iSchedule: Campus-scale HVAC Scheduling via Mobile WiFi Monitoring

Amee Trivedi, Jeremy Gummeson, David Irwin,
Deepak Ganesan and Prashant Shenoy
University of Massachusetts Amherst
amee,gummeson,dganesan,shenoy@cs.umass.edu,irwin@ecs.umass.edu

ABSTRACT

Heating, ventilation, and air conditioning (HVAC) systems account for over 50% of the energy consumed by commercial buildings. While "smart" HVAC technologies, such as learning thermostats, are widely available for residential use, commercial buildings typically rely on legacy systems that are difficult to upgrade and require facility managers to manually set HVAC schedules. In this paper, we propose a novel Machine Learning-driven technique to automatically learn custom occupancy-based HVAC schedules for buildings across a large campus. While our technique is compatible with any occupancy sensor, we leverage the existing wireless networking infrastructure that is omnipresent across any modern campus. We analyze building WiFi activity, specifically from smartphones, to infer detailed spatial occupancy patterns in each building, and present an algorithm that learns from these patterns to derive a custom HVAC schedule. Our approach is adaptive and dynamically adjusts its schedules as occupancy patterns change, much like a learning thermostat. To evaluate our techniques, we analyze data from several thousand WiFi access points deployed in 112 office buildings on a university campus. Our analysis reveals significant differences in occupancy patterns across and within buildings, motivating the need for our adaptive learning-based approach. Compared to the current static approach, our results demonstrate that learning HVAC schedules from mobile WiFi activity across the campus can yield a 37% reduction in waste time, a measure of energy savings, and a 3% reduction in miss time, a measure of user comfort.

1 INTRODUCTION

Buildings consume nearly 40% of the energy and over 70% of the electricity in advanced economies. The dominant component of a building's energy usage, accounting for more than 50%, is its Heating, Ventilation, and Air Conditioning (HVAC) system. In recent years, there has been significant work in the research community and industry to optimize HVAC energy usage in buildings. For example, smart thermostats [12, 17] use sensors to track occupancy patterns within a home, and then analyze these patterns to automatically learn and program a thermostat schedule. Many smart thermostats are now commercially-available for residential use, including the Honeywell Lyric [18], Nest [21], and Ecobee [8]. These smart thermostats eliminate the need for users to manually operate or program their thermostats to save energy.

Prior work on smart thermostats has primarily focused on residential spaces, i.e., homes, with simple HVAC systems and only a few (often one) thermostats. In contrast, there has been much less attention on optimizing HVAC systems for commercial buildings, which in aggregate consume nearly as much energy as the residential sector. The lack of work is due, in part, to the complexity of commercial HVAC systems in large buildings, which often use custom equipment with limited interfaces that are not easily upgraded. These systems are also inherently unique: they are designed for each building's specific characteristics, and include sensor and control points deployed at custom locations. Thus, upgrading these systems requires more than simply installing a few new thermostats. Given this complexity, the interface to these commercial HVAC systems is typically narrow, enabling facility managers the ability to statically program a building's daily HVAC schedule based on expected work patterns, e.g., 9 AM - 5 PM on weekdays. Unlike residential HVAC systems with smart thermostats, commercial systems often do not expose a rich programmatic interface accessible by third-party software.

While facility managers may account for weekends and holidays in setting schedules, once set, they rarely change them. Further, while some rooms may have their own thermostat that permits local occupant control, users cannot control or change the facility manager's schedule. As a result, these static *ad hoc* HVAC schedules are often not aligned with building occupancy, which changes temporally and spatially. For example, occupancy patterns may differ across floors, reducing the effectiveness of using a single schedule for the entire building. In addition, per-building and per-floor occupancy patterns may change over time and across seasons. While there has been some prior work in occupancy-driven HVAC control [1, 2, 4] it is not applicable to the vast majority of commercial buildings that use *schedule-based* HVAC control that requires setting a repeating schedule. Further occupancy-driven control often requires deploying additional occupancy sensors in each building [5, 9], which is not trivial at campus scale.

To address the problem, we propose scalable, data-driven techniques to infer custom HVAC schedules for office buildings across a large campus. Our techniques automatically analyze occupancy information to learn HVAC schedules that optimize for repeating occupancy patterns. Our system, called iSchedule, is designed to operate with most modern building management systems that employ schedule-based HVAC control. Our system currently leverages the existing wireless networking infrastructure that is omnipresent across any modern campus. Similar to prior work on occupancy detection from soft sensors [3, 4, 11], we analyze building WiFi activity, specifically from smartphones, to infer detailed spatial occupancy patterns across the campus, and present an algorithm that learns from these patterns to derive an "optimal" HVAC schedule for each building. Facility managers can either set the derived HVAC schedule manually, or, for those systems with a programmatic interface, use software to dynamically set the schedule. In designing our system, we make the following contributions.

Campus-scale Analysis: We analyze anonymous WiFi activity logs from 4674 access points deployed in over 112 buildings of various ages and types on a university campus. We show that occupancy patterns vary considerably both across types of buildings (e.g. academic department, dormitory, dining hall, library, gymnasium, etc), within a building (floors and rooms), and seasons. Our study motivates the need for an adaptive learning-based approach due to the complexity of manually setting and tuning HVAC schedules for a large building, or a group of campus buildings.

Machine Learning-based HVAC Scheduling: We present a machine learning-based approach based on an ensemble gradient boosting regressor that predicts occupancy for each floor or zone of a building based on observed WiFi activity. This predicted occupancy is used to derive an optimal HVAC schedule for each floor or zone of a building. Although we use WiFi activity as a proxy for occupancy, our approach is easily adapted to occupancy data from other types of sensors. Since occupancy patterns within a building may change over time, our technique is capable of continuously learning new occupancy patterns as they are observed and then making dynamic scheduling adjustments to compensate.

Implementation and Evaluation: We conduct a detailed experimental evaluation of iSchedule using WiFi data from 112 campus buildings to demonstrate that our system can learn schedules that closely follow the observed occupancy patterns in a diverse set of buildings on our campus. We show that our approach is able to reduce waste times by 37% and miss times by 3% across buildings, which reduce energy waste and increase user comfort. Our experiments also demonstrate the ability of our techniques to dynamically detect changes in occupancy patterns and make appropriate adjustments to the HVAC schedule for each building.

2 BACKGROUND AND PROBLEM

Our work focuses on commercial and office buildings located on an organization's campus. Each building's heating and cooling are controlled by a commercial HVAC system. Unlike a residential HVAC system, which is controlled by a thermostat, a commercial HVAC system is typically controlled through a Building Management System (BMS). The building's facility manager interacts with the BMS to set a heating and cooling schedule and temperature setpoints: this
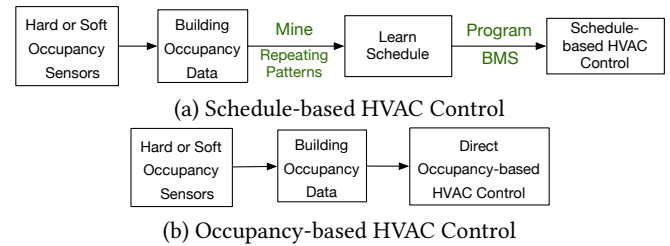

(a) Schedule-based HVAC Control

(b) Occupancy-based HVAC Control

Figure 1: Schedule- versus Occupancy-based HVAC Control

schedule specifies when the HVAC equipment should be turned on over the course of a day and the temperature setpoints for the high and low occupancy periods. The BMS then automatically operates the building's HVAC equipment based on the specified schedule, a process we refer to as *schedule-based HVAC control*.

The schedule is typically determined based on the facility manager's intuitive understanding of the building's occupancy patterns. For instance, in a typical office environment, employees may arrive between 8 am and 9 am and leave for home between 4 pm and 5 pm, and the building may be lightly occupied during non-business hours or on weekends. In this example, the facility manager may program the BMS to heat or cool the building between 8am to 6pm and use a higher cooling or lower heating temperature during the other off-peak hours. Doing so, ensures users are comfortable when the building is highly occupied while saving energy when it is largely unoccupied.

Modern BMS systems enable a different schedule to be set in different parts of a building—e.g., a different schedule on different floors—if the per-floor occupancy patterns differ. However, to fully exploit this functionality, a facility manager needs to determine fine-grain schedules for different parts of a building, and dynamically fine-tune it as occupancy patterns change over time. Such manual operation is cumbersome and error-prone and does not scale across a large campus, where a facility manager may oversee tens-to-hundreds of buildings. As a result, existing schedule-based HVAC control tends to be driven by simple, manually-chosen static schedules, which miss many opportunities for reducing energy use by carefully exploiting temporal and spatial occupancy differences within and across campus buildings.

To address these limitations, we argue that the process of deriving schedules for commercial HVAC systems should be *automated*. To do so, we need a system that monitors occupancy patterns in campus buildings, automatically learns an optimal schedule for each part of a building based on the observed occupancy, and dynamically modifies the schedule as occupancy patterns change. Such a system should be sufficiently robust to tailor its schedules to the different types of *spatial* occupancy patterns seen in different types of campus buildings, e.g. classrooms, academic units, library, dining halls, on a university campus. It should also automatically tailor schedules for *temporal* variations seen across weekdays and weekends and across seasons.

Figure 1(a) depicts the architecture of such a schedule-based HVAC control system—hard or soft sensors are first used to infer occupancy in each building. Occupancy data is then "mined" to extract repeating patterns, from which an "optimal" schedule is learned and fed to the BMS for schedule-based control of the commercial HVAC system.

An essential first step in such a learning-based system is to derive occupancy, which captures *how many* people are present in each part of a building and at *what times*. Hard sensors such as motion or door sensors can be used to track occupancy within each building [1, 2, 6, 22, 25]. However, such instrumentation is not ubiquitous in office buildings and can be expensive and laborious to install in existing buildings. Researchers have shown that occupancy can also be learned through "soft sensors" that are already deployed for other reasons. For example, occupancy can be learned through swipe card door access systems, calendar software, or through wireless network activity [11, 19, 20, 26]. Since WiFi infrastructure is now ubiquitous in offices and campus buildings, our work uses existing wireless networks rather than requiring hard sensors to infer occupancy information. Doing so enables easy deployment of our system in today's campuses without requiring the expensive deployment of hard sensors.

Specifically, our work assumes that most occupants carry mobile smartphones and the presence of a phone in the vicinity of a wireless access point indicates a user (occupant) at that location. We further assume that the exact location of each access point within a building is known a priori. Consequently, simply tracking the number of mobile devices associated with each AP over time is a proxy ("soft sensor") for the number of occupants in that part of the building. We assume the wireless network infrastructure provides a log of when a mobile device connects and disconnects to each access point, which is then used to count the number of active occupants over time. To avoid double counting users, only smartphone log entries are considered and other devices, such as laptops or stationary devices are filtered out[1].

Problem Statement: Given an office campus with multiple buildings, anonymized WiFi association and dissociation logs, and the mappings of the Access Points to specific locations in each building, our goal is to automatically learn an HVAC schedule that optimizes user comfort and energy usage on a fine-grain spatial basis and dynamically adjusts learned schedules when observed occupancy patterns change.

2.1 Relation to Prior Work

Occupancy monitoring and optimizing HVAC efficiency has received significant attention in recent years. This section describes the differences and specific contributions of our work in relation to these past efforts.

Occupancy-driven versus schedule-driven HVAC control: Efforts such as Sentinel and others [1, 2, 4] have shown how occupancy sensors can *directly control* HVAC systems. The basic approach, depicted in Figure1(b), uses observed periods of high and low occupancy to directly control HVAC systems and save energy during off-peak periods. This approach, while novel, is not compatible with most existing BMSs that employ *schedule-driven* control (Figure 1(a)). In the latter approach, occupancy data is first used to learn a repeating schedule, which then is then set in the BMS to control the HVAC system. Thus, occupancy information only indirectly, rather than directly, influences HVAC operation.

While direct occupancy-driven control approaches may be appropriate for buildings with local (e.g., room-specific) HVAC units [13], they are not viable for the majority of centralized commercial HVACs controlled through BMS schedules. In addition, given their experience with schedule-based control, many facility managers may be uncomfortable with ceding direct HVAC control to software. Thus, by deriving repeating occupancy-based schedules, we enable facility managers to retain some control over HVAC usage.

Residential versus Commercial: In residential settings, efforts such as smart thermostat [17], iProgram [12], as well as products such as Nest, Ecobee, and Lyric, have been used to improve HVAC energy-efficiency. Such smart thermostats, as well as all "dumb" programmable thermostats, use schedule-based HVAC control, where occupancy information (from onboard sensors, phone GPS, or even electricity meters [12]) is analyzed to automatically learn a custom schedule. Occupancy sensors may occasionally turn on "away" mode, but they do not exercise direct control. User feedback has also been used to optimize HVAC use [10, 16]. While homes need only binary temporal occupancy larger commercial buildings need spatial occupancy data. Thus, our work can be seen as analogous to these residential efforts but applied to commercial buildings—a more complex problem.

Inferring Occupancy: There has been significant work in deriving occupancy information both for residential and office buildings. Prior work on deriving occupancy information falls into three categories: (i) design of novel occupancy sensors, (ii) use of existing soft sensors [24], and (iii) use of energy analytic methods to learn occupancy [2, 6, 7, 9, 10, 14–16, 20, 23]. However, most approaches only derive occupancy and do not apply it for HVAC control. As shown in Figure 1(a), deriving occupancy data is only a necessary first step for smart HVAC control and is not sufficient for addressing the broader control problem. One closely related technique combines soft sensing with HVAC scheduling [3]; human occupancy is sensed by monitoring human-induced HVAC heat loading and is used as feedback to modify an existing schedule. While our system, iSchedule, also derives HVAC schedules, it instead leverages WiFi-based soft sensors for predicting occupancy. WiFi-based soft sensors can more explicitly derive occupancy counts since there is a direct mapping between numbers of device associations and numbers of occupants. In contrast, our system is easily deployed across an entire campus rather than relying on more specific feedback from advanced HVAC functionality, which could be limited to more recently constructed buildings with newer HVAC units. Furthermore, HVAC-based soft sensors can only operate at the granularity of already defined zones; WiFi access points are typically deployed at a higher spatial density, enabling a building manager with data needed to potentially redefine zones in the future.

3 CAMPUS-SCALE ANALYSIS OF BUILDING OCCUPANCY

A key hypothesis of our work is that occupancy patterns often vary spatially within a building (across floors and zones) and across building types, and temporally across days and seasons.

To validate this hypothesis, we analyzed occupancy patterns across all buildings on our university campus over several months. WiFi connectivity is ubiquitous on our campus—4674 access points

---

[1]Note that the system only needs to count the number of active devices at an AP and does not need to track individual users—identifiable information such as device MAC addresses can thus be anonymized.
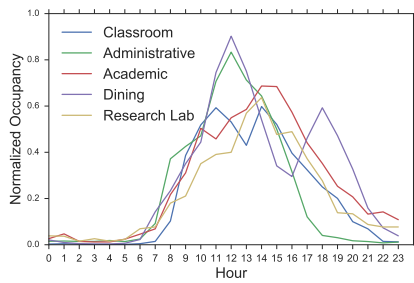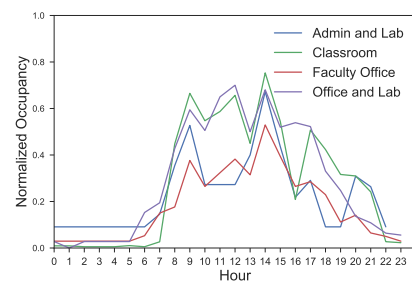
Figure 2: Normalized occupancy across campus buildings.

are deployed in 112 campus buildings. We gathered anonymous WiFi association and disassociation logs for all 4674 APs over a 6 month period ranging from the beginning of Fall to the end of Spring; we included intervals when campus was closed (winter and spring breaks). The mapping of each AP to a building and a specific floor within that building is maintained by our campus IT office, which enabled us to analyze the number of active WiFi users in each building and each floor over the course of the day. We used the number of active smartphone users as a proxy for actual occupancy—a reasonable assumption due to the ubiquity of smartphones today. Due to the scale of our campus level analysis we use occupation computed by counting WiFi devices as our ground truth when evaluating the efficiency of various schedules; this technique has been demonstrated to be sufficiently accurate in prior work [4].
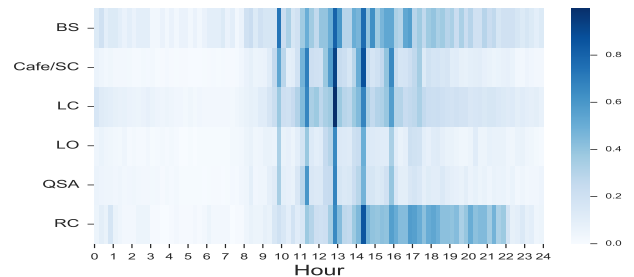
Impact of Building Type: Figure 2 shows the mean weekday occupancy for different hours of the day for several different buildings on our campus. Specifically, the figure shows occupancy for an academic department, a classroom building, an administrative building, a research lab and a dining hall. To ensure comparison across buildings, normalized occupancy is shown and we assume for our discussion here that a building is occupied whenever occupancy levels are more than 20% of the peak, while a building is considered unoccupied if levels are less than 20%.

The figure shows that occupancy patterns vary significantly by building type. The administrative building shows an 8am-5pm occupancy pattern. The academic building, which has student labs/offices has higher evening use as well and shows an 8am-8pm pattern. The occupancy patterns of the classroom building closely follow the lecture schedule. The research labs show a 9am-7pm schedule, while dining halls have peak occupancy during meal hours (e.g., breakfast, lunch and dinner hours). These results show that HVAC schedules need to be aligned with occupancy patterns and building type. The occupancy will vary depending on how the building is used.

Spatial Differences Within Buildings: Next we analyze whether different parts of a building can exhibit different occupancy patterns and by how much. Our analysis of the 112 buildings showed that buildings do exhibit spatial differences in occupancy patterns. This is illustrated in Figure 3, where Figures 3(a) and 3(b) depict occupancy patterns of two buildings on our campus - an academic department and the library. In Figure 3(a), the academic building has one floor comprising of administrative staff, offices, and classrooms, while two other floors comprise faculty offices and research labs. Not surprisingly, occupancy patterns for the floor with staff offices and classrooms is markedly different than those of floors with faculty offices and research labs.
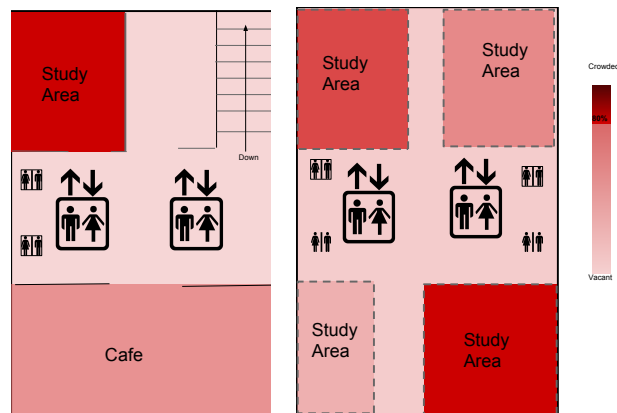

(a) Normalized floor-wise occupancy of an academic building

(b) Floor wise occupancy within the Library

Figure 3: Spatial differences in occupancy inside an academic and library building.

(a) Library Floor 1   (b) Library Floor 5

Figure 4: Library Occupancy depicted as a heatmap on a Tuesday from 10:00 - 11:00 AM shows the differences in occupancy across and within two different floors and motivates the use of unique schedules for different floors and zones.

Similarly, the different floors of our campus library house different functions, as shown in Figure 3(b). One floor consists of the learning commons (LC) that are open 24×7 with student study desks and small breakout rooms for group study. Other floors contain library staff offices (LO), Quiet Study Area (QSA), Resource Center (RC) and checkout desks (Cafe/SC) or aisles of books (BS). As a result, we see very different patterns for these floors.

Finally, we show how occupancy across zones within a given floor of a library differ in Figure 4. In this figure, normalized occupancy is represented as a heatmap and demonstrates how different floors within a building can see different occupancy patterns. In these plots, we show the occupancy of the first and fifth floors

from 10:00 - 11:00 AM on a Tuesday. We note that the top left side of floor 1 (panel a) has a study area that is crowded, while the lower side of floor 1, which has a cafe is less occupied. On Floor 5 (panel b) we see the study areas are occupied but with different occupancy percentage. Zone-based scheduling can further optimize HVAC usage – the top-left portion of Floor 1 has more than 10% occupancy from 9:00 AM - 11:45 PM on the day the heatmap was plotted, while the bottom portion of Floor 1 was more than 10% occupied from 9:30 AM - 5:45 PM. In contrast, the bottom right region of floor 5 sees similar occupancy levels from 9:00 AM - 7:00 PM, while the top most region was occupied from 9:45 AM - 5:15 PM. These observations all motivate the ability to set independent HVAC schedules across both floors and zones.

Overall, our results show that whenever a single building has different types of occupants or houses different types of users in different floors, spatial occupancy patterns will vary.
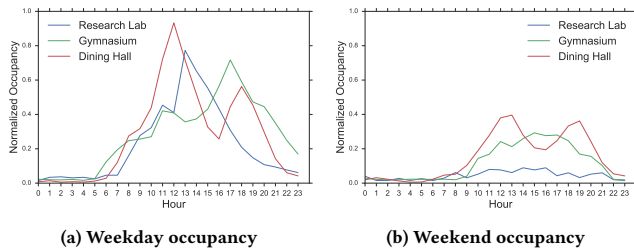

(a) Weekday occupancy   **(b) Weekend occupancy**

Figure 5: Weekday versus weekend occupancy within a research lab building, gymnasium and a dining hall.

(a) Weekday Classroom Occupancy   **(b) Lab Occupancy across semesters**
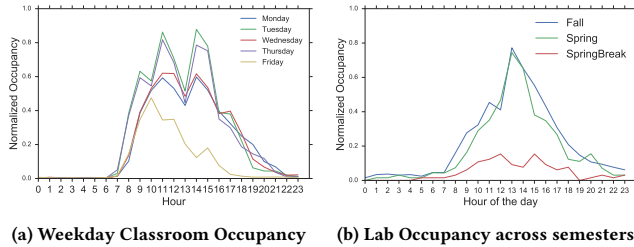
Figure 6: a) Occupancy of the first floor of a classroom building during different days of the week b) Occupancy of a student research lab during Spring, Fall and Spring Break

Temporal Variations: The previous results analyzed spatial occupancy patterns across different types of buildings on a campus and also within a building. Next, we analyze how occupancy patterns vary on a temporal basis over the course of a week and across seasons. Figure 5(a) depicts the normalized occupancy patterns seen in a research lab building, a gymnasium, and a dining hall on a weekday, while Figure 5(b) depicts occupancy patterns for the same buildings on a weekend. Not surprisingly, weekend patterns vary from weekday occupancy; the research lab building sees low weekend occupancy, the gymnasium sees occupancy at different hours on the weekend, while the dining halls see a later "start" to the day on weekends.

Figure 6(a) shows changes in occupancy patterns over the course of a week for the same building on a particular floor. Since this is a classroom building containing a series of lecture halls, the occupancy patterns are highly correlated to lecture schedules, which

vary by the day of the week. Finally, Figure 6(b) shows the occupancy patterns of a research lab in the Fall, Spring, and Spring break. Note that there are seasonal differences in the Fall and Spring semester. Unsurprisingly, we see lower occupancy during breaks when compared to occupancy observed during Fall and Spring.

Due to space constraints, this section depicts only a summary of key results from an extensive analysis of 112 buildings. These results collectively validate our hypothesis that there can be significant differences in occupancy patterns within a building, across buildings, and that the occupancy patterns can change over longer periods (e.g. seasons).

4 LEARNING HVAC SCHEDULES

In this section, we describe iSchedule's data-driven learning algorithm that automatically learns HVAC schedules from occupancy data across campus buildings.

We assume that our system receives a raw log of smartphone association and disassociation information to each access point in the wireless network. Practically every commercial enterprise wireless network product routinely logs such information (e.g. Cisco, HP Aruba). The location of each access point within each building is assumed to be known.[2] Given this data, iSchedule learns schedules as follows:

Step 1: *Compute Temporal Occupancy Per Access Point*
Our system processes the raw WiFi logs to partition the logs on a per-access point basis. It then computes the number of active devices (i.e. users) connected to the AP in each time interval. This is done by incrementing the number of active users upon each new device association and decreasing it for each disassociation event. Doing so yields the number of active users in the vicinity of that AP during each time interval (e.g., every 15 minutes or hourly) over the duration of the log.

Step 2: *Derive Spatial Occupancy within a Building*
Since the location of each AP in a building is known, we can group all AP's spatially to obtain observed occupancy within each part of a building. Any spatial grouping can be chosen (depending on how fine-grain the HVAC control can be). The default grouping is on a per-floor basis—by aggregating the temporal occupancy seen by all APs on each floor, we obtain the number of users that are present on that floor in each time interval over the duration of the WiFi log. This yields a spatial distribution of users across the building and the change in spatial occupancy over time.

Step 3: *Use Predictive Model to Infer Floor/Zone Occupancy*
Next, our system predicts the occupancy of each floor/zone. We use a supervised training technique to predict occupancy. A Gradient Boosting Regressor Ensemble model is trained using the occupancy data computed in the previous step. In the case of university campus buildings, the following features are a strong indicator of occupancy and form our feature set: (i) building name, (ii) building floor or spatial region, (iii) day of the week, (iv) time interval (e.g. hour of day or a 15 minute interval , 9:00 AM to 9:15 AM, etc), (v) semester of the year (vi) month, (vii) holiday and (viii) year. The floor/zone occupancy forms the label set.

---

[2]Since a user may own multiple mobile devices, we avoid double counting by only counting mobile phones connected to an AP and ignoring other device types.

The first two features capture building-specific information, the day of the week captures occupancy variations driven by working versus non-working days, while the time interval captures occupancy at a certain time of that day (e.g. 9:00 AM to 9:15 AM on Mondays). The semester (spring/fall/summer) captures seasonal effects, while holidays capture whether classes are currently in session (e.g. weekends, spring breaks and winter breaks). Note that most of these features are general and can be applied more broadly to any commercial building; specific features, such as semester, can be replaced by a more generic feature, such as the current month.

Our predictive model is based on a regression-based learning approach, which uses a gradient boosting regressor. We use a boosting ensemble method that incrementally builds base estimators so that each sequential estimator is trained to reduce the bias of the earlier estimators. For the model, we used the least squares regression loss function, which is optimized by each estimator. For parameter selection, a 10-fold cross validation technique was used.

Step 4: *Classify Intervals as High/Low Occupancy*
In this step, our system performs a binary classification of each time interval as high or low occupancy based on the occupancy predicted in the previous step. To do so, we first compute a probability distribution of the number of users observed in each part of a building (e.g. probability distribution of users seen on a floor). We define the maximum occupancy of a floor or zone as the high percentile of this distribution. Next, we select a threshold value τ, that represents a fraction of this maximum occupancy – fractional occupancies above or below this threshold are marked as high (H) or low (L) occupancy respectively. This step yields a trace for each portion of a building where each interval is marked H or L over the entire duration of the WiFi trace. As an example, if the max occupancy of a floor is 100 and τ = 10%, then any interval with the floor occupancy exceeding 10 users is marked H and others marked L. Parameter τ can be tuned by the facility manager to chose a suitable tradeoff between building occupant comfort and energy saving. If a high value of τ is chosen then the model becomes aggressive by turning off HVAC equipment more frequently, while a low value of τ causes the model to become less aggressive and leaves HVAC equipment on for longer periods.

Step 5: *Learning a Schedule from the Predictive Model*
In the final step, we derive the actual HVAC schedule. To do so, we consider all seven days of the week and use the model to predict the occupancy for each floor of each building for every time interval of a day. Then, we convert the predicted occupancy in H or L occupancy periods as described in step 4.

We consolidate each sequence of H periods into a single interval where the HVAC must be turned on and consolidate each sequence of L periods into intervals where the HVAC should be turned down. Smoothing can be used to eliminate small periods of H or L periods. This yields a schedule, which gives periods for each day of the week on how the HVAC should be operated on each floor and building (e.g., turn on HVAC from 8:30 AM-5:45 PM on the 3rd floor of the library on Monday and again from 8:00 PM-10:00 PM).

Such a schedule is automatically learned and uses the precise occupancy pattern in each part of a building to compute a custom schedule for different parts of a building.

5 DYNAMIC ADAPTATION OF LEARNED SCHEDULES

The previous section described our learning algorithm, which automatically learns a customized HVAC schedule for each spatial region of a building based on the occupancy patterns observed in that region. However, occupancy patterns are not stationary and will slowly (or abruptly) change over time. These changes in occupancy patterns may occur for a number of reasons: the building or floor may get re-purposed for a different class of users. For example, an academic building may become administrative space with new types of users moving in or there may be subtle changes in occupancy patterns with different types of users over time (e.g., due to changing of class schedules or different user patterns).

Regardless of the cause, the learned schedules cannot remain static—they must adapt and evolve with changing occupancy patterns. In other words, once learned, the HVAC schedule must be dynamically and periodically recomputed and adjusted. The algorithm presented in the previous section can be enhanced in one of two ways to support adaptation.

Continuous Adaptation: In this method, WiFi activity data is ingested every day and spatial occupancy observed within each building during that day is added to the historical trace. The predictive model is re-learned using all data, including the newly ingested information, and the HVAC schedule (step 5) is re-computed. The frequency with which the schedule is recomputed is configurable (e.g., daily, weekly, monthly, etc).

On-demand Adaptation: A limitation of the continuous adaptation approach is that it wastes computational resources when no significant changes to occupancy are observed, as the model is re-trained periodically, regardless of whether it is necessary. On-demand adaptation is an alternate approach that triggers re-training only when the prediction deviates from observed occupancy.

As before, new WiFi activity data arrives continuously and is added to the historical data repository. The system then periodically invokes the previously learned predictive model to predict high and low occupancy labels for a recent time interval. The model predictions are compared to the actual occupancy levels observed in the newly captured WiFi-based occupancy data. If the model predictions match the observed levels, then the occupancy patterns are same as before and neither the model nor the HVAC schedules need to be adjusted. On the other hand, if the recently observed occupancy levels begin deviating from model predictions, then our system triggers a re-training of the predictive model and uses the new model to recompute the HVAC schedules.

Thus, a new model is learned only when needed and only for those buildings (or parts of a building) where significantly different occupancy patterns are observed. The threshold error ϵ between model predicted and actual observations that trigger a re-learning is configurable: a smaller ϵ triggers more frequent re-computations and schedule adjustments and vice versa.

5.1 Discussion

Selecting τ : The value of τ can be selected by the facility manager to choose a suitable tradeoff between building occupant comfort and energy saving. Since our algorithm learns building occupancy independent of a particular schedule, our model supports any value
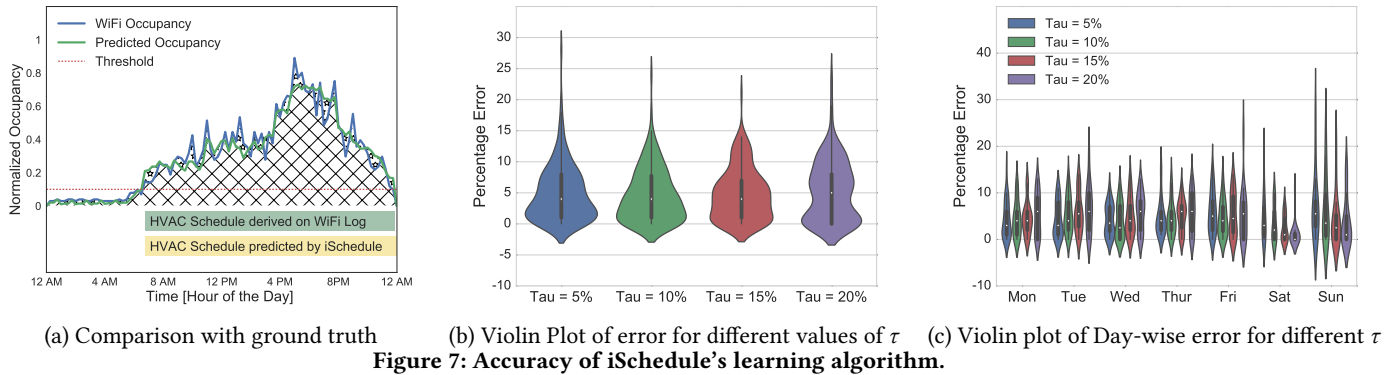
(a) Comparison with ground truth (b) Violin Plot of error for different values of τ (c) Violin plot of Day-wise error for different τ

Figure 7: Accuracy of iSchedule's learning algorithm.

Building Type	Weekday	Weekend
Classroom	8:00 AM to 8:00 PM	Off
Administrative	8:00 AM to 6:00 PM	Off
Academic	8:00 AM to 6:00 PM	10:00 AM to 4:00 PM
Dining	7:00 AM to 10:00 PM	7:00 AM to 10:00 PM
Research Lab	8:00 AM to 7:00 PM	10:00 AM to 4:00 PM
Library	24 hours	8:00 AM to 8:00 PM
Student dorm	24 hours	24 hours
Student Union	8:00 AM to 8:00 PM	10:00 AM to 6:00 PM

Table 1: Statically determined schedules for when the HVAC should be turned on in different types of buildings

of τ. Selecting high values of τ results in more aggressive schedule since HVAC is turned off for periods of occupancy lower than the threshold. Such schedules have a high energy saving but may not necessarily have high comfort. On the other hand, lower values of τ results in a non-aggressive schedule. Therefore, we allow this parameter to be adjusted based on building occupant feedback.

Residential Buildings: While our approach works well for a broad range of office buildings, any campus building that sees "residential" usage requires special handling. Specifically, our technique assumes that *lack of WiFi activity corresponds to a lack of occupants.* In campus buildings with residential environments such as student dorm or campus hotel, students or hotel guests sleep at night and, thus, are present in the environment despite a lack of WiFi activity. Thus, turning down heating and cooling overnight due to lack of observed activity will result in incorrect schedules for these buildings. A simple enhancement can be made to our algorithm by adding a new binary feature called "Sleeping Zones". The computed schedules are then adjusted to keep the HVAC system operational during night hours (e.g. 11 PM - 6 AM) in all areas marked as sleeping zones.

6 EXPERIMENTAL EVALUATION

In this section, we experimentally evaluate the efficacy of our learning-based algorithm and its ability to dynamically adjust schedules based on changing occupancy patterns. We use data from 112 buildings on our university campus for our experimental evaluation. We compare the schedule derived from iSchedule against those derived from WiFi-based occupancy and static pre-set schedules. As a baseline for comparison, we assume the set of static schedules that are shown in Table 1; these static schedules are based on a facility manager's expectation of how different buildings are used.

For the purposes of our evaluation, we use the WiFi occupancy data and our learning-based algorithm to learn a schedule for each building and various floors of each building. We compare the learned schedules, which are based on actual occupancy data, to the statically chosen schedules, and also quantify whether the learned schedule yields increased energy savings and user comfort.

Energy savings are computed as the reduction in *waste time* (WT) when compared to the static schedule, where waste time is defined as the duration for which the HVAC system is turned on even though an area is in a low occupancy period. The increase in user comfort is computed as the reduction in *miss time* (MT) when compared to the static schedule, where miss time is defined as the duration for which the HVAC system is off while an area is in a high occupancy period (causing user discomfort). We formally define waste and miss time in the Appendix.

6.1 Accuracy of our algorithm

In Figure 7(a) we compare the generated HVAC schedule to the actual occupancy. We see that based on the WiFi occupancy detected we find the low and high occupancy periods as marked by H or L Occupancy. This string of H or L Occupancy generated by the system is then smoothed to remove any short intervals of H or L and the smoothened HVAC schedule is obtained. We find that the HVAC schedule generated by iSchedule closely matches the HVAC schedule generated from WiFi data. Further, Figure 7(b) shows the model error computed for a wide range of building types for different values of τ. We trained our model on the historic training dataset and predicted the HVAC schedule for the next 15 days with the adaptation feature disabled. The error was computed against the WiFi building occupancy. We find that our model has a high accuracy of 95.35% with a coefficient of variation of 3.15%. Finally, Figure 7(c) shows the error computed for a wide range of building types for different values of τ for each day of the week. We find that highest variation in error occurs on Sunday for all values of τ. Also, for all weekdays the mean error range was 0 - 7% for different types of buildings and different values of τ.

6.2 Efficacy of Learned Schedules

Table 3 shows weekday and weekend schedules learned for several different buildings on our campus. These schedules correspond to the observed occupancy and we observe several differences between

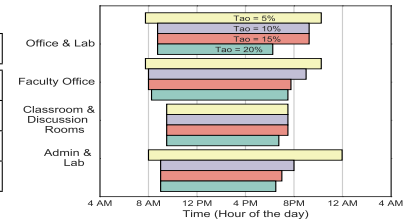| Floor Type	τ = 5%	τ = 10%	τ = 15%	τ = 20%
Office & Lab	8:00 to 23:59	9:00 to 19:59	9:00 to 18:59	9:00 to 18:29
Faculty Office	9:30 to 19:29	9:30 to 19:29	9:30 to 19:29	9:30 to 18:44
Classroom & Discussion Rooms	7:45 to 22:14	8:00 to 20:59	8:00 to 19:44	8:15 to 19:29
Admin & Lab	7:45 to 22:14	8:45 to 21:14	8:45 to 21:14	8:45 to 18:14

Table 2: Comparison of Predicted HVAC schedule for each floor of an academic building, derived for τ = 5%,10%,15%,20%.
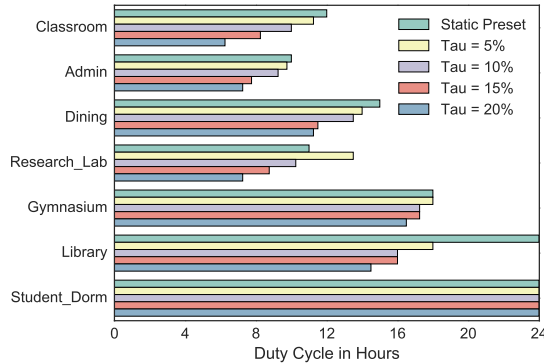

Figure 8: Duty Cycle of learned weekday HVAC schedules for different types of buildings for different thresholds.

Building Type	Weekday	Weekend
Classroom	8:15 AM to 6:14 PM	off
Administrative	8:00 AM to 4:44 PM	off
Dining	8:30 AM to 9:14 PM	11:15 AM to 7:59 PM
Research lab	8:00AM to 6:59PM	off
Library	8:45 AM to 11:59 PM	1:00 PM to 10:59 PM
Student dorm	24 hours	24 hours
Gymnasium	6:45 AM to 11:44 PM	11:15 AM to 7:59 PM

Table 3: Learned weekday and weekend HVAC schedules for different types of buildings computed with τ = 5%.

the learned schedules and the statically set ones, which demonstrates the ability of our approach to capture fine grain occupancy differences. Figure 8 shows the weekday duty cycle for the same set of buildings, as shown in Table 3, for different values of τ. These schedules are derived by iSchedule and we observe several differences between the duty cycles of the learned and static schedules. The duty cycles for derived schedules are highest for τ = 5% and are lowest for τ = 20%; as τ increases, the model is more aggressive in turning off HVAC equipment. We also observe that the duty cycle of static schedules are highest for non-residential buildings, which demonstrates that there is energy wastage by conditioning buildings when occupancy is low.

Building Type	Monday Schedules
Classroom	8:29 AM to 6:14 PM
Administrative	7:30 AM to 5:14 PM
Dining	7:45 AM to 9:14 PM
Research Lab	9:45 AM to 6:59 PM

Table 4: Learned HVAC schedules for a Monday.

Next, Table 4 shows the schedules learned for a specific weekday (Monday) by our algorithm for several types of buildings on our campus – a threshold of τ = 20% was used to compute these schedules. The table also reveals differences from the static schedules, which imply that they incur either more waste or miss time.
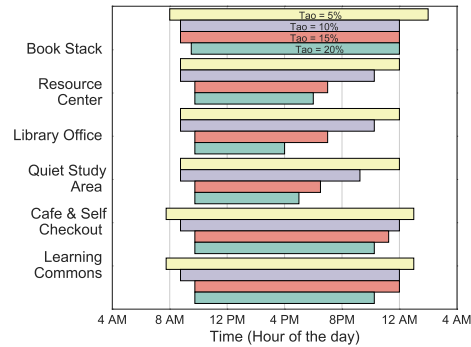

Figure 9: Predicted weekday HVAC schedule derived with different thresholds for a library building with different types of occupancy on each floor.

Table 2 and Figure 9 show schedules learned for different floors of an academic building and library. Since there are spatial differences across floors of each building, a manual schedule that uses a single schedule for the entire building is sub-optimal. Our approach can exploit the observed differences in spatial occupancy and choose different schedules for each floor of a building. We again observe the floor-specific schedules are in line with observed occupancy shown in Figure 3.
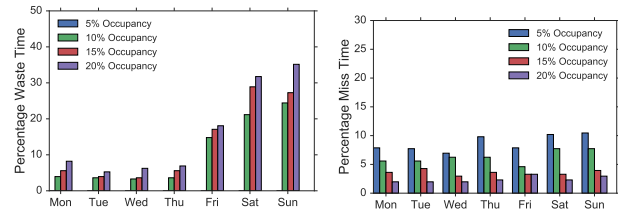
Handling residential areas on campus: For buildings such as student dorms with sleeping residents, our basic algorithm will turn off the HVAC equipment 12 AM to 8 AM on weekdays and 2 AM to 10 AM on weekends in student dorms due to lack of WiFi activity in the night. However, our enhanced algorithm can handle sleeping zones and, as shown in Table 3, leaves HVAC equipment on for 24 hours on weekdays and weekends when the semester is in session, while reverting to a normal schedule during summer breaks when the residence halls are vacant.

6.3 Impact on Energy Use and User Comfort

While the previous results highlight the ability of our approach to automatically derive schedules that closely match observed occupancy, we now quantify the benefits of these derived schedules in terms of energy saving and user comfort. We vary the threshold τ that determines the low occupancy period for each building on our campus and use our algorithm to generate a schedule for that τ. We

compare the derived schedule to the static schedule and compute the increased energy savings and user comfort.

Figure 10 depicts the percentage of waste time for the entire campus (112 buildings) for different days of the week for varying values of τ (τ = 5%, 10%, 15% and 20%).

(a) Reduction in waste time across the campus for varying values of τ

(b) Reduction in miss time across the campus for varying values of τ

Figure 10: Waste Time and Miss Time

Figure 10(a) shows an increasing reduction in waste time with increasing value of τ; the figure depicts the average reduction in waste time across all 112 campus buildings for different days of the week. This occurs because, as the threshold τ is increased, our algorithm is more aggressive in turning off the HVAC equipment via the learnt schedules at higher levels of occupancy. The percentage reduction in waste time is around 3-20% for τ = 10% and increases to 15-37% for τ = 20%.

Figure 10(b) shows the schedules computed by our approach are also able to increase user comfort, which is achieved by reducing miss times. The figure depicts the average reduction in miss time across all 112 campus buildings for different days of the week. Unlike energy savings, user comfort shows a *decreasing* trend with increases in τ. This occurs because, with higher τ, the HVAC equipment is on for fewer hours, which reduces the opportunity to simultaneously increase user comfort. The average reduction in miss times is around 17% for τ = 10% and around 9% for τ = 15%.

Together the results show that across 112 buildings with varied use, automatically learning HVAC schedules using occupancy data yields energy savings while also providing a more comfortable ambient environment to users.
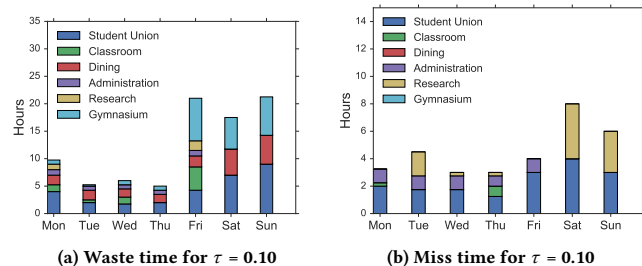

(a) Waste time for τ = 0.10

(b) Miss time for τ = 0.10

Figure 11: Daily reduction of waste and miss time (in number of hours) for a selection of campus buildings

Figure 11(a) and (b) show a breakdown of energy saving and comfort for different types of building for τ = 10%. The greatest gains are observed where learned schedules and actual occupancy varies the most with the static manual schedules. For example, more energy
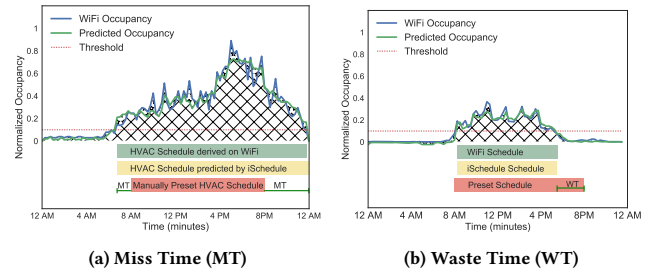

(a) Miss Time (MT)

(b) Waste Time (WT)

Figure 12: Comparison of Schedule derived from WiFi Occupancy, iSchedule and Static schedule, where MT is Miss Time and WT is Waste Time
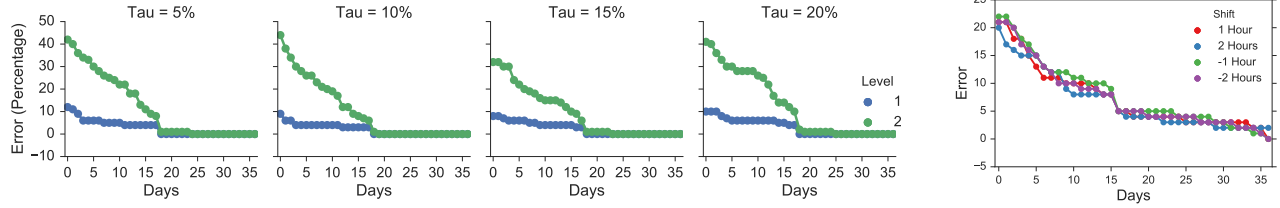
savings are seen on weekends than weekdays. On weekdays, the student union building sees the most savings. Classroom buildings see more savings on Fridays than other days due to a shorter lecture schedule on Fridays. Dining shows high energy savings on Fridays and weekends. Administrative buildings show increased comfort on weekdays, while research labs show an increase in comfort on weekends by following the dynamic schedule.

Finally, Figure 12(a) depicts the normalized occupancy of one illustrative campus building on our campus (the student union) on Friday. As can be seen, the learned schedule is better aligned with observed occupancy on that day and the ground truth occupancy derived schedule. Also, we can see that the derived schedule results in improved user comfort during the evening hours (8 PM-12 AM); the static schedule turns down the HVAC even though the building is at 30% occupancy, while the learned schedule keeps the HVAC system running to maintain comfort for building users. Finally, figure 12(b) depicts the normalized occupancy of one illustrative campus building on our campus (Learning Center, with Classrooms) on Friday – the learned schedule shows substantial savings over a static schedule by reducing waste time. It shows the energy savings during the evening hours (where the static schedule keeps heating or cooling the building later than necessary).

6.4 Efficacy of Dynamic Adjustments

Finally, we evaluate the efficacy of our technique to adjust to dynamic changes in occupancy that may occur in a building. We use WiFi activity data from an academic building and synthetically modify the trace data to emulate two types of changes. First, we shift the observed occupancy to earlier hours, which reflects users arriving to the office earlier than previously observed data. We study the impact of users arriving 1 hour and 2 hours earlier than usual and leaving proportionately sooner, as well as the impact of users arriving 1 - 2 hours later than usual and leaving proportionately later. Second, we swap every Monday and Friday for a set of different building types to simulate a change in working hours, since Monday and Friday occupancy patterns are very different.

Figure 13 and Figure 14 depict our results. In Figure 13(a), we show the change in error for the first 35 days of a building with 2 levels each having a different type of floor occupancy. Level 1 has an almost stable occupancy schedule due to administrative offices while Level 2 has a very dynamic occupancy pattern that differs each day of the week due to the presence of Classrooms and Discussion Rooms. We observe that Level 1 converges quickly as compared to

(a) System Error across different types of floors of a building for 1 Hour Shift   (b) System Error for different shifts

Figure 13: Adaptability of iSchedule's learning algorithm for change in occupancy pattern.
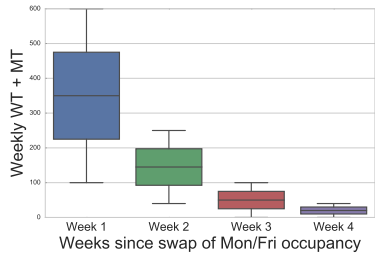

Figure 14: Weekly Waste Time + Miss Time of iSchedule compared to WiFi when Monday and Friday occupancy is swapped for an academic building and a dining hall.
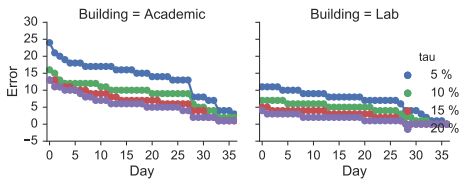

Figure 15: Model Error for learning HVAC schedules of a newly constructed building (a) Academic (b) Research Lab

Level 2. Also, we see high error in the first week after the change in occupancy; this triggers re-training of the model each day which reduces error. The model learns the new occupancy pattern over time and achieves accuracy improvements by the end of the second week – this demonstrates the ability of our approach to adapt to non-transient changes in occupancy patterns for different types of floor dynamics. This experiment also demonstrates that our model adapts to occupancy changes for buildings that have nearly identical occupancy throughout the week or a highly fluctuating occupancy across each day. Figure 13(b) shows that for varying values of shift in schedules the model error converges by the end of week 2 resulting in accuracy of more than 90%. Figure 14 shows that for the first week the error is highest resulting in a high MT + WT value and decreases as the model retrains with new data.

Figure 15 depicts how quickly our algorithm can compute HVAC schedules for two new buildings that lacks historical occupancy patterns. In particular, we consider a newly built academic building and research lab.

In this case, as new WiFi data arrives, the model is retrained and HVAC schedules updated daily. An academic building has multiple floors and each floor has a varying schedule over weekdays and weekends, while the research lab has an almost fixed schedule across weekdays and weekends. As can be seen, our techniques can
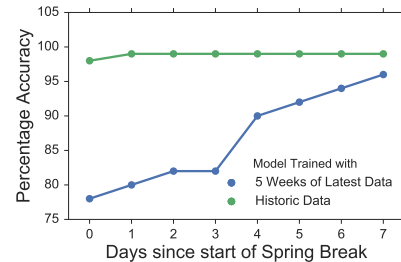

Figure 16: Utility of historic data

still compute schedules with only a few days of occupancy data and its accuracy improves gradually as more occupancy data becomes available – 10 to 28 days of data seems to be sufficient to converge to good schedules, which shows the agility of our technique.

Finally, Figure 16 shows the model accuracy during a change of semester (start of Spring Break) for a dining hall. The model, when trained on only five most recent weeks of data, shows lower accuracy than the model trained on historic data. From the start of Spring Break, the model trained on historic data has a high accuracy of the predicted HVAC schedule, whereas the model trained on the latest 5 weeks of data shows lower accuracy but gradually learns.

7 CONCLUSION

In this paper, we presented a system for campus-scale HVAC scheduling using mobile WiFi data. Our campus-scale analysis showed spatial and temporal variations in occupancy within and across buildings and motivated the need for an automated approach for learning HVAC schedules in campus buildings. We presented iSchedule's supervised learning algorithms and show its efficacy and accuracy across a large university campus. Our future work will focus on shifting HVAC schedules to account for a building's thermal inertia, predicted weather conditions, and thermal communication across zones; these optimizations will further improve user comfort beyond the presented results. We are also implementing fingerprint-based WiFi occupancy detection to further improve the accuracy of zone level scheduling.

APPENDIX

Formal definition of Waste and Miss Time

Formally, we define the miss time and waste time in terms of *conditioning period* (CP) below for N time periods with normalized occupancy N(t) for threshold τ.

$$O(t) = \begin{cases} 0, & N(t) < \tau, \\ 1, & N(t) \geq \tau. \end{cases} \quad (1)$$

$$CP(t) = \begin{cases} 1, & \text{if the zone is conditioned at time t,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Given CP(t), we then define the average daily miss time and waste time over a time period N, as shown below.

$$MT = \frac{\sum_t (O(t) - CP(t))}{N} \quad \forall\ t \text{ where } O(t) = 1 \quad (3)$$

$$WT = \frac{\sum_t (CP(t) - O(t))}{N} \quad \forall\ t \text{ where } CP(t) = 1 \quad (4)$$

REFERENCES

[1] Y. Agarwal, B. Balaji, S. Dutta, R.K. Gupta, and T. Weng. 2011. Duty-cycling Buildings Aggressively: The Next Frontier in HVAC Control. In *IPSN*.

[2] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng. 2010. Occupancy-Driven Energy Management for Smart Building Automation. In *BuildSys*.

[3] Omid Ardakanian, Arka Bhattacharya, and David Culler. 2016. Non-Intrusive Techniques for Establishing Occupancy Related Energy Savings in Commercial Buildings. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 21–30.

[4] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, and Y. Agarwal. 2013. Sentinel: Occupancy Based HVAC Actuation Using Existing WiFi Infrastructure Within Commercial Buildings. In *SenSys*.

[5] A. Beltran and A. Cerpa. 2014. Optimal HVAC Building Control with Occupancy Prediction. In *BuildSys*.

[6] Alex Beltran, Varick L. Erickson, and Alberto E. Cerpa. 2013. ThermoSense: Occupancy Thermal Based Sensing for HVAC Control. In *BuildSys*. ACM, 11:1–11:8.

[7] Adrian Clear, Adrian Friday, Mike Hazas, and Carolynne Lord. 2014. Catch My Drift?: Achieving Comfort More Sustainably in Conventionally Heated Buildings. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 1015–1024. https://doi.org/10.1145/2598510.2598529

[8] ecobee 2015. Ecobee. http://ecobee.com. (January 2015).

[9] V. Erickson, M. Carreira-Perpinan, and A. Cerpa. 2011. OBSERVE: Occupancy-based System for Efficient Reduction of HVAC Energy. In *IPSN*.

[10] Peter Xiang Gao and Srinivasan Keshav. 2013. SPOT: a smart personalized office thermal control system. In *e-Energy*. ACM, 237–246.

[11] Sunil Kumar Ghai, Lakshmi V Thanayankizil, Deva P Seetharam, and Dipanjan Chakraborty. 2012. Occupancy detection in commercial buildings using opportunistic context sources. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*. IEEE, 463–466.

[12] Srinivasan Iyengar, Sandeep Kalra, Anushree Ghosh, David E. Irwin, Prashant J. Shenoy, and Benjamin Marlin. 2015. iProgram: Inferring Smart Schedules for Dumb Thermostats. In *BuildSys*. ACM, 211–220.

[13] G Karmakar, A Kabra, and K Ramamritham. 2015. Maintaining thermal comfort in buildings: feasibility, algorithms, implementation, evaluation. *Real-Time Systems* 51, 5 (2015), 485–525.

[14] W. Kleiminger, C. Beckel, T. Staake, and S. Santini. 2013. Occupancy Detection from Electricity Consumption Data. In *BuildSys*.

[15] Wilhelm Kleiminger, Silvia Santini, and Friedemann Mattern. 2014. Smart Heating Control with Occupancy Prediction: How Much Can One Save?. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 947–954. https://doi.org/10.1145/2638728.2641555

[16] Abraham Hang-Yat Lam, Yi Yuan, and Dan Wang. 2014. An occupant-participatory approach for thermal comfort enhancement and energy conservation in buildings. In *e-Energy*. ACM, 133–143.

[17] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse. 2010. The Smart Thermostat: Using Occupancy Sensors to Save Energy in Homes. In *SenSys*.

[18] lyric 2015. Honeywell Lyric. http://lyric.honeywell.com. (January 2015).

[19] R. Melfi, B. Rosenblum, B. Nordman, and K. Christensen. 2011. Measuring Building Occupancy Using Existing Network Infrastructure. In *Proceedings of the 2011 International Green Computing Conference and Workshops (IGCC '11)*. IEEE Computer Society, Washington, DC, USA, 1–8. https://doi.org/10.1109/IGCC.2011.6008560

[20] Marija Milenkovic and Oliver Amft. 2013. An Opportunistic Activity-sensing Approach to Save Energy in Office Buildings. In *Proceedings of the Fourth International Conference on Future Energy Systems (e-Energy '13)*. ACM, New York, NY, USA, 247–258. https://doi.org/10.1145/2487166.2487194

[21] nest 2015. Nest. http://nest.com. (January 2015).

[22] G.R. Newsham and B.J. Birt. 2010. Building-level Occupancy Data to Improve ARIMA-based Electricity Use Forecasts. In *BuildSys*.

[23] Devika Pisharoty, Rayoung Yang, Mark W. Newman, and Kamin Whitehouse. 2015. ThermoCoach: Reducing Home Energy Consumption with Personalized Thermostat Recommendations. In *BuildSys*. ACM, 201–210.

[24] A. Prakash, V. Prakash, B. Doshi, U. Arote, P. Sahu, and K. Ramamritham. 2015. Fusing Sensors for Occupancy Sensing in Smart Buildings. In *ICDCIT*.

[25] E. Tapia, S. Intille, and K. Larson. 2004. Activity Recognition in the Home Setting using Simple and Ubiquitous Sensors. In *Pervasive*.

[26] K. Ting, R. Yu, and M. Srivastava. 2013. Occupancy Inferencing from Non-Intrusive Data Sources. In *BuildSys*.