Investigating the correlation dynamism between WiFi connection counts and camera-based occupancy counts

Nastaran Alishahi¹, Mohamed M. Ouf¹, Mazdak Nik-Bakht¹
¹Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada

Abstract

WiFi connection counts were previously suggested as a reliable proxy for occupancy pattern identification, to optimize building systems’ operation. However, validating and calibrating WiFi counts with actual occupancy count data have remained limited in the literature, due to the costly and time-consuming process of collecting ground-truth data. This study uses camera-based image recognition counters as a source of continuous ground-truth data, to examine the correlation between WiFi connection counts and actual occupancy counts. The results show a significant dependency for the ratio of the two values, on the time of day and occupancy level. Therefore, these factors must be accounted for, when using WiFi counts to adapt building operations to variable occupancy.

Key Innovations

- Using camera-based occupancy counts as a continuous data stream for occupancy counts
- Investigating the variation in the ratio of WiFi connection counts to camera-based occupancy counts over a long duration

Practical Implications

Extracting accurate occupancy information from WiFi connection counts can help to build operators and customize the schedule of building systems accordingly.

Introduction

Accurate building occupancy information is key for optimizing building systems’ operation. Previous studies have shown adjusting the operation schedules of Heating, Ventilation, and Air Conditioning (HVAC) systems based on occupancy information can reduce the building energy consumption by 10-40% (Pérez-Lombard et al., 2008; Trêka & Hensen, 2010). This fact necessitates building systems to be operated according to the unique characteristics of their occupancy. However, usually due to the lack of accurate occupancy information, most building systems are operated based on pre-programmed schedules with the assumption of full or nearly full occupancy (Ouf et al., 2019).

Previous studies employed different technologies to obtain accurate occupancy information with the aim of optimizing building energy consumption. These technologies mostly included environmental sensors measuring the CO₂ concentration, temperature, humidity, and pressure, or cameras, radio frequency identification (RFID) tags, Bluetooth Low Energy (BLE) beacons, and Passive Infrared (PIR) sensors. Sun, Zhao, and Zou (2020) and Chen, Jiang, and Xie (2018) have conducted a comprehensive literature review on the pros and cons of these technologies. They mostly have different limitations in terms of intrusiveness, accuracy, privacy violation, additional installation and maintenance costs, etc. Recent studies proposed WiFi connection counts as a reliable proxy for occupancy estimation. Since this technology is widely available in most buildings, especially offices and educational buildings, and is commonly used by occupants in these types of buildings, it is considered a non-intrusive approach with relatively low or no additional cost.

Although WiFi connection counts have been previously proposed and tested as a proxy for real-time building occupancy, validation with ground-truth actual occupancy count data has been limited. Since obtaining ground-truth occupancy count data is challenging, many studies typically relied on short-term manual counting at a room- or zone-level. At such levels, occupancy is less dynamic compared to the full building-level.

Furthermore, due to limited ground-truth data, most previous studies proposed fixed values as the ratio between WiFi connection counts to occupancy counts. Possible changes in this ratio over time as well as the influencing factors on its variation are rarely investigated.

To this end, this study uses camera-based occupancy counters to establish a continuous ground-truth data stream for validating the correlation between WiFi connection counts and actual building occupancy counts. The camera-based counters use image recognition to detect the number of people (heads) going in and out of a building without any facial recognition to avoid privacy issues. The objectives of this study include (i) identifying stationary devices connected to WiFi to be used as a baseline that does not correspond to occupancy; (ii) identifying temporal changes in the ratio of non-stationary WiFi connection counts to occupancy counts over time, and (iii) investigating the influencing factors on the identified ratios. K-means clustering, and other statistical analysis methods were applied to WiFi connection counts and camera-based occupancy counts data collected from a university library building in Montreal to achieve these objectives.
The rest of this paper is organized as follows: A summary of previous works is presented in Background; The study method is described in Methodology; followed by an introduction on Case Study; Results and Discussion present the results of the experiment and discuss the major findings; and finally, Conclusion provides the conclusion and discusses limitations and future works.

### Background

Although several previous studies showed a strong statistically significant correlation between WiFi connection counts and actual occupancy counts (Mohottige et al., 2018; Ouf et al., 2017), they mostly relied on limited ground-truth data to account for stationary devices and occupants carrying more or less than one WiFi-connected device. Obtaining ground-truth data through manual counting is challenging, costly, intermittent, inaccurate, and time-consuming. Therefore, a majority of studies have focused on occupancy estimation, using short-term manual counting. Besides manual counting, some studies proposed video recordings as a source of ground-truth data (Petersen et al., 2016) which can result in privacy issues. For example, Hobson et al. (2019) recorded a total of 208.5 hours of ground-truth observations from different floors of an academic office building during different periods. This data was collected by manually reviewing videos recorded using cameras installed at the entrance and exit points of each floor. They developed Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) models for different floors using only a part of this ground-truth dataset. Wang, Chen, and Hong (2018) also collected three-day ground-truth data by manually observing cameras in a graduate student office to develop ANN, k-Nearest Neighbors (kNN), and SVM models for estimating occupancy counts. Wang et al. (2019) collected camera-based data from two floors of an office building with maximum occupancy counts of 20 for five weeks. The accuracy of this data was validated through two hours of manual counting at each of the entrances and then it was used as ground-truth data to estimate occupancy through Random Forest, Deep Learning Neural Network, and Long-term Short-term Memory networks (LSTMs).

Most of the studies employed short-term manual counting to develop a Linear Regression model for translating WiFi connection counts to actual occupancy counts. For example, Ouf et al. (2017) identified 0.79 as the coefficient of WiFi connection counts with a bias of 1.4 in their regression model developed for a classroom with one week of ground-truth data. In another study on a lecture room, Jagadeesh Simma, Mammoli, and Bogus (2019) developed different models to investigate the correlation of WiFi connection counts and occupancy counts and reached a coefficient for actual occupancy counts ranging between 0.75 and 0.99.

Some studies have shown the variation of ratio based on the time and the space type but since the ground-truth data is limited, one fixed value was proposed as the final ratio. For example, Wang and Shao (2018) suggested 1.16 as the ratio of WiFi connection counts to occupancy counts in a typical study room which was the average of WiFi connection counts divided by occupancy counts during a period of 14 hours. In another study, Mohottige et al. (2018) calculated a proportionality factor of average devices per occupant to be 1.3 in an experiment on a university campus while 37 samples of actual occupancy counts were collected from 4 classrooms. Longo, Redondi, and Cesana (2019) calculated the average ratio in different types of rooms in a university building by dividing the number of WiFi connections by ground-truth occupancy value showing a higher ratio of devices to occupants in a wireless network laboratory compared to classrooms.

In addition to ratio, some studies that had access to devices’ Media Access Control (MAC) address and connections history, were able to remove multiple devices of a unique user and also exclude devices with long duration of connection as stationary devices as well as devices showing a short duration of connection (Mohottige et al., 2018; Z. Wang et al., 2019). However, accessing this level of detail is not always feasible due to privacy concerns.

Although short-term manual counting cannot provide insight into the variation of the ratio, it at least can help developing models that estimate occupancy counts from WiFi connection counts with reasonable accuracy. However, for developing models that predict future occupancy counts, a longer duration of ground-truth data is needed. Different approaches were proposed to overcome the challenge of collecting ground-truth data in

### Table 1: Linear Regression models developed for translating WiFi connection counts to actual occupancy counts.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Linear Regression Model</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jagadeesh Simma, Mammoli, and Bogus (2019)</td>
<td>Y (WiFi) = 0.9885X (Sensor/Camera) – 0.5363</td>
<td>R²: 0.963</td>
</tr>
<tr>
<td></td>
<td>Y (WiFi) = 0.9585X (Sensor/Camera) + 0.5252</td>
<td>R²: 0.958</td>
</tr>
<tr>
<td></td>
<td>Y (WiFi) = 0.842X (Sensor/Camera) – 0.5135</td>
<td>R²: 0.940</td>
</tr>
<tr>
<td></td>
<td>Y (WiFi) = 0.8339X (Sensor/Camera) + 0.5374</td>
<td>R²: 0.916</td>
</tr>
<tr>
<td></td>
<td>Y (WiFi) = 0.8925X (Sensor/Camera) + 0.1607</td>
<td>R²: 0.910</td>
</tr>
<tr>
<td></td>
<td>Y (WiFi) = 0.7516X (Sensor/Camera) + 0.3936</td>
<td>R²: 0.905</td>
</tr>
<tr>
<td>Ashouri et al. (2019)</td>
<td>Bias = 27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y (Occupancy counts) = 1.27X (WiFi)</td>
<td>R²: 0.900</td>
</tr>
<tr>
<td>Hobson et al. (2020)</td>
<td>Y (WiFi) = 1.2X (Occupancy counts)</td>
<td>R²: 0.845</td>
</tr>
<tr>
<td>Ouf et al. (2017)</td>
<td>Y (Occupancy counts) = 0.79X (WiFi) + 1.4</td>
<td>R²: 0.703</td>
</tr>
</tbody>
</table>
occupancy prediction studies. Ashouri et al. (2019) created ground-truth data for almost 9 weeks with the help of a linear regression model that was developed with limited ground-truth data collected in an office building. This regression model translated WiFi connection counts to occupancy counts with 1.27 as the coefficient of WiFi connection count. They counted 27 stationary devices during the night when minimum occupancy was expected. In this study, day-ahead prediction models were then trained on the created 9-week ground-truth data. In another study, Hobson et al. (2020) also produced ground-truth data for a 7-month study through a regression model trained on eight-and-a-half weekdays of ground-truth data with a coefficient of 1.2 for actual occupancy counts. The original ground-truth data was collected by manually reviewing the cameras installed at the entrances and exits of an academic office building. Finally, the created ground-truth time-series data was used to develop a classification tree for predicting day-ahead occupancy day type. A summary on Linear regression models developed based on WiFi connection counts and actual occupancy counts are presented in Table 1.

### Methodology

This study consisted of three main steps to identify the correlation between WiFi connection counts and actual occupancy counts over a longer period of time, which are discussed in this section. These steps were then implemented in a case study which is described in detail.

#### Data collection and preparation

The first step entailed collecting the number of devices connected or trying to connect to each WiFi Access Point (AP) in the building. This number, called ‘associated’ device counts, can be retrieved on an hourly or sub-hourly basis from most of the typical WiFi network management platforms. Since this data includes only aggregated device counts, it would not raise any privacy concerns. On the other hand, this study also relied on data from camera-based occupancy counters as a source of continuous ground-truth for the entire duration of collected WiFi data. For this approach, cameras installed at the main entrances/ exits of the building were used. The accuracy of these counters was first tested through manual counting at random points over time. In the next step, both datasets were cleaned, transformed into the proper format, and finally synchronized with each other. To identify outliers among daily profiles, the Kruskal-Wallis H test was used to investigate whether statistically significant differences existed between different daily profiles of each day of the week, as well as all days together.

Furthermore, since WiFi connection counts are strongly correlated with occupancy counts, it is expected that both datasets follow similar patterns. Clustering daily profiles of both datasets can provide insights about daily patterns and peak occupancy times which were used to validate the synchronization of the two data streams. K-means clustering based on Euclidian distance (as the measure of dissimilarity) was selected to extract daily profiles, and Davies Bouldin Index (DBI) was used as the mathematical performance measure for finding the optimal number of clusters. In addition to WiFi and camera data, the building’s contextual information including floor areas, HVAC system’s layout as well as APs’ location attributed to each zone was retrieved from the drawings of the building.

#### Establishing a Baseline

The collected WiFi connection counts data includes stationary devices such as printers, servers, etc. which are connected to the network for almost entire days. Although the number of these devices can be dynamic throughout the day due to changing into idle mode, the level of such dynamism is considerably lower than the counts from the occupants’ devices. A fixed value can be recognized as the baseline to account for these stationary devices for weekdays and weekends. These values should be subtracted from the WiFi connection counts at each interval to decrease its deviation from actual occupancy counts.

To find this fixed value, firstly, the hour at which minimum occupancy counts occurred on each day was extracted. It is expected that during these hours, the deviation of WiFi connection counts from actual occupancy counts was minimal. Secondly, the differences between WiFi connection counts and occupancy counts during these hours were considered as the number of stationary devices. After removing outliers from the dataset of differences, the most frequent value was considered as the fixed baseline for WiFi connection counts.

#### The ratio between WiFi connection counts and camera-based occupancy counts

The next step entailed calculating the ratio of WiFi connection counts to occupancy counts. A time-series of ratios was produced as the quotient of WiFi connection counts and occupancy counts at each time interval. The ranges of these ratios at different hours of the day were extracted to investigate their variation at each hour. Then, daily profiles of ratios were clustered, using k-means clustering, to provide insights about the dominant patterns of ratios and the influence of time and occupancy level on these patterns.

#### Case Study

The case study is a 13-story university building located in Montreal, Canada. The experiment was conducted on four stories (total area of 19 180 m²) of this building which are used as a library. These four floors are very similar in layout and space types and mostly contain reading rooms. A summary of the area of the different spaces identified in the selected four floors is presented in Table 2.

**Table 2: The total area of different spaces in the investigated four floors of the case study building.**

<table>
<thead>
<tr>
<th>Space type</th>
<th>Area (m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office</td>
<td>2 671</td>
</tr>
<tr>
<td>Corridor/Lobby</td>
<td>6 087</td>
</tr>
<tr>
<td>Reading room</td>
<td>7 110</td>
</tr>
<tr>
<td>Group study room/Lecture room</td>
<td>630</td>
</tr>
<tr>
<td>Collection spaces</td>
<td>2 680</td>
</tr>
</tbody>
</table>
Data collection and preparation

WiFi connection count data was collected from a total of 152 Cisco Aironet AIR-CAP3702I and Cisco Aironet 2802I APs that were located throughout the four investigated floors of the case study building. The data was provided in 3-min intervals for the period of nine weeks, from January 13, 2020, to March 12, 2020, which was aggregated into hourly and flattened into one dataset for the building-level analysis. Camera-based occupancy count data including counts of occupants entering or leaving the investigated four floors of the library building was also collected in 5-min intervals for the same time. The counts were provided by two cameras, located at the two main gateways of the library on the second floor. All visitors are obliged to pass through these gates, but staff may use other access points. After synchronizing collected data from two cameras with actual time, the occupancy count at each interval was calculated which was then aggregated into hourly counts at the building-level. The cumulative occupancy during this period revealed that the total numbers of occupants entering and leaving the investigated floors did not match at the end of each day, especially during weekdays. This might be due to technical faults in the camera or because some of the staff used other access points. These daily differences were accumulated and caused the camera count time-series to show negative values after some days. Therefore, the camera-based occupancy count data was calibrated by setting the minimum occupancy value of each day at zero. In another word, the daily profiles of camera-based occupancy count were shifted vertically to start from zero every day. For the rest of the study, the prepared camera-based occupancy count data was used as the ground-truth.

Results and Discussion

This section presents the descriptive data analytics and discusses the case study results. The discussion focuses on identifying the baseline and investigating changes in the ratio of WiFi connection counts to occupancy counts.

Descriptive Analytics

Figure 1 shows the time-series for WiFi connection count and camera-based occupancy count during the nine weeks of the experiment. Although week no. 7 (the week of the university’s winter break) shows a lower level of occupancy compared to the weeks before and after, based on the results of the Kruskal-Wallis H test, no statistically significant differences were observed among the days of this week with other weeks. Descriptive statistics of WiFi connection count and camera-based occupancy count data are presented in Table 3.

The daily profiles of WiFi connection counts and camera-based occupancy counts were clustered through the k-means clustering algorithm and two clusters were recognized for each of the datasets, based on DBI metrics. Figure 2 shows the two clusters with their centroids, for both camera and WiFi counts. In both patterns for each dataset, counts start increasing at around 8 a.m., then reach a peak between 3 and 5 p.m. when they start to decrease until midnight. The main differences between patterns of the two clusters for each dataset are their average values and the time at which the peak occurs.

Figure 3 illustrates the membership of days of the week to each cluster, for WiFi and camera count data. As seen in this figure, pattern 0, having a comparatively higher level of occupancy, is mostly allocated to weekdays while weekends mainly follow pattern 1 with a lower level of occupancy. The similarity between extracted clusters for each dataset, in terms of the memberships as well as the number of daily profiles in each cluster, is also noticeable. However, despite the similarity of patterns in terms of shape and distribution between different days of the week, a one-hour difference was observed between the peak times of WiFi connection count and camera-based occupancy count data. While the peak occupancy time for patterns 0 and 1 of the WiFi counts happens between 4 p.m. and 5 p.m., for the camera-based occupancy counts, it occurs between 3 p.m. and 4 p.m. Since camera-based occupancy was validated through manual counting, it appears that WiFi connection count data was generated with a one-hour lag. This error might be due to WiFi network management platform failure in reporting real-time connection counts.

Table 3: Descriptive statistics of WiFi connection count and camera-based occupancy count data

<table>
<thead>
<tr>
<th>Descriptive statistics</th>
<th>WiFi connection count data</th>
<th>Camera-based occupancy count data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>Max</td>
<td>3449</td>
<td>2180</td>
</tr>
<tr>
<td>Median</td>
<td>631</td>
<td>350.5</td>
</tr>
<tr>
<td>Mode</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td>Mean +/- SD&quot;</td>
<td>(24.8, 1787.4)</td>
<td>(-13, 1140.2)</td>
</tr>
</tbody>
</table>

* Standard deviation
To validate this assumption, the root mean squared error, as well as the average absolute error of the difference between WiFi connection counts and camera-based occupancy counts were calculated. These metrics were also calculated if WiFi connection counts were shifted backward by one hour. For this comparison, all three datasets were normalized.

According to Table 4, shifting WiFi connection counts backward by one hour resulted in significant error reduction, thus suggesting that the shifted data was more similar to the camera-based occupancy counts and more representative of occupancy patterns.

![Figure 2](image-url)  
*Figure 2: Two clusters of daily occupancy patterns with their centroids for (a) WiFi connection counts, (b) camera-based occupancy counts.*

![Figure 3](image-url)  
*Figure 3: Membership of days of the week to each cluster for (a) WiFi connection counts, (b) camera-based occupancy counts.*

**Table 4:** The descriptive statistics of the two comparisons between camera-based occupancy count vs. WiFi connection count data and camera-based occupancy count vs. one-hour backward shifted WiFi connection count data.

<table>
<thead>
<tr>
<th>Descriptive statistics</th>
<th>Normalized camera-based occupancy count data vs. Normalized WiFi connection count data</th>
<th>Normalized camera-based occupancy count data vs. Normalized one-hour backward shifted WiFi connection count data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root Mean Squared Error</td>
<td>0.083</td>
<td>0.059</td>
</tr>
<tr>
<td>Average absolute error +/- SD</td>
<td>(0.004, 0.116)</td>
<td>(0.001, 0.083)</td>
</tr>
<tr>
<td>Residual distribution</td>
<td><img src="image-url" alt="Residuals" /></td>
<td><img src="image-url" alt="Residuals" /></td>
</tr>
</tbody>
</table>

Proceedings of the 17th IBPSA Conference  
Bruges, Belgium, Sept. 1-3, 2021  
https://doi.org/10.26868/25222708.2021.31057
The rest of the study was conducted after WiFi connection counts were shifted backward by one hour. Figure 4 shows the correlation between WiFi connection and camera-based occupancy counts in this case. The figure reveals a statistically significant positive correlation, with a Pearson product-moment correlation coefficient \( r \) of 0.987.

According to this figure, a high variation in ratios was observed during the early morning, which is due to the low number of occupants and the high level of noise in WiFi connection counts data.

Accordingly, the temporal analysis was limited to intervals with a standard deviation of less than 0.3. This resulted in focusing on 14 consecutive hours between 9 a.m. and 10 p.m. These were the times at which the library had the most occupancy. Daily profiles of ratios during these hours were then clustered using k-means clustering. These profiles, along with two centroids of recognized clusters are plotted in Figure 7.

Establishing a Baseline

The hour at which camera-based occupancy counts were minimal was extracted for each day. These hours ranged between 4 a.m. and 7 a.m. through the entire experiment period while the minimum counts varied between 1 and 15. It was expected that during these hours, the differences between WiFi connection counts and camera-based occupancy counts would be attributed to stationary devices with minimum error. To further investigate these differences, Figure 5 shows their frequency distribution throughout the experiment duration. Accordingly, the frequent value for the number of stationary devices on weekdays and weekends was 29 and 18, respectively. Therefore, For the rest of this study, 29 and 18 connections were subtracted from WiFi connection counts data at each interval of weekdays and weekends, respectively.

The ratio between WiFi connection counts and camera-based occupancy counts

The ratio between WiFi connection counts and camera-based occupancy counts was calculated for each hour during the experiment period, and the range of these ratios for each hour is plotted in Figure 6.

Accordingly, the temporal analysis was limited to intervals with a standard deviation of less than 0.3. This resulted in focusing on 14 consecutive hours between 9 a.m. and 10 p.m. These were the times at which the library had the most occupancy. Daily profiles of ratios during these hours were then clustered using k-means clustering. These profiles, along with two centroids of recognized clusters are plotted in Figure 7.

Figure 4: The correlation of WiFi connection counts and camera-based occupancy counts.

Figure 5: The frequency distribution of the differences between WiFi connection counts and camera-based occupancy counts during weekdays and weekends.

Figure 6: The hourly range of ratio between WiFi connection counts and camera-based occupancy counts.

Figure 7: Profiles and clusters of daily ratios of WiFi connection counts to camera-based occupancy counts between 9 a.m. and 10 p.m.
Although the ratio of WiFi connection counts to camera-based occupancy counts changed through the time of the day, the pattern of the change was almost the same on different days. Both clusters followed descending patterns that were slightly different only at the beginning and at the end. Based on the membership of days of the week to each cluster, weekends mostly followed pattern 1 with a lower level of ratio at the beginning and the end of the investigated hours. The convergence of all profiles between 3 to 4 p.m. revealed that at the time of peak occupancy, the ratios of different days were almost similar.

In addition to time, the influence of occupancy level was also observed as shown in Figure 8. This plot shows the relationship between camera-based occupancy counts, the ratio, and time. Based on this plot, as the level of occupancy grew, the ratio experienced less variation, and finally, it converged to a steady value (i.e. almost 1.5) during the time of peak occupancy. One of the reasons for a more stable ratio during the peak time might be that all stationary devices were actively connected during these hours due to the high level of occupancy. Therefore, the part of WiFi connection counts’ deviation from actual occupancy counts that can be attributed to stationary devices is almost stable during these hours on all days. In the same plot, a high variation in ratios was identified at the lower level of occupancy (i.e. camera-based occupancy counts ranging between 0 to 500). Despite this variation, two distinct clusters of ratios were recognized at this level of occupancy including (i) a cluster of ratios with a lower average, mostly happening during the nighttime (between 6 p.m. to 10 p.m.); and (ii) a cluster of ratios with a higher average, mostly happening during the day time (between 9 a.m. to 12 p.m.). Since the level of camera-based occupancy counts was similar between these two clusters, the difference between ratios of these two clusters resulted from the difference in their WiFi connection counts level. This means WiFi connection counts were lower in the first cluster and higher in the second cluster. This might be due to the operation of more stationary devices in morning hours, compared to night time when offices are closed or more devices are in idle mode. Furthermore, there might be a difference between the behaviour of occupants in terms of the number of devices they usually connect to the network during daytime (before the peak) and nighttime (after the peak). This assumption needs to be investigated on-site through questionnaires or semi-structured interviews.

**Conclusion**

Several former studies used ground-truth data to validate using WiFi connection counts for occupancy count estimation and prediction. However, due to the limited duration of collected ground-truth data, they typically reported a single (fixed) value as the ratio of WiFi connection counts to actual occupancy counts. However, the temporal variations in this ratio and the effect of occupancy levels remained an unanswered question. This study addressed this gap by collecting longitudinal ground-truth data via camera-based occupancy counters for validating the correlation between WiFi counts and actual building occupancy counts. This can improve the accuracy of occupancy prediction for optimizing buildings’ system operation. In this regard, clustering and other statistical analysis techniques were applied to firstly identify stationary devices, and then investigate the correlations between WiFi connection counts and occupancy counts. Finally, the influencing factors on the temporal behavior of this ratio were identified. The developed methodology was implemented in a library building and showed WiFi connection count is an accurate indicator of occupancy counts. The case study results showed the possibility of extracting the number of stationary devices without invading occupants’ privacy by accessing information such as their MAC (Media Access Control) addresses.

Subtracting the stationary devices from WiFi connection count data only showed a slight improvement in the correlation with occupancy counts. The results also showed that the ratio of WiFi connection counts to camera-based occupancy counts is significantly affected by the time of day and occupancy level. Therefore, considering a fixed value as the ratio of WiFi connection counts to occupancy counts at different hours and occupancy levels can result in inaccurate interpretations, especially if WiFi counts are used as a proxy of occupancy for the operation of building systems such as ventilation.

Despite the significant findings of this study, some limitations should be addressed in future research. Although visitors were obliged to enter and leave the investigated floors from the two main gateways that were covered by the camera counters, staff may have used other access points which influenced the accuracy of these counts. Moreover, although some repetitive daily patterns of ratio were identified for the investigated building, the applicability of these findings to other buildings was not studied as part of this work. Considering that previous studies identified different ratios in even similar space types, these analyses should be performed in other buildings with similar or different space types in order to investigate the extent to which similar patterns can be observed. Future work should also focus on developing models to predict actual occupancy counts based on
historical WiFi traffic data, which can be used to adjust the operational schedules of various building systems according to their unique occupancy patterns.

Acknowledgment
This research is supported by the research funding provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors acknowledge the support of three teams in this study including the Instructional and Information Technology Services (IITS) team of Concordia University by providing the WiFi connection count data; the Information Systems and Technology team of Concordia University by providing access to the camera-based image recognition occupancy counters; and the Security and Facilities Management departments of Concordia University by providing the drawings of the building.

References


