DOI 10.20357/B75G6M

Lawrence Berkeley National Laboratory

Harvesting the low-hanging fruit of high energy savings -- Virtual Occupancy using Wi-Fi Data

Callie Clark[1], Anand Prakash[1], Marco Pritoni[1], Margarita Kloss[1], Pranav Gupta[1], Bruce Nordman[1], Mary Ann Piette[1], Michel Kamel[2], Tony S. Semaan[2], Anne Eisele[3], Dotty Hage[3], and Pat Flannery[3]

[1]Lawrence Berkeley National Laboratory, [2]MelRok LLC and [3]Pomona College

Energy Technologies Area
August 2020

Harvesting the low-hanging fruit of high energy savings -- Virtual Occupancy using Wi-Fi Data

Callie Clark, Anand Prakash, Marco Pritoni, Margarita Kloss, Pranav Gupta, Bruce Nordman, Mary Ann Piette, Lawrence Berkeley National Laboratory
Michel Kamel, Tony S. Semaan, MelRok LLC
Anne Eisele, Dotty Hage, Pat Flannery, Pomona College

Abstract

Approximately 20% of primary energy consumed in the U.S. is attributed to HVAC use. Ideally, HVAC operation would be driven by actual building occupancy, but lack of reliable occupancy information often results in the use of conservative static schedules. This disparity is even more pronounced in a college campus, where the function of each space differs by building (classrooms, offices, libraries) and the class schedules change frequently -- every semester, day of week, and hour. While several research papers propose the use of counts of the Wi-Fi connections (e.g., phones, computers) as a proxy for occupancy, few real-world implementations exist. This paper describes the development and deployment of an open-source Wi-Fi-to-Occupancy software library in 65 buildings of a college campus, and the planned integration with the building energy management and control system at the building scale. Over a year of Wi-Fi data was gathered into distinct academic periods, including fall and spring semester, academic breaks, and summer sessions. Patterns such as students moving between classrooms, closing laptops before exams, etc., can be visualized from the data. Approximating occupancy from Wi-Fi data presents challenges which we address in this project -- for example, identifying static devices, or estimating the ratio of devices per person. Utilizing real-time occupancy data to inform optimal HVAC schedules and ventilation rates creates the potential to identify and reduce energy waste. Other potential applications include forecasting occupancy, and using Wi-Fi data to predict peak demands. Finally, the paper discusses how to easily scale these tools to other buildings.

Introduction

People require services from building -- lighting, computing, air conditioning, heat, ventilation, hot water, communication, and others; these services consume energy. When buildings are empty, these services are not needed and the energy use should decrease correspondingly. Studies have demonstrated that occupancy-derived HVAC schedules reduce runtime of the HVAC system by 3-37% (Trivedi 2017) and occupancy-driven HVAC systems have produced energy savings up to 42% (Erickson 2011). Despite the importance of knowing whether a building or sub-zone has people in it, energy information systems for buildings do not typically track occupants' presence, or location (Price et al. 2015). Conventional occupancy sensors (e.g. infrared, ultrasound, or CO_2) have been expensive to install and to connect to a central information system, especially in an existing building, and have often had questionable reliability. Recently, cameras have shown a high accuracy when used for occupancy detection (Petersen, 2016), however they are costly, may trigger privacy concerns (Wang, 2019; Chen, 2018, Zou, 2017), and require significant computational resources, especially when the data are anonymized.

Wi-Fi network data have been proposed as a means of estimating occupancy in buildings (Pritoni, 2017), however, most of the previous research on occupancy detection focused on algorithms (Wang 2019; Zou, 2018, Shen et al., 2017) or advanced control applications (Wang 2019, Erickson 2011). The use of advanced algorithms to predict occupancy from Wi-Fi both increases the difficulty of implementation and requires training; an academic schedule with short periods of distinct behavior patterns adds further complications (Wang 2019, Shen et al., 2017). A few methods address the issue of changing patterns through continuous integration of new data into the model to capture recent trends (Zou 2018, Trivedi 2017), but they do not address the nature of distinct periods and events. Other research correlates Wi-Fi data to occupancy (Ouf 2017, Rashi 2015), but lacks discussion on the data process required to transform raw data into an occupancy proxy, or the study requires personal information and Mac Addresses to map Wi-Fi to Occupancy (Balaji 2013, Zou 2018, Ardakanian 2018). A startup company[1] based at the University of British Columbia uses Wi-Fi data to determine occupancy and adjust HVAC controls accordingly; they have several pilot projects underway, but results have not been made publicly available. Overall, the literature lacks papers describing installations of Wi-Fi sensing at scale and the process of utilizing raw Wi-Fi data to create occupancy-driven applications.

This paper describes the development and deployment of an open-source Wi-Fi-to-Occupancy software library in 65 buildings at Pomona College, and how this intelligence is integrated with the building energy management and control system. Over a year of Wi-Fi data were gathered into distinct academic periods, including fall and spring semester, academic breaks, and summer sessions. Occupancy based on Wi-Fi counts was integrated with the Automated Cloud-based Continuously Optimizing Building Energy Management System (ACCO-BEMS)[2], a building energy management system developed by MelRok LLC (MelRok), which continuously and automatically monitors the performance of a building's energy systems, and optimizes and dictates operational adjustments. By gathering information on the real-time occupancy of a building, energy use can be carefully matched to provide only the needed services while the building is being used. The California Energy Commission (CEC) sponsored this research to deploy ACCO-BEMS in 12 buildings (a subset of the 65) of varying activity types, such as classrooms, offices, a data center, the campus student center, and a music/performing arts center on the Pomona College campus in Claremont, California, part of Claremont Colleges Consortium. The project lead is the Zero Net Energy (ZNE) Alliance .

These sections follow below: Data Collection, Processing and Analysis; Applications; Discussion; Conclusions and Next Steps; the paper closes with Acknowledgements and References.

Data Collection, Processing and Analysis

Data collecting

The Pomona College campus Wi-Fi network supports the wireless connectivity needs of about 1,700 students, faculty, staff, and visitors. More than 1,000 Wi-Fi access points (APs)

---

[1] Sensible Building Science https://sensiblebuildingscience.com/
[2] For more information: https://www.znealliance.org/acco-bems

report to a central Cisco Wireless Local Area Network (WLAN) controller. The Wi-Fi infrastructure is shared with other colleges in the consortium.

As part of this CEC-funded project, the research team developed an open-source software application, called COUNT (Counting Occupants Using Network Technology) to collect occupancy information using the Pomona Wi-Fi network infrastructure. The architecture of this software is shown in Figure 1. The software is developed in Python 3 and publicly available under a modified BSD license. To comply with cybersecurity requirements from the campus IT department, this software was deployed on a Virtual Machine (VM) hosted locally on the campus network. This VM has the necessary permissions to communicate with the WLAN controller and can push data to an external data store.

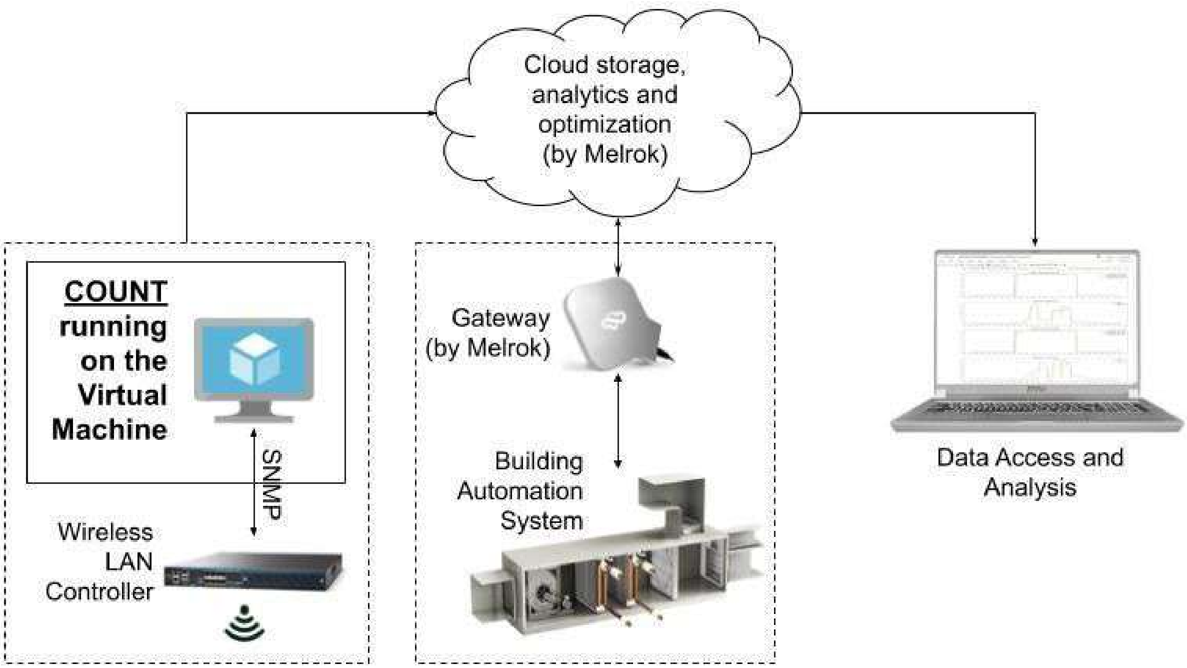

Figure 1: Architecture of the COUNT software installed in Pomona Campus.

The software has two components: one module queries the campus Wi-Fi controller using the Simple Network Management Protocol (SNMP) (Stallings, 1993) and saves the collected data to a local database which serves as a data buffer to prevent data loss. The SNMP query is used to query the list of devices connected to each access point at a given moment. This is the "raw data" that is pushed to the local data buffer. Before storing it, the COUNT software ensures that all personal information, such as the MAC address of the user device has been removed and that the data is completely anonymized. The second module reads this data from the local buffer and tries to push it to the external cloud data store. Upon a successful data push, it removes data from the buffer. Otherwise, the second module detects an unsuccessful data push to the database and it keeps the data in the buffer. The software attempts to push the data again, once the network connectivity is reestablished. The decoupling of the software modules prevents loss of data in case of loss of network connectivity between the VM and cloud database. As the modules are set up via Linux cron jobs, they terminate after execution and restart periodically.

The system outputs the raw sum of connected devices (e.g., laptops, phones) for each building. While it is possible to obtain more granular data, such as the count of devices for each access point or floor, locating the exact position of a device in a building and associating it to the correct HVAC zones is challenging and difficult to scale. For this reason, the research team used building-level occupancy information in the applications described below. Scalability challenges related to using more granular data is addressed in the discussion session. More than a year of Wi-Fi data were gathered using this platform and were used for the analysis and applications below.

Automatic data cleaning

The raw data collected from the Wi-Fi system need to be pre-processed to provide reliable occupancy-proxy values, necessary to any application (e.g., schedule adjustments, measurement and verification, etc.). In order to create an automatic procedure to pre-process the data, the project team conducted extensive data exploration and identified several issues with the raw data, including irregular sampling frequency, null and zero values, outliers, and static devices that are always connected. The algorithm shown in Figure 2 was designed to address these issues and was implemented using a script in Python 3.

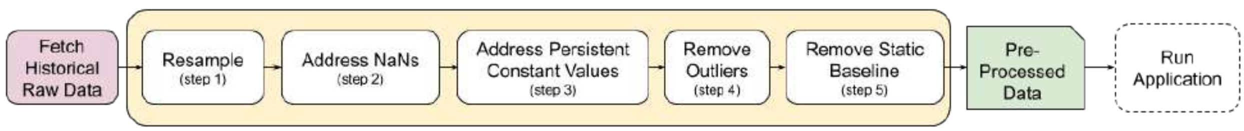

Figure 2: Data pre-processing algorithm (yellow), its inputs (red) and outputs (green).

The output of the algorithm is the pre-processed data that is used by other applications. Steps 1 through 4 are necessary to address values in the data that do not represent device count, and step 5 removes devices that are not associated with people so that the data serves as a better proxy for a count of occupants.

Step 1. Address Irregular Sampling Frequency. The COUNT software currently collects data every minute for the Cisco system. However, the sampling time is not exact due to delays caused by network traffic and computation time of the virtual machine which hosts the software. All the data queried from the central Wi-Fi controller are time-stamped only when the data collection process for all the access points is completed. As a result, the actual difference in time between two consecutive scans is not constant. To address this issue, we resampled the data at consistent intervals. The data were aggregated using the maximum value of each time resample interval (e.g., 5 minutes) to maintain whole numbers and any period without a sample point is assigned a NaN (Not-A-Number) value.

Step 2. Address Null and Zero values. NaNs are present in the data for several reasons: 1) Data missing for an entire building (e.g., during the transition between the previous Aruba Wi-Fi system to the current Cisco system); 2) COUNT system offline; 3) Access points disconnected from power or Wi-Fi network; 4) Access points working correctly, but no clients connected to them; and 5) NaN values generated during the resampling step. Cases one and two were left as NaN values; case three is addressed in step 3. All the instances of case four were converted to zero-values, and case five was set to a zero value unless the data points directly preceding and following were non-zero values. In that case a linear interpolation was applied to the values.

Step 3. Identify Constant Values. Data that remains constant over a long period of time need to be flagged as a possible anomaly. We set the threshold at three days to account for the lack of occupancy change over the weekends and the different class schedules on Fridays. All instances of uninterrupted constant data for a period exceeding the threshold are identified and the values are converted to NaNs. This process accounts for the case three identified in Step 2 of the anomaly detection process. After the NaNs and zeros are sorted, the sampling rate is standardized and the constant values are dropped.

Step 4. Detect and Remove Outliers. The Wi-Fi data approximates occupancy trends. As stated previously, on a college campus, the occupancy trends vary greatly based on the academic calendar. In order to identify and remove outliers, these trends need to be considered, since an outlier for a period with low occupancy (e.g., 100 occupants in a building during summer break) may not be identified if the same criteria is applied to the building during periods of medium occupancy (e.g., during the academic season). Five academic periods were identified for Pomona College: spring term, fall term, finals week, summer session, and all breaks. Within the identified periods, the daily occupancy profile also varied at a daily level. Depending on the function of the building, the daily occupancy profile either varied according to class schedules or with respect to the working week schedule. Examples of the profiles are depicted in Figures 3 and 4. Figure 3 shows the average daily device count profiles for the five academic periods in Building 1, which houses offices and classrooms. Fall and Spring show the characteristic dip in occupancy around lunch time.

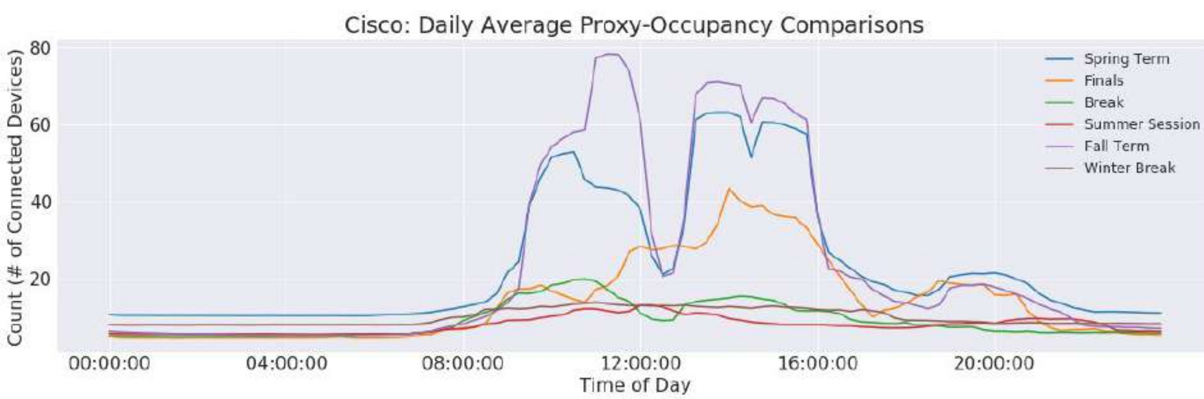

Figure 3: Average occupancy patterns during different academic periods for the Building 1

Figure 4 shows the device count daily profile for the Building 1 during the Spring Semester and it aggregates (1) Monday (M) and Wednesdays (W); (2) Tuesdays (T) and Thursdays(R); (3) Fridays (F); and (4) weekends separately, reflecting patterns in class schedules. Comparing the dotted blue line and the daily profiles, it is apparent that aggregating all weekdays creates an inaccurate occupancy daily profile both with respect to magnitude and timing. The impacts of this disaggregation on both HVAC energy use reduction and cost savings likely will be substantial.

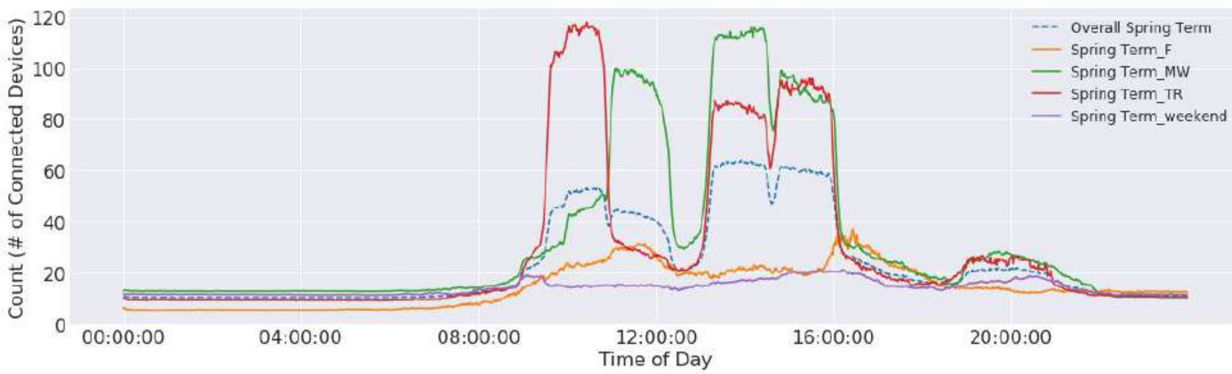

Figure 4: Average occupancy patterns during different days for the spring term for Building 1

The research team designed an outlier detection method to accommodate the variability in daily and seasonal occupancy trends.

Step 5. Remove static devices. At this point in the process, all identified anomalies have been removed but the data may still show connected devices, even when the building is unoccupied. We call these devices "static." Static devices include things that are always connected to the Wi-Fi network such as desk phones, printers, and Internet of Things (IoT) devices such as plug controllers, thermostats, etc. As most of these devices are not associated with the presence of an occupant, their inclusion would distort the estimate of occupancy.

A detailed exploration of the data revealed that the number of static devices is not constant over time. In fact, the number of static devices tends to vary based on the different time periods, and does not seem to be correlated to the day of the week. For Pomona College, the variability is significant, with some buildings having less than half the number of static devices during the summer, when compared against other periods.

To address this variability, the research team created a model to dynamically determine the number of static devices per building. The model groups the data to assess each of the five academic periods separately, and then filters by the hours between midnight and 4 AM. The frequency of the device count is used to fit a distribution. A value from this distribution (e.g., the median) is then used to estimate the number of static devices in each building for each academic period. Using the median value in the model, the static device count as a fraction of the peak hour of occupancy ranged from 4.6% to 16.9% across the buildings in the Fall semester. The static device count is then subtracted from the base Wi-Fi count. Picking higher values (e.g., third quartile) for static devices may under-estimate actual occupancy, but also allows for more aggressive control strategies. Lower values for static devices reduces the chances of undercounting occupants, but may lead to more conservative control actions. This approach allows energy managers to change this adjustment as needed.

Over-/under-counting evaluation

The model discussed above produces an estimated device count which does not directly correspond to the number of occupants. For instance, a single person may have multiple connected devices or may carry no connected devices. The discrepancy between device count and human occupancy can create scenarios of under-counting or over-counting actual occupants.

To understand this error we investigated the relationship between devices and occupancy count using carbon dioxide sensors, class schedules, and physical counting of building occupancy.

Comparison with course schedules. The team collected the list of classes in Building 1, the weekly schedule, and the number of students enrolled in each course from Pomona College. From this schedule, the researchers calculated the expected number of students that would be present in the building at any point in the day, for each day of the week (Monday-Friday). Differences between these values and the exact count include:

- not all students who are enrolled might attend a class session
- students or others who are not enrolled may attend a class session
- the enrollment could change over the period of the semester and we have only a snapshot
- students could come early for a class session or remain in the building after their class session has ended
- some occupants are present in the building for reasons other than to attend a class session (people in offices, common spaces, etc.)

The blue line in Figure 5a represents the total number of enrolled students that are expected to be in Building 1 on a Monday and the orange line represents the average number of devices connected to access points in the same building across all Mondays in the period. It can be seen that changes in occupancy from the two datasets coincide, although the magnitudes do not. The factors listed above and the range of devices per person could explain this difference in magnitude.

Comparison to ground truth counts. On Sept 16, 2019, LBNL visited Pomona College and conducted ground-truthing of the occupancy count to compare to the Wi-Fi data at Building 1. Ground-truthing refers to directly observing the relevant data. In this case, it required the counting of people entering and leaving Building 1, and enabled a comparison of actual occupancy with the Wi-Fi data meant to indicate occupancy. At 7:30 am the team began monitoring the two access points on the east and west sides of the building. Researchers saw a facilities worker physically unlock the building; Building 1 also has a key card lock that faculty and staff can use to access the building outside of working hours, so there is no certainty that the building was unoccupied when the team began recording. Utilizing a google form accessed via phones, the LBNL team continuously recorded time-stamped entries of the number of people entering and exiting the building from 7:30 until approximately noon. Figure 5c shows the number of people in the building ("occupancy" as estimated from the visual counts) and the number of devices connected to the access points ("devices_adj"). The device per person ratio over the time where the ground-truthing occurred averages about 1.2 devices per person, although it is apparent in Figure 5c that the number of devices vary when the number of people inside the building remains constant. This variation shows the dynamic nature of the multiplier, and highlights the difficulty in evaluating whether a constant multiplier accurately represents the ratio of devices per person. Furthermore, around 11:45 am the count of devices suddenly drops without a corresponding change in actual occupancy. The research team speculated that this could be due to people closing laptops to leave the building or before an exam.

Comparison with a Carbon Dioxide Sensor. A test was performed with one calibrated CO2 sensor installed in a large classroom in Building 1. We used the CO2 level measurements from this sensor to compare its variation with the variation in number of connected devices to the nearest Wi-Fi access point(s). Since the classroom is isolated and large, the assumption is that the only devices connected to the access points are the ones in the classroom. Figure 5b is an illustration of this comparison with the CO2 level in blue and the number of connected devices in red. It is encouraging to see that the peaks and dips in CO2 level and the number of connected devices nearly coincide, although the relative magnitude of these peaks/dips are different. The correlation coefficient between the carbon dioxide levels and device count, calculated for a week in November of 2019, is 0.57. It should also be noted that the access point might also count devices belonging to occupants who do not contribute to the CO2 levels measured by the sensor, for instance people who sit in nearby rooms.

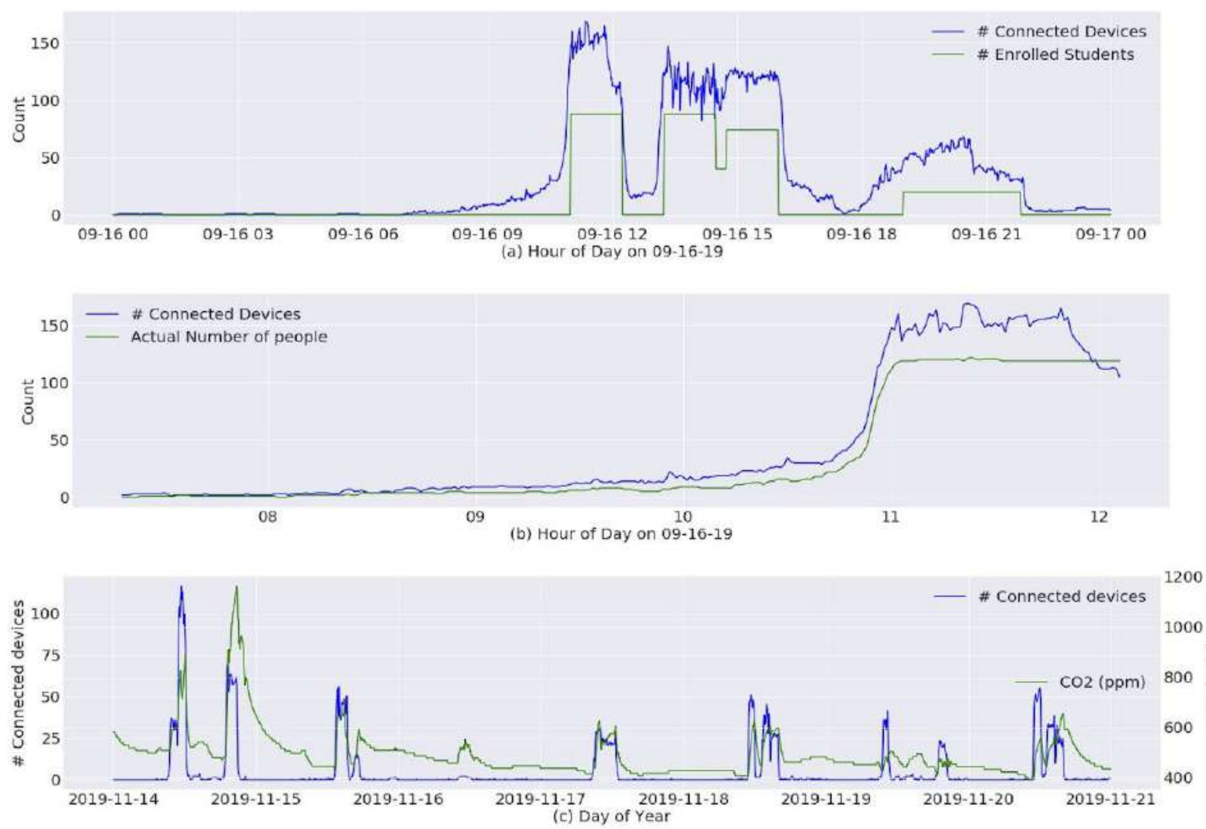

Figure 5.Validation of occupancy counts. a) comparison with class schedules; b) ground truthing by counting people; c) comparison with CO2 sensor

Overall, Wi-Fi trends correlate well with other indicators of occupancy such as class schedules, CO_2 sensors, and manual counts of people. While these tests are not conclusive, they provide some confidence on the usefulness of the Wi-Fi approach. Furthermore, for buildings without sensors, Wi-Fi offers an untapped opportunity to compare occupancy profiles with scheduled building operation.

Applications

Understanding the occupancy patterns of a building enables the application of building-specific HVAC schedules and ventilation rates to reduce energy waste, both for general trends and real-time observations.

Occupancy-aware HVAC schedules

Similar to other universities or campuses in our experience, Pomona facilities set their HVAC schedules to operate during generic weekday business hours and to turn off for weekends and some predetermined holidays. However, the actual schedules are not always up to date. For example some holidays were not programmed into the schedule and several scheduled events outside of the typical operation continued to be scheduled even after cancellation. Additionally, utilizing the same hours of operation for every building, day of the week, and season is a non-optimal oversimplification.

To address these issues, the research team created an application that utilizes the occupancy data to create occupancy-aware HVAC schedules and do dynamic event identification.

- The occupancy-aware HVAC schedule model inputs consist of the enrollment schedule, building type, two weeks of Wi-Fi data, and dates of period, and then outputs the on/off HVAC schedule for each day of week. The resulting schedule accounts for the weekly class or work schedule of the occupants.
- The dynamic event identification model inputs consist of the Wi-Fi data and expected occupancy profile, and produces an event classification, returning yes or no. This model detects and corrects discrepancies between the occupancy-informed (static) HVAC schedule and what is occurring in real time. For example, if a holiday occurs that is not in the schedule and the real time occupancy is below the expected occupancy by a certain threshold, then this event is flagged **Unoccupied** . An example can be seen in the figure below where Thanksgiving break was not programmed as a holiday and the building operated as usual even when the campus was closed. Conversely, if there is a one-time meeting or study session scheduled and the real time occupancy is higher than the expected occupancy by a certain threshold, an event will be triggered **Occupied**.

The facilities department is in the process of adapting schedules based on this new information.

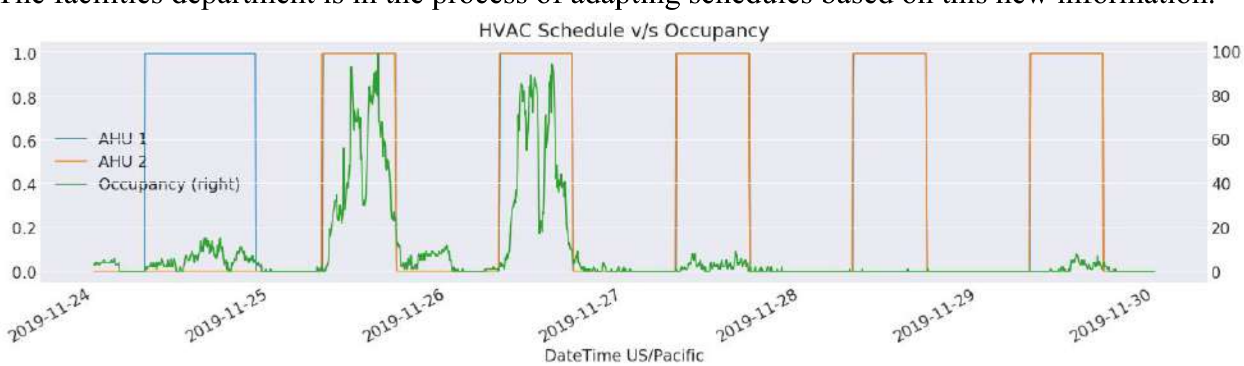

Figure 6: Occupancy and HVAC schedule at Building 2 before (Sunday-Tuesday) and during the Thanksgiving break (Wednesday-Friday)

Occupancy-aware Ventilation Rates

According to ASHRAE 90.1, building ventilation should be based on both the size of the building and the number of occupants. However, most buildings do not have occupancy data so the ventilation rates are usually determined with occupancy-based design parameters, or rule of thumb. With the Wi-Fi based occupancy information, we are able to calculate a continuously varying recommended ventilation rate, as shown by the green line in Figure 7.

To evaluate the discrepancy between the implemented ventilation rates and the recommended ventilation rates, the project team calculated the actual ventilation from the supply air rates at the variable air volume box (VAV[3]) level and outdoor air (OA) damper[4] position in a air handle unit (AHU[5]) and compared these values to the code-compliant ventilation rate. Figure 7 shows one of the buildings that was found to be under-ventilated due to OA dampers in two AHUs being closed almost all the time (until 2019-12-05). This problem was corrected after we reported this issue. The new behavior of the damper is illustrated in Figure 7 (after 2019-12-06) and the corresponding change in ventilation is shown. The correction to this problem led to an increase in air quality and the next steps to this work will involve implementing a dynamic change in damper positions with occupancy.

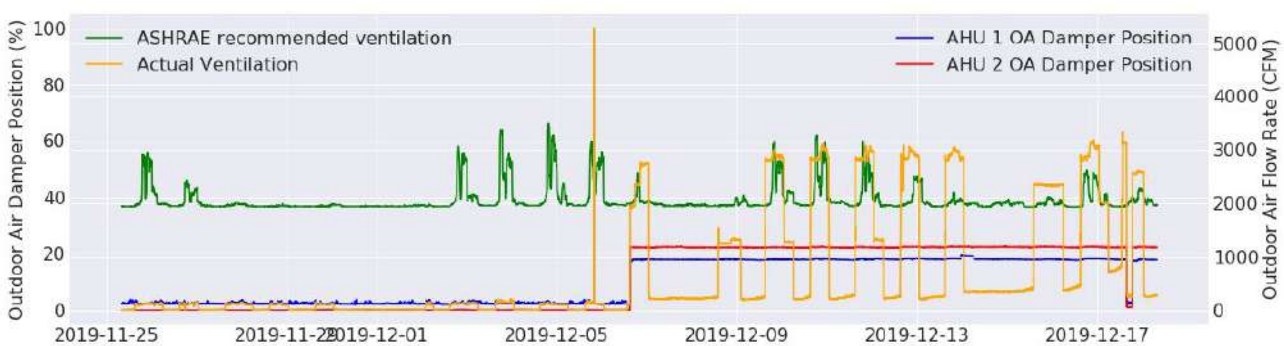

Figure 7: Outdoor damper positions of several AHUs in Building 1 before and after correction and the continuously varying recommended and actual ventilation rates

Improved peak demand forecast with occupancy data

Most utilities include demand charges and time of use prices in their tariffs for commercial buildings. Thus, forecasting building load is useful for implementing strategies (e.g., precooling) that minimize customer bills and, at the same time, minimize stress on the grid. There are multiple ways for forecasting peak demand (Yunsun et al., 2019; Chirag et al., 2017; Grant et al., 2014), with a large number of these methods making use of only power and weather data. Given the availability of occupancy data, the team tested whether the forecast of peak demand would improve by adding this new variable to existing models. To do this, the team ran

[3] VAV: Variable Air Volume box. A type of HVAC equipment that supply air to each zone.
[4] OA damper: a damper that controls access of outdoor air to the building and is located in the Air-Handler
[5] AHU: Air Handling Unit. A type of HVAC equipment that distributes air through ductwork.

a linear regression model[6] with about five months of data (spanning two semesters and summer break) to forecast peak demand for the next 24 hours on a building in Pomona College. First, the model was run with just one independent variable - weather and other time-dependent variables such as time-of-day and day-of-week. Then, Wi-Fi occupancy was added as another independent variable to see if there were any improvements.Data was sampled at 15 min intervals. Occupancy and weather prediction data for the next 24 hours were assumed to be correct, and these were used to forecast the power through the two models.

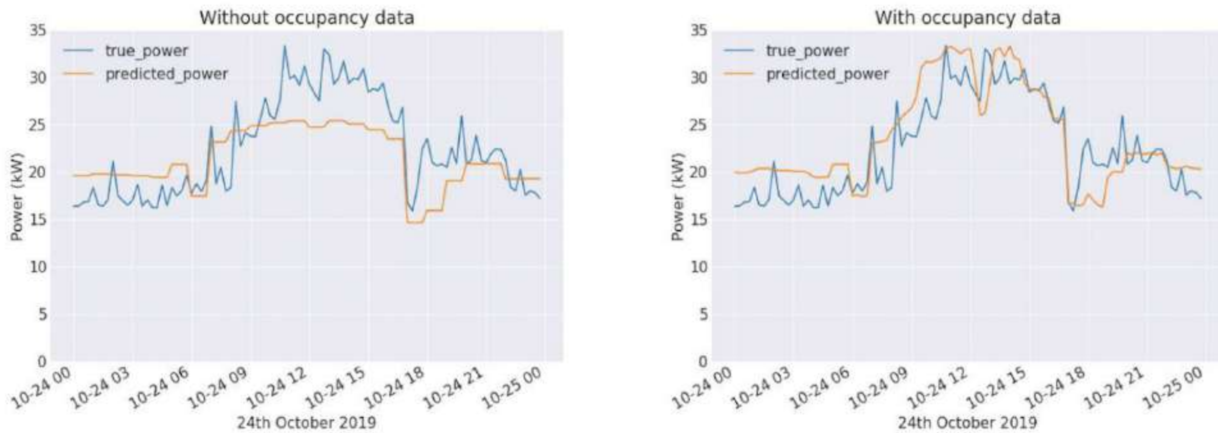

Figure 8: Actual (blue line) power consumption vs forecasted (orange line) power consumption (W) without and with occupancy data in Pomona's student center.

Figure 8 shows the results on a single day (October 24, 2019) of this test. As demonstrated by the figure, occupancy data allows the model to better fit the peaks and troughs. Several other models such as random forest and artificial neural networks were run on this dataset as well. In both cases, occupancy data increased R2 values by around 0.05, thereby showing that occupancy data plays a significant role in improving the load profile forecast and that it can be used for improving current state-of-the-art models.Once peak demand can be reliably forecasted, strategies can be developed and compared to minimize energy cost, as well as relieve stress on the grid.

Discussion

Evaluation of Energy Savings

Isolating the savings attributed to Wi-Fi occupancy detection as part of the ACCO-BEMS platform is outside the scope of this paper, however the research team is in the process of estimating the energy impact of adapting occupancy schedules and airflow rates, for at least one Pomona building. Due to the closure of campus for the COVID-19 pandemic, the above applications have not been implemented in the Spring Semester as planned and an evaluation of the energy savings cannot occur until campus is re-opened.

[6] Linear Regression: A method used for finding relationship between target and one or more predictors.

Lessons Learned

Deployment of this technology in a large campus highlighted important challenges in scaling up this approach to new buildings.

The Pomona IT department voiced legitimate **cybersecurity and privacy concerns** at the outset of the project. The research-grade software developed at LBNL did not initially meet the requirements set by the IT department for deploying software on the IT network. In order to reduce cybersecurity risks, during the initial design and implementation process, the LBNL software was embedded in a virtual machine hosted by Pomona IT, but managed by MelRok. Wi-Fi data were pushed to the MelRok cloud store and accessed via the MelRok platform, which had already been authorized by the IT staff. Due to this multi-stakeholder organizational structure, any changes and updates to the COUNT software required all three institutions to be involved and this led to a slow development cycle. While these initial delays did not impact the overall schedule of the project, new installations should expect similar tight cybersecurity requirements and the project plan should include adequate time for collaborating and coordinating with the local IT group, especially in systems managed by large organizations.

Throughout the project many activities required significant **coordination and cooperation** among people at different institutions. The installation, configuration, and upgrades of the occupancy detection software needed active collaboration among LBNL, MelRok, and Pomona IT. LBNL revised the software package and its configuration due to changes in the Wi-Fi infrastructure (from Aruba and Cisco in the middle of the project); additional functionalities were added to the software (e.g., ability to track single devices anonymously); and there were upgrades to the ACCO-BEMS system and underlying control infrastructure. MelRok developed new features to integrate the new occupancy inputs with the other control and monitoring systems. Pomona IT managed the new software and made sure it did not negatively impact the operation of the network (e.g., cause network traffic congestion). During the integration of the occupancy data into the active management of the buildings, the Facilities and Campus Services Department made sure changes to the HVAC operation would preserve occupant comfort and safety in addition to generating savings. No progress would have been possible without the collaboration of all these partners. Developing a system that streamlines coordination is key to scaling up this solution beyond a research and development project. At the same time, the organization sponsoring the development of such a system should make sure all the parties have the right incentive to cooperate.

Aside from organizational challenges, **integrating a new source of data with existing building automation systems** is a significant enterprise, given the uniqueness of each installation. For instance, to compare actual ventilation with code-compliant rates, as described in our second application, one has to map VAV boxes to the AHU(s) connected to it. The VAV box sensors are used to estimate the amount of air that gets to each zone, while sensors in the AHU are utilized to calculate what fraction of that air is “fresh.” This mapping does not currently exist for all buildings and some of the outdoor air dampers do not report correct data. As a result, to enable this application on all the buildings some physical inspection and software updates may have to be performed on different buildings. A similar issue occurs when mapping APs to VAV boxes; this process requires maps including the location of both types of equipment. Additionally, the range of each AP is not a set zone but depends on the strength of the signal and other factors including how many devices are connected. Therefore, even with an accurate map of the equipment, there is not a definite relation between the AP that a device connects to and the zone that the corresponding person occupies. As a result of this ambiguity, and other factors involving the tradeoff of granularity and complexity, we performed the occupancy analysis at the building level. Future work will include increasing the granularity to the floor or zone level.

Conclusion

This paper describes the development and deployment of an open-source software library to estimate occupancy using Wi-Fi data. The software collects data from tens of buildings in Pomona College campus in Southern California. The library is integrated with the native building automation systems as part of a CEC-funded project in collaboration with MelRok and Pomona College. Over a year of Wi-Fi data was gathered into distinct academic periods, including fall and spring semester, academic breaks, summer sessions. The paper illustrates the data collection system and the automatic data cleaning process, and describes some qualitative tests to understand the potential for over- and under-counting people. Three applications are also highlighted. Two utilize real-time occupancy data to inform optimal HVAC schedules and ventilation rates to identify and reduce energy waste. A third one shows how predicting peak demand can improve by using occupancy data. Finally, the paper discusses lessons learned from the project. While significant progress has been made by the research community in demonstrating the potential of Wi-Fi-based occupancy detection, further work needs to be done to facilitate its deployment in the field. First, guidelines and tools to ensure cybersecurity and data privacy needs to be developed, to reduce risk and facilitate discussion between IT and the facilities departments. Second, additional field work should demonstrate the energy savings potential of this new sensing technology at scale. Third, the integration of Wi-Fi occupancy sensing with Energy Management and Information Systems (EMIS) should be simplified and clearly documented. For instance researchers should streamline the process to integrate Wi-Fi occupancy sensors with a building automation system using BACnet, with new Fault Detection and Diagnostic (FDD) tools, and with emerging advanced supervisory control software.

Acknowledgements

This work was supported by the California Energy Commission under the Electric Program Investment Charge (EPIC) Program, Solicitation Grant Funding Opportunity Number: EPC-16-034, entitled “Automated Cloud-based Continuously Optimizing Building Energy Management System (ACCO-BEMS). LBNL is supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Thanks to Pomona College for hosting this project. We would also like to thank the project lead, Zero Net Energy (ZNE) Alliance ; as well as our partners and collaborators -- MelRok LLC.

References

Ardakanian, O., Bhattacharya, A., & Culler, D. (2018). Non-intrusive occupancy monitoring for energy conservation in commercial buildings. Energy and Buildings, 179, 311-323. https://doi:10.1016/j.enbuild.2018.09.033

Chen Z., Jiang C., Xie L. (2018). "Building occupancy estimation and detection: A review", Energy and Buildings, https://doi.org/10.1016/j.enbuild.2018.03.084.

Balaji, B., Xu, J., Nwokafor, A., Gupta, R., & Agarwal, Y. (2013). Sentinel. Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems - SenSys 13. https://doi:10.1145/2517351.2517370

Chirag D., Fan Z., Junjing Y, Siew E.L. Kwok W.S. (2017). "A review on time series forecasting techniques for building energy consumption". Renewable and Sustainable Energy Reviews, Volume 74. https://www.sciencedirect.com/science/article/pii/S1364032117303155

Erickson V.L., Carreira-Perpiñán M. Á., Cerpa A. E. (2011) "OBSERVE: Occupancy-based system for efficient reduction of HVAC energy," Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks.

Grant J., Eltoukhy M., Asfour S. (2014). "Short-Term Electrical Peak Demand Forecasting in a Large Government Building Using Artificial Neural Networks". Energies, Volume 7. https://www.mdpi.com/1996-1073/7/4/1935

Ouf, M. M., Issa, M. H., Azzouz, A., & Sadick, A. (2017). Effectiveness of using WiFi technologies to detect and predict building occupancy. Sustainable Buildings, 2, 7. https://doi:10.1051/sbuild/2017005

Petersen S., Pedersen T.H, Nielsen K.U, Knudsen M.D. (2016). "Establishing an image-based ground truth for validation of sensor data-based room occupancy detection". Energy and Buildings, https://doi.org/10.1016/j.enbuild.2016.09.009.

Price P.,B. Nordman,M.A. Piette, R. Brown,J. Page, S. Lanzisera, J. Granderson. (2015). "Automated Measurement and Verification and Innovative Occupancy Detection Technologies" September 20, 2015. LBNL-1007182. Available at https://cbs.lbl.gov/publications/automated-measurement-and-1.

Pritoni M., Nordman B., Piette M.A. (2017). "Accessing Wi-Fi Data for Occupancy Sensing". LBNL-2001053. http://escholarship.org/uc/item/7sm90837

Rashid, H., Singh, P., & Singh, A. (2019). I-BLEND, a campus-scale commercial and residential buildings electrical energy dataset. Scientific Data, 6(1). https://doi:10.1038/sdata.2019.15

Shen, W., Newsham, G. and Gunay, B., (2017). "Leveraging existing occupancy-related data for optimal control of commercial office buildings: A review". Advanced Engineering Informatics.Stallings W. (1993). SNMP, SNMPv2, and CMIP: the practical guide to network management. Addison-Wesley Longman Publishing Co., Inc., USA.

Trivedi, A., Gummeson, J., Irwin, D., Ganesan, D., & Shenoy, P. (2017). ISchedule. Proceedings of the Eighth International Conference on Future Energy Systems. https://doi.org/10.1145/3077839.3077846

Wang Z., Hong T., Piette M. A. Pritoni M. (2019). "Inferring occupant counts from Wi-Fi data through machine learning". Building and Environment Building and Environment, Volume 158, 2019, Pages 281-294,ISSN 0360-1323 https://doi.org/10.1016/j.buildenv.2019.05.015

Yunsun K., Heung-gu S., Sahm K. (2019). "Short term electricity load forecasting for institutional buildings". Energy Reports Volume 5. http://www.sciencedirect.com/science/article/pii/S2352484719300435

Zou H, Zhou Y., Jiang H., Chien S., Xie L., Spanos C. (2018). "WinLight: A WiFi-based occupancy-driven lighting control system for smart building." Energy and Buildings. https://doi.org/10.1016/j.enbuild.2017.09.001

Zou J, Zhao Q., Yang W., Wang F. (2017). "Occupancy detection in the office by analyzing surveillance videos and its application to building energy conservation." Energy and Buildings. https://doi.org/10.1016/j.enbuild.2017.07.064