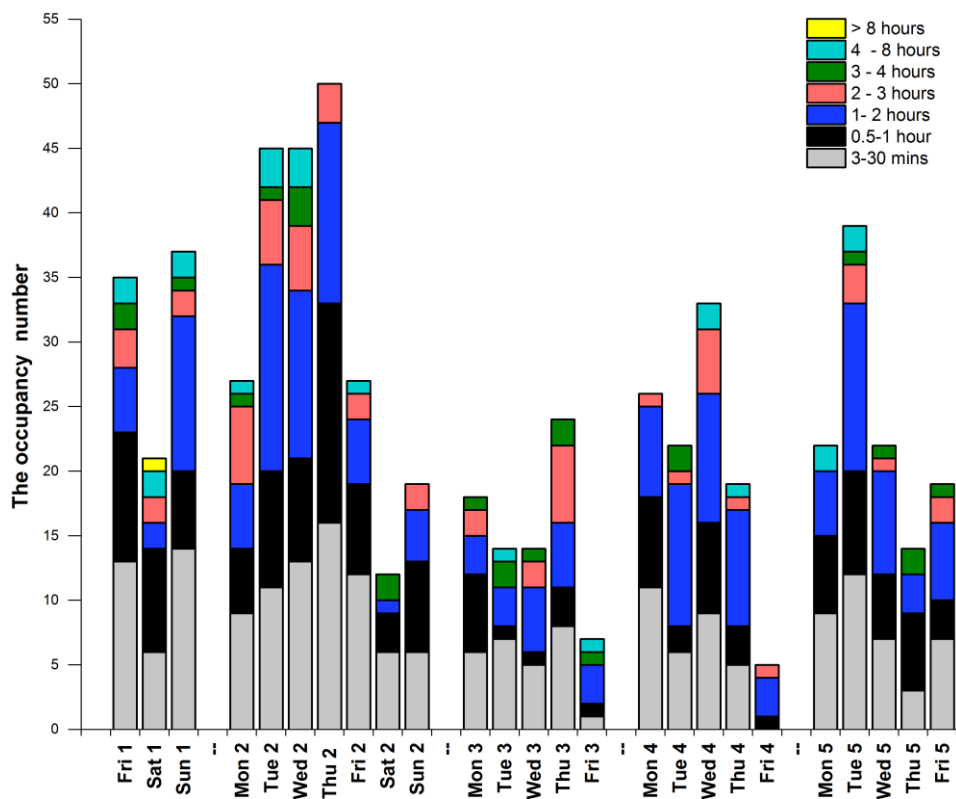Figure 10. Distribution of total occupancy durations for each test day

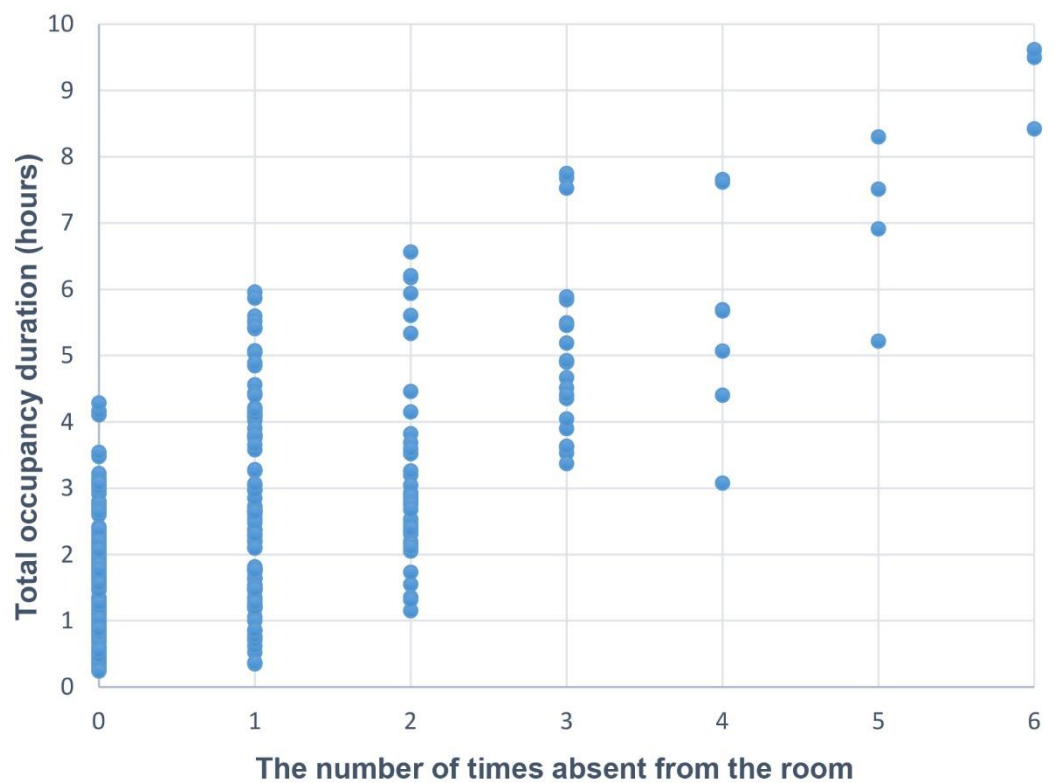Figure 11. Distribution of durations of individual occupancy periods in each test day

Figure 12. The number of times absent from the room in relation to

users' total occupancy duration in a day

List of captions for all tables

Table 1. Library opening schedule during the test period.

Table 2. Sampling test arrangement.

Table 3. Occupancy and absence durations for each occupancy pattern.

Table 4. The visiting frequency with corresponding device numbers.

Table 1. Library opening schedule during the test period

Time period	Opening timetable
Summer tem (27th May-10th June, 2016)	24-hour open except from 9 pm on Saturday to 8.30 am on Sunday
Summer vacation (11th- 26th June, 2016)	Only open from 8.30 am to 5 pm on weekdays

Table 2. Sampling test arrangement

Sampling test	Date	Time
1	27th May	10am-11am
2	28th May	3pm-4pm
3	29th May	7pm-8pm
4	1st June	8pm-9pm
5	2nd June	8am-9am
6	4th June	1pm-2pm
7	7th June	1am-2am
8	8th June	6am-7am
9	10th June	10pm-11pm
10	13th June	10am-11am
11	15th June	12pm-1pm
12	16th June	3pm-4pm
13	21st June	2pm-3pm
14	22nd June	9am-10am

Table 3. Occupancy and absence durations for each occupancy pattern

Category	Main range
Pattern A occupancy duration	0.5 to 1 minute
Pattern B occupancy duration	0.5 to 1 hour
Pattern C individual occupancy duration	0.5 to 1 minute
Pattern C individual absence duration	2 to 8 hours
Pattern D individual occupancy duration	0.5 to 1 hour
Pattern D individual absence duration	10 minutes to 2 hours

Table 4. The visiting frequency with corresponding device numbers

Frequency (in days)	Device number
1	1091
2	85
3	59
4	26
5	12
6	2
7	5
8	1
9	1