Measuring building occupancy through ICT data streams

Bianca Howard
Department of Civil and Environmental Engineering
Imperial College London
London, SW7 2AZ
United Kingdom
b.howard@imperial.ac.uk

Nilay Shah
Department of Chemical Engineering
Imperial College London
London, SW7 2AZ
United Kingdom
n.shah@imperial.ac.uk

Salvador Acha
Department of Chemical Engineering
Imperial College London
London, SW7 2AZ
United Kingdom
s.acha@imperial.ac.uk

John Polak
Department of Civil and Environmental Engineering
Imperial College London
London, SW7 2AZ
United Kingdom
j.polak@imperial.ac.uk

Keywords

offices, user behaviour, building technology, occupancy

Abstract

Understanding building occupancy is becoming important for sustainable building operations. Specifically measuring and predicting building occupancy are required for demand controlled ventilation, dynamic scheduling of occupancy states, as well as predicting disturbances in model predictive control schemes. Proposed methods for directly or indirectly measuring occupancy have often required installation of numerous additional sensors. However, data collected through information and communication technology (ICT) systems have the potential to provide the same information without the need for additional infrastructure. This work accesses how well various ICT data streams can reflect building occupancy counts as compared to a commercial occupancy counter.

Three separate ICT data sets reflecting Wi-Fi connected mobile devices, desktop computer activity, and security access control counts were gathered for an office building in London. These data sets were used to develop ICT based estimates of occupancy and were then compared to values reported by a commercial occupancy counter for a week in December. The comparison showed that each of the ICT data sets reflects the overall pattern of occupancy usage. The magnitude of the counts, however, differed from that of the commercial counter. The measurements show that the computer activity data severely under estimates occupancy, upwards of 40 % during peak hours. The Wi-Fi connected mobile devices and repaired security access control counts, however, consistently report occupancy counts within the error of the commercial counter during high occupancy periods, indicating that these ICT data sets could be used to measure building occupancy.

Introduction

Energy efficiency gains in buildings primarily result from minimizing, at appropriate times, losses and gains from the building fabric as well as increasing the efficiency and minimizing the use of appliances, lighting and heating, ventilation and air conditioning (HVAC) systems. Minimizing the use of building equipment requires occupants to change their behaviours or for building systems to sense and learn occupant behaviours and react accordingly. Here we focus on the latter.

Gains in automation have been made in the area of lighting, where passive infrared sensors determine the occupancy state of room or corridor allowing the lighting system only be used when the space is occupied. This has been shown to reduce energy consumption for lighting up 86 % depending on the space use type [1]. Achieving the same reductions in practice for building HVAC systems has not been as easy, as the dynamics of the systems are more complex.

This has resulted in a large body of research literature around modulating building HVAC systems to provide dynamic levels of fresh air (called demand response ventilation) and change temperature set points based on room occupancy status. Additionally advanced control schemes aimed at increasing the efficiency of building plant services often require accurate prediction of building occupancy. Simulations [2] have shown that these strategies can reduce HVAC energy consumption by

23 %. Key to large scale use of these approaches, however, is determining how to best measure occupancy in heterogeneous building environments.

Researchers have looked into various methods for measuring occupancy. A survey of occupancy measurement in commercial buildings [3] present the following as possible occupancy measurement systems: carbon dioxide detection, passive infrared devices, ultrasonic detection systems, image detection systems, sound detection systems, electromagnetic signal detection (RFID) systems, building access control systems, and computer activity detection systems. A recent review [4] describing state-of-the-art data acquisition technologies for occupant behavoir modelling also include wireless sensor networks capturing a slew of indoor environmental factors and detecting occupants through mobile devices connected through Wi-Fi. While each method has its benefits and drawbacks, a major delineator is the cost of additional infrastructure and maintenance for the building facility. To increase the viability of these systems in buildings, the solutions should aim to be as simple as possible.

A promising direction are those occupancy detection methods that rely on systems currently deployed in buildings, requiring little adjustment or no adjustment. The only detection methods previously mentioned that fit that criteria are those detection systems based on information and communication technology (ICT) systems: Wi-Fi connected mobile devices, access control logs, and computer activity data.

In this work, the aim is to answer a fundamental underlying question: How well can these ICT datasets measure occupancy with respect to a commercial occupancy counter?

Before continuing, it is important to define what is meant by occupancy. Occupancy can be defined at various levels such as presence, location, count, activity, identity, and track [3]. In this work, occupancy is defined as presence (occupied or unoccupied) and count (number of people) at the building scale.

To evaluate the ability of these ICT datasets to measure occupancy, an office building in London was used as a case study. The remainder of the paper discusses the case study building, the commercial occupancy counter, the methods used to derive occupancy counts from each of the ICT data sets, a discussion of the results and general conclusions.

Case Study Building Description

The case study building is the administrative hub of a University. There are approximately 350 people assigned working space in both open plan and closed offices within the four floors of the building. The building also contains three meeting rooms capable of accommodating 60 people in aggregate. As it is the administrative hub, many employees enter and exit the building throughout the day to go to other meetings throughout the campus. The building also houses secure space to store bicycles for those commuting to work.

Access is gained to the building through two entry/exit points: the main entrance and the rear bike rack entrance. The main entrance is equipped with entry gates that release upon the presence an ID card with appropriate permissions. For the rear bike rack entrance, presenting an ID card to the reader unlocks a set of doors.

Within the building, there are 321 desktop computers equipped with a power management software to track their power consuming activities. Specifically the software reports how many minutes a computer was active, the computer was on, and the monitor was on.

Lastly to access the wireless network, the building is equipped with eight wireless access points, two on each floor of the building. Users can connect to the University's internal network or also to a more general educational system network.

Commercial Occupancy Counter

Many organizations have an interest in understanding occupancy within their spaces, especially in the retail sector. Therefore several companies have developed sensors to count and track visitors based on various technologies such as thermal imaging, infrared beams, and 3D stereo video. The commercial counters used in this work are 3D stereo video cameras equipped with proprietary image processing software to detect the height, direction, mass and velocity of people. As the cameras have a limited field of vision, to measure building occupancy they must be placed at each entry and exit point to ensure all traffic in and out of the building is recorded. Occupancy is then determined by a cumulative sum of the net flows of traffic. For the case study building, the occupancy counters were placed at the main entrance and the rear bike rack entrance. The counters have a 95 % accuracy on the net counts of people coming in or out of the building.

Overall building occupancy is determined by

$$O_t = \sum_{k=1}^{t} f_{k,s} - \sum_{k=1}^{t} r_{k,s}$$

And the error associated with the measurement is

$$e_t = 0.05\left(\sum_{k=1}^{t} f_{k,s} + r_{k,s}\right)$$

where 0_t is the occupancy at the end of time period t, e_t is the error at time t, $f_{k,s}$ is the number of people measured entering within time period k from entry s, and $r_{k,s}$ is the number of people measuring exiting within time period k from entry s. The software accompanying the commercial counter allows for reporting of occupancy at the end of 15 minute intervals. To allow for comparison with the ICT data sets the occupancy was reported at hourly intervals. The total cost of the two sensors including installation was £2,883.

Methodology for Estimating Occupancy from ICT Data Sets

Each ICT data set provides different types of information that can be used to estimate building occupancy. The following sections describe the data sets and methodologies used to provide final occupancy estimates. Each data set was currently available within different areas of the University's ICT department, requiring no additional cost to gather the information.

COMPUTER ACTIVITY DATA

For each desktop computer in the case study building, for power management purposes, the computer activity data is tracked. Specifically, the system logs three different states: the minutes in each hour the computer was active "computer active", the computer was on "computer on", and the monitor was on "monitor on". Occupancy from this data set was defined in equivalent person hours,

$$O_t = \sum_{n=1}^{N} m_{n,t}$$

where O_t is the occupancy at time t in equivalent hours, $m_{n,t}$ is the number of activity minutes for desktop n in time t and N is the total number of desktop computers.

There is still a need, however, to determine which activity status, "computer on", "computer active", and "monitor on" to use for the estimation, as each data set has the potential to miss information. Figure 1 depicts the equivalent person hours calculated for each activity state for a week in September.

Each status provides the general trend of occupancy with sharp increases and decreases in the morning and evening hours. However, each status provides a range of peak occupancy with "Computer Active" and "Computer On" consistently reporting approximately a 75 person difference. During the evening hours, the "Computer On" and "Monitor On" provide high values of occupancy. Given typical office building patterns, it is unlikely that there is a large percentage of employees working through the evening and early morning hours. A more plausible explanation is that the computers and monitors were left on in the evening hours. Therefore for the current analysis, the equivalent person hours estimated from the "Computer Active" occupancy status was used for comparison.

WI-FI CONNECTED MOBILE DEVICES

The case study building has 8 wireless access points, 2 on each level, that allow occupants to connect to the wireless network. For a device to send data through the network, it must first undergo two steps: an authentication and an association. The authentication ensures that the mobile device and wireless network are compatible and the association allocates a wireless access point to a device. Counts of the number of devices authenticated and associated with the network were sampled hourly. Figure 2 depicts the hourly counts of devices authenticated and associated to the wireless network for a week in September. There is minimal difference between two statuses, although there are consistently more devices associated with the network than authenticated. In the current analysis, the hourly counts of associated devices were used as a measure of occupancy, assuming that the devices associated with the network are still physically present within the building.

ACCESS CONTROL DATA

Access to the building is provided by presenting an ID card with appropriate permissions. The access control system records each entry and exit in a data log which contains the direction of entry (IN or OUT), a unique ID for each person, the campus building to which they are assigned, and the location of their office with in that building.

Upon initial data exploration, it was clear that there were missing data points, in that there were times when people would be registered entering the building but not exiting or vice versa. After observation of each entrance and exit, these missing data points can be attributed to the "tailgating" phenomena, where a single person may present an ID but several people may follow behind, resulting in their entry or exit not being logged.

In the data set, this tailgating led to a fictional accumulation of occupancy at the end of the day. For a single weekday, upwards of 80 occupants would appear to remain in the building at midnight. The data log was not usable in this form and required repair using the historical records. The data records

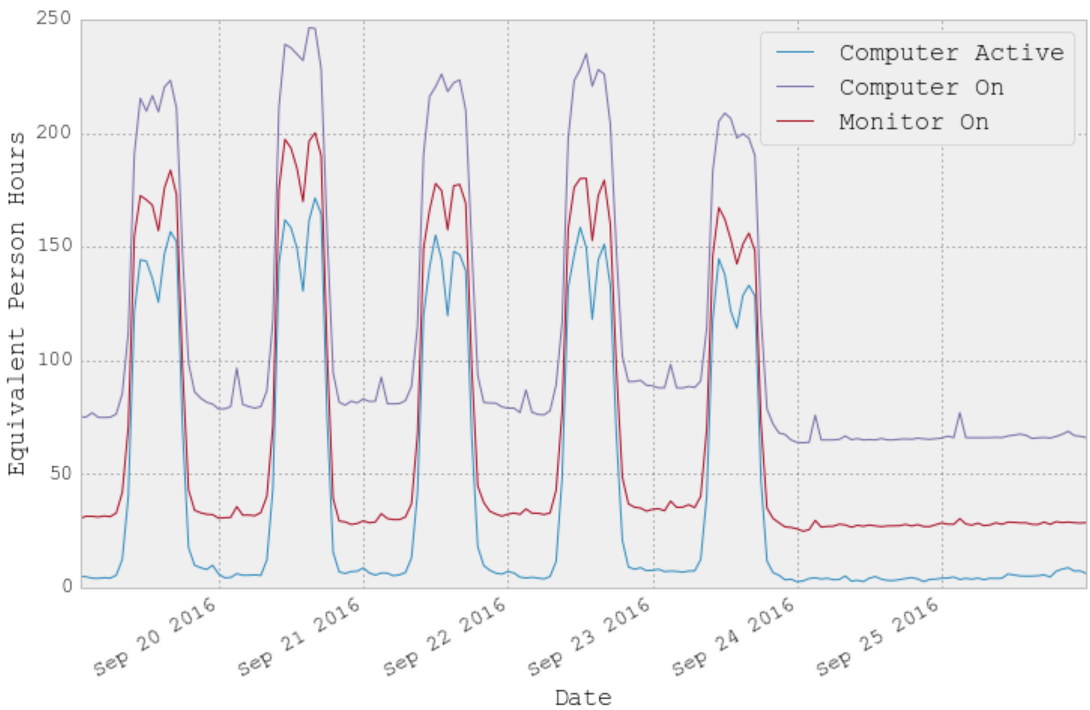

Figure 1. Equivalent Person Hours for activity states "Computer Active", "Computer On" and "Monitor On" for a week in September 2016.

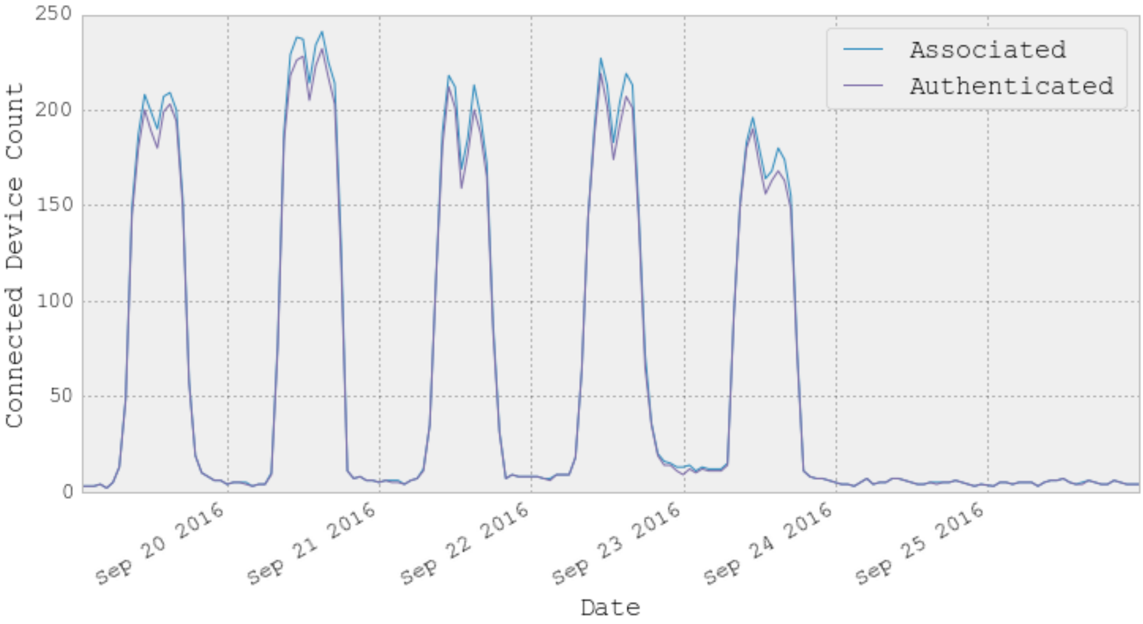

Figure 2. Occupancy count derived from Wi-Fi connected mobile devices.

were repaired by developing assumed entry or exit logs based on typical times individuals entered or exited the building.

Repair Methodology

To repair the data, an imputation strategy was used to repeatedly develop assumed entry and exit points from empirical distributions developed from historical data. Figure 2 shows an example of the imputation strategy for single individual on 3rd October 2016.

A complete data record must comprise of at least one set of an "IN" record followed by an "OUT" record. On this day, the individual is missing an "IN" record that must occur before 12:15:10 (Figure 3a). To develop an imputed record for this gap, all of the times this individual entered the case study before 12:15:10 were collected. If this data set holds more than 10 data points, then a kernel density estimation was performed to develop an empirical distribution (Figure 3b) of this person's entry. A kernel density estimation provides the same information as a histogram, but estimates the probability density function with a smooth function. The kernel density estimate was made with the StatsModels python package with a Gaussian kernel and normal reference bandwidth. This distribution is then sampled to collect an imputed record (Figure 3c). If there is less than 10 historical data points in the gap period, then a new record was sampled from a uniform distribution.

The imputation strategy was performed over the entire data set 100 times. The average of the 100 imputations, averaged to the nearest integer, resulted in the repaired data set. From the repaired data, set the occupancy was defined as

$$O_t = \sum_{k=1}^{t} f_{k,p} - \sum_{k=1}^{t} r_{k,p}$$

where 0_t is the occupancy at time t, $f_{k,p}$ is the number of people entering during time period k from access point p and $r_{k,p}$ is the number of people exiting during time k from access point p.

Figure 4 depicts the average occupancy counts for a week in December from the repaired access control data as well as the full range of estimates from the imputation strategy. The imputation methodology led to a small deviation in the estimates of the hourly occupancy.

Results and Discussion

Figure 5 depicts the occupancy counts from the commercial counters, the Wi-Fi connected mobile devices, repaired access control data, and computer activity data at hourly intervals. Each of the data sets reflects the same overall pattern of occupancy. On weekdays, individuals enter the building between 7am and 9am, a first peak is reached between 10am and 11am, followed by a reduction over typical lunch time hours, a second peak after lunch, and a rapid reduction in occupancy between 5pm and 7pm. On Friday, there is a lower magnitude of occupancy than the other weekdays. On the weekends, the building is practically unoccupied. There are, however, discrepancies in the magnitude of the occupancy counts by each data source.

The computer activity data severely underestimates the number of people in the building. During peak times, the occupancy counts derived from the computer activity data are up to 40 % less than the occupancy measured by the commercial counters.

Making conclusions about the accuracy of the Wi-Fi connected mobile device and access control data, however, is more nuanced due to the large cumulative error inherent in the commercial occupancy counters. In a 24 hour period, over 1500 people can enter or exit the building. This high traffic in relation to the occupancy values leads to the large cumulative error, upwards of plus or minus 100 people by the end of the day.

Figure 6 depicts the difference between the commercial counters and the ICT data sets as well as the cumulative error for the commercial counter for Monday December 12th 2016.

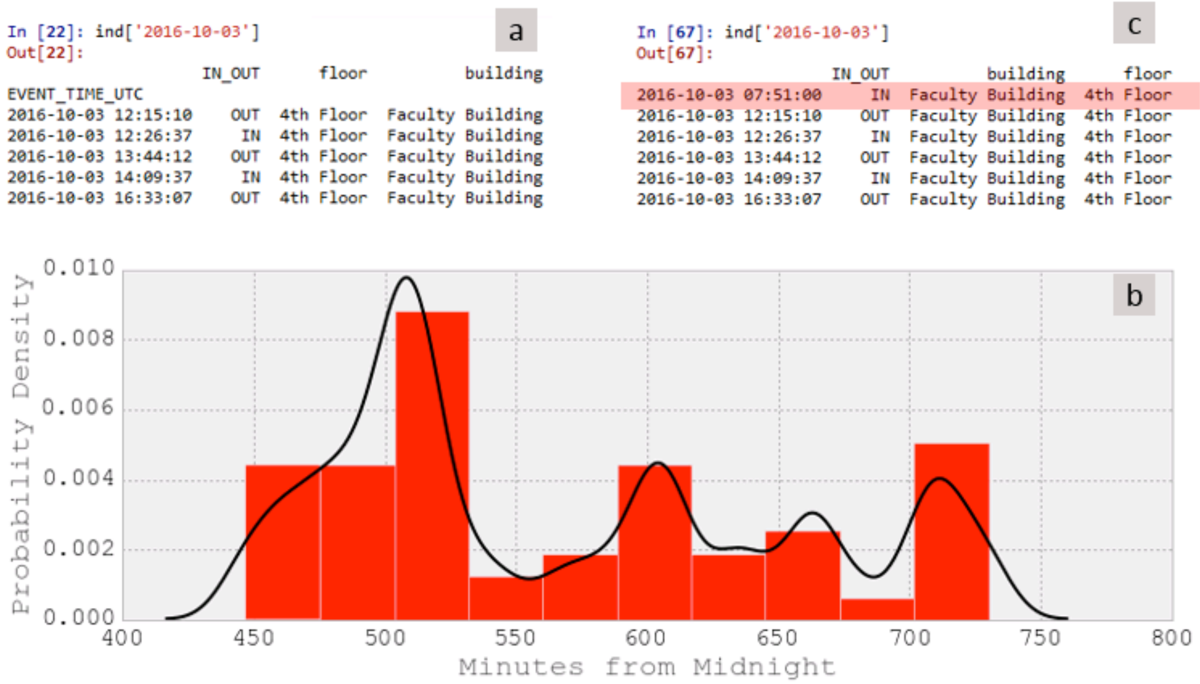

Figure 3. Example of Imputation Strategy. a: Access Data Logs for a single person (X) on October 3rd, 2016. b: Histogram (red bars) and Estimation of Empircal distrubtion (black line) for 'IN' times for person "X" between 00:00:00 and 12:15:10 in minutes from midnight. c: Repaired Access Data Log.

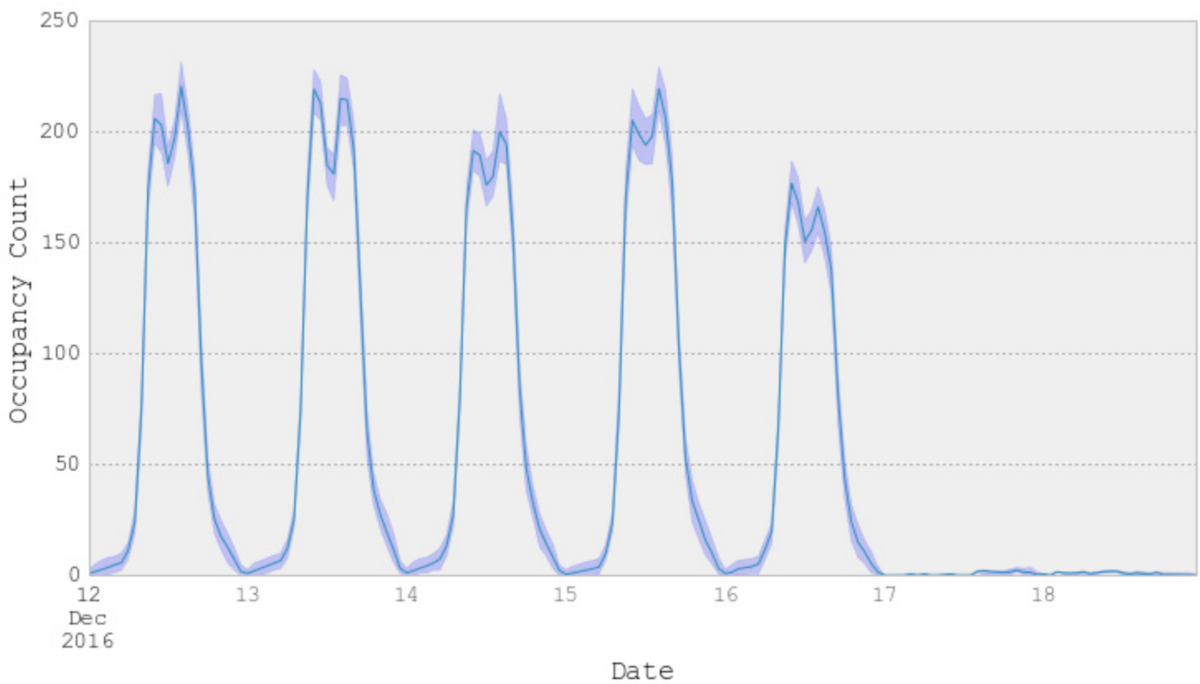

Figure 4. Average occupancy count from repaired access control data. The full range of the 100 simulated repaired values shaded in blue.

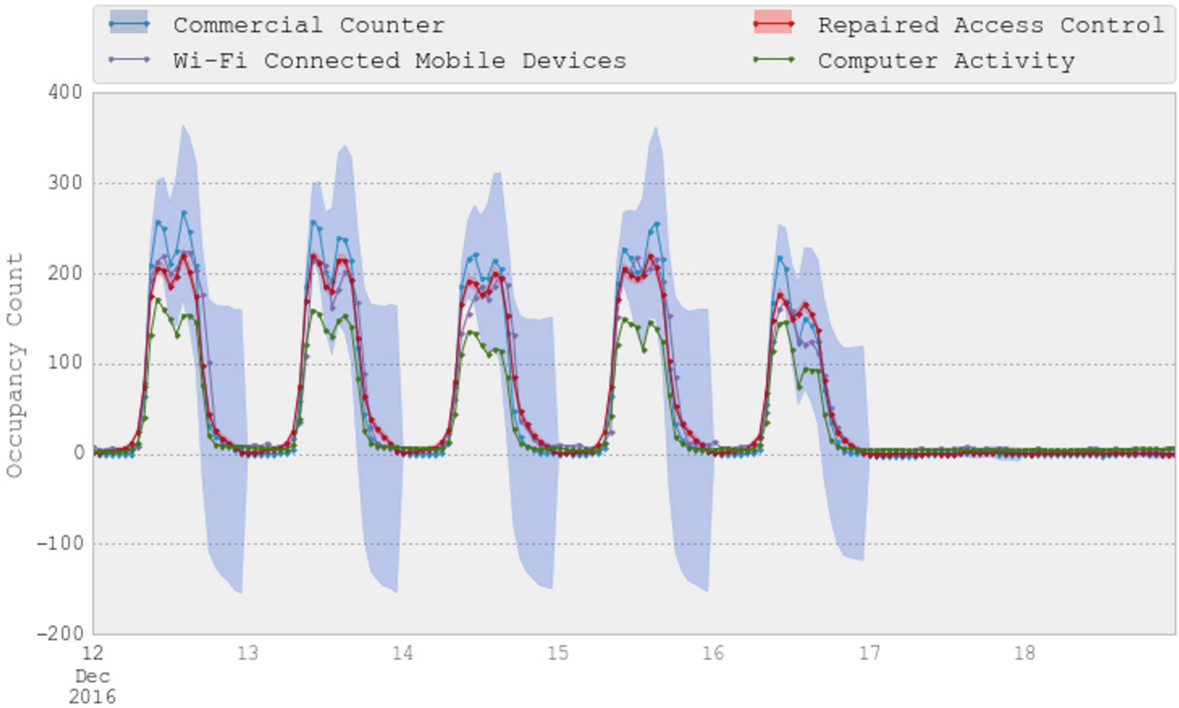

Figure 5. Occupancy counts for the commerical counter (including error in shaded region), Wi-Fi connected mobile devices, repaired access control (including imputation range in shaded region), and computer activity data sets for a week in December.

Table 1. Percentage of Weekday Hours that the ICT datasets are within the error of the commercial counter.

	Wi-Fi Connected Mobile Devices	Repaired Access Control Data	Computer Activity Data
% of Weekday Hours within Error of Commercial Counter	57 %	61 %	43 %
% of Weekday Hours after 6am within Error of Commercial Counter	82 %	87 %	63 %

During peak occupancy periods, the counts from Wi-Fi connected mobile devices and the access control data can underestimate occupancy by 20 to 50 people with respect to the commercial counter. However due to the cumulative error in the occupancy counter, this difference is not significant past 8am for the Wi-Fi connected mobile devices and past 11am for the repaired access control data. In the early morning hours, all of the ICT data sets over estimate occupancy, with each measuring less than 10 people remaining in the building.

Table 1 reports the number of hours the occupancy, as measured by each of the ICT data sets, falls within the error of the commercial counter during the one week period. Considering all hours, the Wi-Fi Connected Mobile Devices and the repaired access control data report values within the error of the commercial counter for the majority of hours, 57 % and 61 % respectively. If one considers only hours after 6am, these two data sets provide values within the error of the commercial counter over 80 % of the time.

Conclusions

The estimated occupancy derived from computer activity data severely underestimates occupancy, upwards of 40 % during peak occupancy times, therefore it should not be used directly as a measure of occupancy. However, in the current analysis, the "Computer Active" status was used as the measure of occupancy which reported the lowest occupancy counts with respect to "Computer On" and "Monitor On". A linear combination of these statuses could provide a more accurate estimate of occupancy.

The occupancy estimates derived from the Wi-Fi connected mobile devices and the repaired access control data provide building level occupancy counts within the error of the commercial counter for the majority of high occupancy hours during the one week test period. While within the error of the commercial counter, there is a consistent underestimation bias. Depending on the desired use for the occupancy measurements, both the Wi-Fi connected mobile devices and repaired access control data could be used as a measure of occupancy.

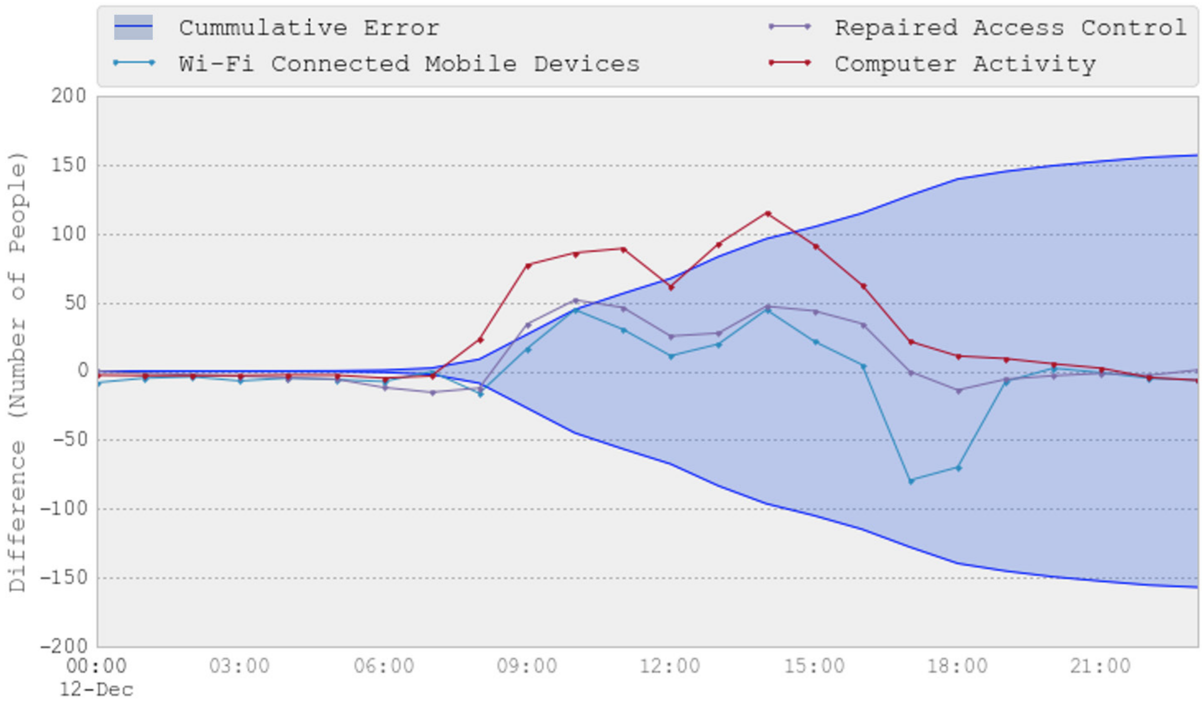

Figure 6. Comparison of difference between commerical counter estimates and ICT data sets as well as the commerical counter error for a day in December. Postivie Difference: Underestimation; Negative Difference Overestimation.

Using the number of Wi-Fi connected mobile devices to estimate building occupancy potentially has the advantage of being able to provide high spatial resolution occupancy data. For an HVAC system, these occupancy measurements could be used to determine the number of air exchanges for demand controlled ventilation or used to predict occupied and unoccupied spaces throughout the building. Attempting to perform the same with the commercial counters would require installation of additional sensors incurring more costs. For example, if floor level estimates of occupancy were desired for the case study building, the total sensor cost would increase to approximately £13,000 to cover each entry and exit. Whereas gathering occupancy counts from Wi-Fi connected mobile devices would require no more additional costs. However further work is needed to determine how the bias in the counts derived from the Wi-Fi connected mobile devices would affect the resulting use cases.

References

[1] X. Guo, DK Tiller, GP Henze, CE Waters, The performance of occupancy-based lighting control systems: A review, *Lighting Research & Technology,* Vol 42, Issue 4, pp. 415–431.

[2] J Zhang, G Liu,RG Lutes, MR Brambley, Energy Savings for Occupancy-Based Control (OBC) of Variable-Air-Volume (VAV) Systems, January 2013, Pacific Northwest National Laboratory, Richland, Washington 99352.

[3] Timilehin Labeodan, Wim Zeiler, Gert Boxem, Yang Zhao, Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation, Energy and Buildings, Volume 93, 15 April 2015, Pages 303-314, ISSN 0378-7788

[4] Wilhelm Kleiminger, Friedemann Mattern, Silvia Santini, Predicting household occupancy for smart heating control: A comparative performance analysis of state-of-the-art approaches, Energy and Buildings, Volume 85, December 2014, Pages 493-505, ISSN 0378-7788.

Acknowledgements

This work has been funded by the United Kingdom's Engineering and Physical Sciences research Council (EPSRC) under the project Future Proofing Facilities Management (Future FM) Reference EP/L02442X/1