

Softgreen : Towards Energy Management of Green Office Buildings with Soft Sensors

Lakshmi V Thanayankizil[†]
Georgia Institute of Technology, USA
lakshmi@gatech.edu

Sunil Kumar Ghai, Dipanjan Chakraborty and Deva P. Seetharam
IBM Research, India
{sunilkrghai, cdipanjan, dseetharam}@in.ibm.com

Abstract—This paper describes an approach for saving energy in commercial buildings, based on the information gathered from pre-existing opportunistic context sources. Most energy management systems rely on a *heavy instrumentation strategy* to infer occupancies, and unfortunately ignore already available opportunistic context sources, that can provide significant information about occupancy. We present models to conduct a *Context Profiling* with available context sources, to infer spatial occupancy measures. Further, we model electrical loads of several types to infer potential energy savings. Through a pilot study of a building with 5 users for 30 days, we identify intra-building areas where additional instrumentation of occupancy sensors is not necessary and demonstrate potential for significant reduction in energy consumption. We believe such Context Profiling can provide insights to significantly reduce deployment and management costs for future occupancy detection and energy management systems.

I. INTRODUCTION

Office energy consumption is a significant contributor to greenhouse emissions. In particular, space conditioning and lighting loads (L-HVAC - Light, Heat, Ventilation and Air Conditioning) together account for 70% of all energy consumed in a typical office building [15]. Several Building Energy Management Schemes (BEMS) are being proposed to reduce office energy consumption. Many of these schemes adopt a *heavy instrumentation* strategy, installing many kinds of sensors (e.g. camera, augmented PIRs, CO_2 etc) to monitor environment occupancy [8], [5], [13], and control electrical loads.

Heavy instrumentation strategies essentially install sensors with a strategy of optimally covering the monitored area to achieve desired accuracy in detecting occupancy. Although inexpensive yet accurate deeply-coupled wireless sensors have been developed [1], [3], [10], costs for deployment and management of additional sensing infrastructure remain significant. Such costs can be avoided if the pre-existing IT and Security infrastructures of a building (ubiquitously present in most office buildings today) can be used to determine occupancy. These infrastructures include several opportunistic context sources (we call them *soft sensors*). For example, many offices have an ID badge scanning system, Wi-Fi access points/Ethernet ports, and several additional context sources on devices carried by users – instant messaging systems, online

calendar, device activity status, etc, which can provide valuable cues to an employee's location within a building.

A central research question that drives our work is: *To what extent can opportunistic context sources available in commercial buildings help us in achieving energy savings?* We believe this is an important question because such a hardware sensor-less approach can be economical and is easier to adopt as there are no additional costs for deployment, operation and maintenance. It is important to note that this approach does not preclude additional occupancy detection sensors. We advocate a hybrid strategy, where a smart deployment should only install additional sensors at spaces where current opportunistic context sources are insufficient. Such a strategy can accelerate commercial scale adoption of occupancy detection systems.

Our contributions are two-fold: (1) Development of a Context Profiling and Energy Management System (dubbed *SoftGreen*) that conducts a profiling of office spaces, and detects areas where occupancy can be detected with existing Soft Sensors. (2) Implementation of a generic model of electrical loads based on detected occupancy levels that enables us to compute energy savings. We evaluated SoftGreen by conducting a pilot in an office building for 30 days. Our results indicate that we do not need deep coupling and heavy instrumentation of office buildings, at least in many parts of it. Opportunistic Context sources seem to be sufficient for detecting occupancy with reasonable accuracy.

II. BACKGROUND

There are high precision (accuracies of around 1-2 m) and costly localization sensors [11] available which can be used for indoor localization and subsequent occupancy detection. However when used for Building Energy Management, errors as low as 1-2 m could place a user under the coverage area of the wrong electrical load. The tradeoff between cost vs. accuracy forms the central reason for investigating alternative approaches and many recent works focus on these. For example, some modern buildings have explored installing cheap Passive Infrared (PIR) based motion sensors to detect user presence and control L-HVAC [4]. However, we agree with [10] who points out - “Motion sensors are notoriously poor occupancy sensors and have long been a source of frustration for users of occupancy-based lighting systems, which often turn the lights off when a room is still occupied”.

[†]The work was done when the author was an intern at IBM Research.

Research works have considered alternatives like CO_2 -based occupancy detection [12], sonar-based methods [13], camera networks [14], [8], [5]. These are better detectors of long-term occupancy than PIR sensors. CO_2 sensors however are slow to detect change of events and hence perhaps suitable for less dynamic environments (e.g. residences) than large offices. Large-scale deployment of Camera or sonar-based surveillance networks (costlier than PIR sensors) would incur substantial deployment costs and maintenance overhead, also bring up privacy issues as they detect more than what is required. Our work is complimentary to these approaches. We argue for selective deployment of such occupancy sensors only in areas where opportunistic context sources are inadequate.

To reduce deployment overhead, more recent works consider using wireless networks [1], [3], [6], to aggregate data from multiple occupancy sensors. Cheap sensors can be installed in large numbers, leading to better occupancy predictions. For example, Delaney et al.[3] use PIR based wireless occupancy sensors to measure wasted energy in lighting when there are no occupants. Erickson et al.[6] propose a more sophisticated wireless network of cameras (which have the aforementioned privacy and cost issues) to determine coarse-grained floor-level occupancy detection. A different work by the same authors consider occupancy prediction [5] in addition to real time occupancy monitoring using cameras to save energy while maintaining building comfort standards. A more recent work [1], uses low cost and incrementally deployable wireless PIR sensors with reed switches and evaluates the accuracy of their occupancy sensor. They show that their sensors make accurate predictions and save over 18% of the energy in comparison with a motion sensor-based scheme.

Note that, these works do not consider any pre-existing context sources, the devices being carried by users, and to what extent can occupancy be detected by these. The focus is primarily on reducing deployment and management cost of the parallel infrastructure. Moreover, many of the sensors work only for closed spaces (e.g. fitting sensors on doors), whereas, many large commercial offices have open spaces (e.g. low-wall cubicles). Our research augments this body of work by advocating a prior context profiling, and smart identification of areas where occupancies can be detected using existing *soft sensor data*.

III. PROBLEM DETAILS AND CHALLENGES

Offices have many possible layouts, can be mixed use/multi-function buildings as well. We focus on two key abstractions: (1) *Closed spaces*: Spaces such as meeting rooms, conference rooms, that are used on-demand; has a door for entry/exit and dedicated L-HVAC loads. (2) *Open Spaces*: Spaces with low (such as cubicles with less than 5-feet high separators) or no enclosing walls where L-HVAC equipment are shared between areas. Usually, a significant portion of many large-scale commercial sites are often open spaces [7]. In fact, about 85% of our candidate office is comprised of open spaces while the remaining area belongs to closed spaces. Determining occupancy in open spaces (e.g. a layout of 30

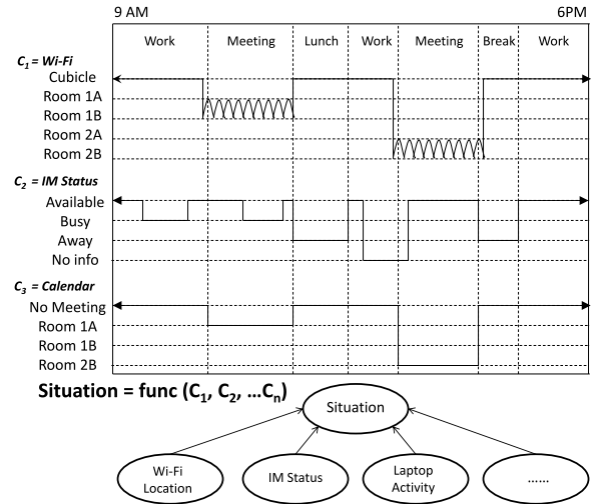


Fig. 1. Situation Inference in Softgreen. Situation is a function of inputs from multiple context sources.

cubicles in a large hall) is challenging as there are multiple entries/exits and occupancy densities vary across the area.

It is non-trivial to determine occupancies of open and closed spaces with imprecise context sources. Some of the issues we need to handle are:

- 1) *Unreliability of Context Sources*: Many of the infrastructures (e.g. Instant Messaging, Calendar entries) are meant for other purposes (e.g. chatting or maintaining appointments) and are at best fuzzy indicators of an employee's state (e.g. user is away from his laptop, user is in a meeting). Wi-Fi based fingerprints of user's laptops are often coarse-grained and dependent on the base station density, which is usually enough to ensure coverage. Installations are not streamlined to localize, in most cases. Moreover, it is not a certain indicator of the user's occupancy in an area. This is because employees often leave their laptops in their cubicles, while going for other activities (meeting, break etc). An accurate *location state* of the user needs to be derived considering these multiple context cues.
- 2) *Conflicts*: We need to address conflicts amongst sources providing two different location estimates for a user. For example, imagine an employee attending a meeting in Room X (obtained from his calendar entry), leaving his laptop in his cubicle (open area). The Wi-Fi context source would indicate occupancy in the open area, whereas the online calendar would indicate an occupancy in Room X.

Hence, intelligent learning models need to be designed to fuse different context source outputs, to infer an employee's location state and derive occupancy of areas. In order to detect occupancy from soft sensor data, we model the context of an user as a *situation*, which in turn is a function of the context cues coming from these sources (detailed later). Figure 1 illustrates a figurative example of how data from

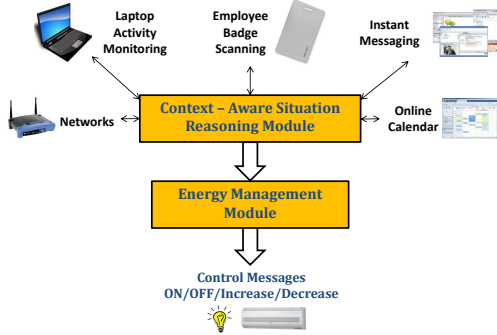


Fig. 2. SoftGreen: Context Profiling and Energy Management System.

multiple context sources (including the Wi-Fi fingerprint from the user’s laptop) vary during an employee’s daily routine (Y axis) and the corresponding situations s/he is in (Time axis). The two curves in Wi-Fi fingerprint show that Wi-Fi is unable to differentiate between two close meeting rooms (e.g. Room 1A and Room 1B), if the user is attending a meeting in one of the two rooms.

In the next section, we describe different models that we employ in SoftGreen to conduct the context (and energy) profiling.

IV. SOFTGREEN SYSTEM AND MODELS

As shown in Figure 2, the SoftGreen system consists of two tightly coupled modules. The Situation Inference Module accepts context cues and determines *relevant* user situations. The Energy Management Module accepts inferred user situations as a time series, determines occupancy levels in areas belonging to electrical loads. There after, it determines ideal operating conditions of loads (e.g. switch ON/OFF, dim/brighten) based on load type and constructs an energy savings profile for the area. Next we describe the two modules in detail.

A. Situation Inference Module

Discrete Values of Context Cues: Every soft-sensor reports a set of discrete context cues. For example, IM status could take one of the values : {Available, Busy, Away, No Info}. Similarly, possible sets for other context sources are shown in Table I.

Situation Definition: Situation is the fundamental unit in our model and is a higher level abstraction of the user location and activity. We define situation $S = S_1, \dots, S_M$ as a function of data observed from N context sources $C = C_1, \dots, C_N$. i.e. $S = f(C)$.

For each user, we consider three situations that help to infer area occupancy: (1) User is in her cubicle (*available*($user_i, cubicle_j$)); (2) She is in a meeting room(*meeting*($user_i, room_j$)); (3) She is outside (*outside*($user_i$)). Situation (3) essentially indicates when User is not in any location of interest. For example, she could have gone to lunch.

As shown in Figure 1, context source values change corresponding to user situations (shown along x-axis), leading to

Soft-Sensor	Discrete values set
WiFi	{Cubicle, Meeting Room 1A, Meeting Room 1B, ..., Outside}
Calendar	{No Meeting, Meeting in room 1A, ..., Call in number}
System Monitoring	{Active, Inactive}

TABLE I
DIFFERENT DISCRETE VALUES OF SOFT-SENSORS

a multi-dimensional (one dimension per source) synchronized time series of readings coming from N context sources.

Observation vector: For N context sources, given a set of context observations $C = [C_{1t}, C_{2t} \dots C_{Nt}]$ at time t , where C_{it} represents the value observed from the context source C_i , we want to infer the user situation S_t at that instant. An example of a set of observations from the context sources could be {WiFi => Cubicle, Calendar => No Meeting, IM Status => Available, and System Activity => Active}.

Now we explain two supervised learning approaches to infer user situation from these observations. Both these approaches utilize the training data that correlates situations (‘Cubicle’, ‘Meeting’, ‘Outside’) with context cues from different sources.

1) *Maximum Likelihood Model:* This model is expressed as:

$$S_t = \underset{j}{\operatorname{argmax}} \mathcal{L}(S_j | C_{1t}, C_{2t} \dots C_{Nt}) \quad (1)$$

where $\mathcal{L}(S|C)$ is the likelihood of the situation S given observed values in C (from context sources). Since the values from context sources are correlated (for example, a user might set her IM status to ‘Busy’ while in a meeting or when she is in a meeting, the WiFi location could point to one of the meeting rooms), equation 1 can be expanded as

$$\begin{aligned} P(S|C_{1t}, C_{2t} \dots C_{Nt}) &= P(S|C_{1t}) \cdot P(C_{1t}|C_{2t} \dots C_{Nt}) \\ &= P(C_{1t}|C_{2t} \dots C_{Nt}) = \\ &= P(C_{1t}|C_{2t}) \cdot P(C_{2t}|C_{3t}) \dots P(C_{(N-1)t}|C_{Nt}) \end{aligned}$$

The conditional probabilities on the right side of the equations are computed using the training data. This approach worked reasonably well for many office spaces. However, we observed that it suffers from over estimation if a dominant location fingerprint is present in the training data¹. For instance, during the training phase, if a user mostly left her laptop in her cubicle while attending meetings or during breaks, such a training data would bias the situation estimation to ‘cubicle’ during the operational phase. Considering the limitation, we propose a regression-based model.

2) *Regression-based Model:* In this model, we consider relative contributions of each context cues to each of the situations and combine them as follows:

$$w_1 \cdot C_{1t} + w_2 \cdot C_{2t} + \dots + w_N \cdot C_{Nt} = S_j \quad (2)$$

where $[w_1, w_2 \dots w_N]$ are the unknown weights assigned to observations $[C_{1t}, C_{2t} \dots C_{Nt}]$ for situation S_j . Before we solve $W \cdot C = S$ using standard techniques to get W , we need

¹We validated this through experiments, though we do not present the results in this paper due to shortage of space.

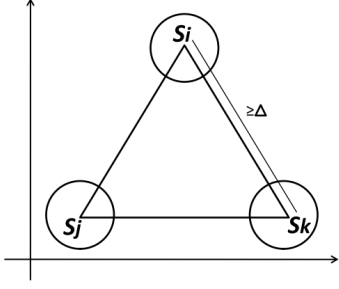


Fig. 3. Conversion of situation vector to numeric values

to assign numeric values to the observation vector C and the situation vector S .

Assigning numeric values to observation vector C : We assign $P(C_{it}|S_j) * idf(C_{it})$ as the numeric value to corresponding elements in the C vector ($i = 1 \dots N$), where $P(C_{it}|S_j)$ is computed from the training data as the Probability Mass Function (PMF) of the observations from context source C_i , for the situation S_j ; $idf(C_{it})$ is Inverse Document Frequency (IDF) measure of C_{it} . The IDF measure enables us to de-emphasize same context cue that occur in multiple situations.

Assigning numeric values to situation vector S : For S , we transform each S_i to a Cartesian point in (x, y) (say S'_i) plane, as shown in Figure 3, such that for any other situation S_j , $|S_i - S_j| \geq \Delta$. S'_i sets target point for situation S_i during test time and the value of Δ defines the distance between two target points. We associate a threshold value τ around each S'_i (radius of each circle) to classify between an ‘inferred’ and a ‘not inferred’ state. Essentially, as τ increases more states could be inferred, but, at the cost of precision.

Finally, at test time, given an input observation $C(t')$ (context values at time t'), and the W vector (precomputed using training data), we assign a situation S'_i , if $Euclidean\ distance(S'_i, C) \leq \tau$ (i.e. a point lies within one of the three circles). For the rest, we mark them *not inferred*.

Learning approaches : The ground truth is essential for any learning and validation. However, it is arduous for employees and often the data collected is not sufficient for learning in supervised approaches. We are currently investigating unsupervised techniques such as Latent Dirichlet Allocation [2], to reduce the burden of human tagging to identify situations.

B. Energy Management Module

For the design of the Energy Management Module, we consider 3 layers : *electrical load, area and user*, where we use area as a key midlevel abstraction for combining multiple context sources.

For an electrical load, L , the state of operation (ON/OFF) is modeled using a binary function $f(L) \in \{0, 1\}$, such that the load L at time t

$$L = \begin{cases} ON & \text{if } f(L) = 1 \text{ at time } t \\ OFF & \text{if } f(L) = 0 \text{ for time period } \Theta_t = t' \rightarrow t \end{cases} \quad (3)$$

This model currently follows a lazy strategy to switch off loads if no occupancy has been detected in the associated area for a certain time Θ_t , and a fast trigger (switch ON) if presence of any user is detected.

An area associated with a load L is defined as a collection of n objects covered (spatially) by that load (where the object could be a meeting room or a cubicle). For loads like lights inside a cubicle $n = 1$, while $n > 1$, for loads like AC, lights that are common to many cubicles. The output of the binary function $f(L)$ is determined by the object/objects of the area associated with that load and is modeled as follows:

$$f(L) = \begin{cases} 1 & \text{if } \sum_{i=1}^n A_{i,L} \geq 1 \\ 0 & \text{if } A_{i,L} = 0, i = 1 \dots n \end{cases} \quad (4)$$

where $A_{i,L} \in \{0, 1\}$. $A_{i,L}$ is the i^{th} object of the associated electrical load L and the value depends on the ‘inferred situation’ of the user(s) associated with $A_{i,L}$. We define $Z_{i,A_{i,L}}$ to be the set of all employees associated with the i^{th} associated object of load L . $Z_{i,A_{i,L}}$ has a cardinality one if $Z_{i,A_{i,L}}$ has only one user associated with it (for example, cubicle), whereas for objects like ‘meeting rooms’ there are multiple users associated with each object.

For the object $A_{i,L}$, let $S_{k,A_{i,L}}$ be the inferred situation of the k^{th} user, where $k \in Z_{i,A_{i,L}}$. Then, $A_{i,L} = 1$, if $S_{k,A_{i,L}} = 1$ for any user $k \in Z_{i,A_{i,L}}$. Further, we have the flexibility to model open spaces as well, apart from closed spaces, and provide a formalism to connect the inferred user situations with the appropriate loads. Note that extending L to be n-ary to represent tunable loads can be done by considering the relation of the load levels with the occupancy numbers.

V. IMPLEMENTATION AND EXPERIMENTS

We implemented a monitoring agent in SoftGreen using Visual Basic, that continuously recorded context sources observations from a user’s laptop. We considered the following available context sources in the candidate office: Online Calendar, Wi-Fi Access Points seen by the user’s laptop, Chat Client, activity status of the user’s laptop. This multi-dimensional time series data was transferred to a back-end server periodically for running SoftGreen analytics. The analytics platform was built using Python and Matlab. HORUS [16] – a Wi-Fi based location estimation mechanism was used to compute significant locations traversed by users, based on an offline collection of radio maps of the candidate areas. In our office set-up, we obtained a fuzzy location granularity approximately of 6-10 meters.

Data Collection: Data was collected from 5 office users, in a corporate office building for 30 working days². The users were chosen from multiple different teams having different mobility behaviors. Users sat in open spaces (in their cubicles) and performed several activities (attending meetings, phone calls, lunch, breaks etc). Each floor typically consisted of a large open space and a few closed spaces (meeting rooms).

²Though we ran the campaign with ≈ 15 users, only 5 users had reasonable ground truth logged for any fair learning and validation.

Figure 4 shows the floor map of the building; the red circles indicate the meeting rooms and the blue squares indicate the cubicles for some monitored users. Finally data of electrical load positions were collected to run SoftGreen profiling. Further, we collected ground truth of activity logs from

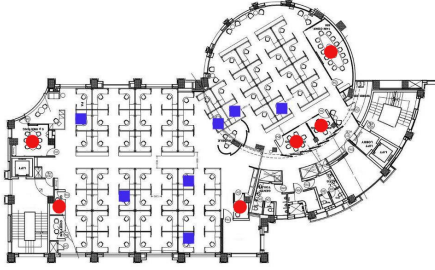


Fig. 4. Floor map of one of the floors that were monitored showing monitored open areas and closed areas.

monitored users. Each user logged entries for the following situations, at the start of the situation (1) User in his cubicle ($available(user_i, cubicle_j)$); (2) User is in a meeting room ($meeting(user_i, room_j)$); (3) User is outside ($outside(user_i)$).

At the preprocessing stage, we discard all data chunks for which complete set of observations from all the sources were not recorded (could be due to several reasons, including failure of the agent before it recovered again). For learning of situation models, we considered a document size of 30 minutes on the processed data, starting from each user tag (i.e. data from tag time: $t \rightarrow t+30$). This is because due to the onerous nature of tagging, users often forget to tag exactly at the start time of each situation. Our model learns the W_i vector for each user i , and uses a user-specific threshold τ_i to infer situations. Not all situations are present with all users. For example, one user never attended meetings (network team). Next, we present results on (1) SoftGreen’s Context Profile and Occupancy prediction; (2) Insights on Energy Savings.

A. Spatial Context Profile and Occupancy Measures

For testing, we compute SoftGreen Situation Inference on each unique observation from the context sources (e.g. C_{1t}, C_{2t}, C_{3t}) for the complete time series of each user for 30 days, using a 10-fold cross-validation approach. The inferred situations give us occupancy statistics for users in cubicles and meeting rooms in our candidate office. Occupancy measures for loads spanning across multiple users (e.g. an overhead light or an AC vent) is a trivial addition of the measures obtained for each user in that area at time t .

Let $correct(S_i, U_j) = \text{Num. of correctly predicted entries of } S_i \text{ for user } U_j \text{ across the observation period.}$ Let $ground_truth(S_i, U_j) = \text{Total Num. of log entries of } S_i \text{ from } U_j.$ We measure *accuracy* for situation S_i for user U_j as $accuracy(S_i, U_j) = \frac{correct(S_i, U_j)}{ground_truth(S_i, U_j)}$.

Figure 5 demonstrates the spatial distribution of accuracy measures obtained for open areas (spanning cubicles of these users) and closed areas (meeting rooms).

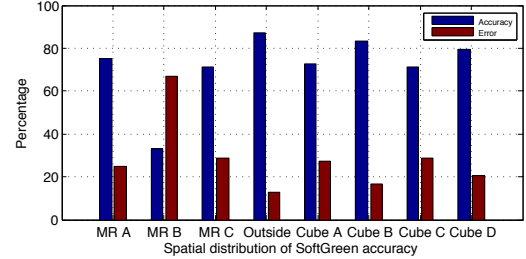


Fig. 5. Spatial Distribution of the average Softgreen accuracy for some spaces in our candidate office. X-axis indicates different areas.

From these results, we observe that SoftGreen predictions have a strong spatial correlation. For example, average accuracies in many open spaces are quite high, while one closed space is also well detected. This implies that these areas can elegantly use the existing Soft Sensor infrastructure for occupancy detection. Further, we can also infer which spaces need additional instrumentation (e.g. MR B, Cube A where accuracies are below 80%). Considering that it is challenging to instrument open spaces, we believe our results show strong motivation for BEMS to adopt such profiling strategies.

Figure 6 shows a sample experiment comparing SoftGreen’s occupancy prediction, with that of a commercial motion-sensor. In this experiment, we placed the sensor in one of the user’s cubicles for a day. It was surrounded by other cubicles and meeting rooms. The typical sensitivity was 5 meters. We observe user-centric learning (by SoftGreen) performs well to infer occupancy and avoids many false positives (and negatives) possible in such environments.

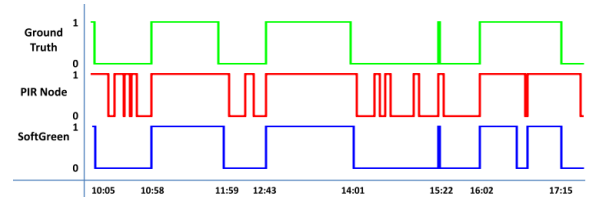


Fig. 6. A sample experiment comparing SoftGreen’s occupancy prediction w.r.t. a motion sensor placed in an open cubicle. ‘1’ indicates ‘occupied’ and ‘0’ indicates ‘unoccupied’ in the Y-axis.

The richness of profiling would depend in turn on the number of context sources available, dynamicity of the user, which vary for offices. As our models and computing strategies are independent of the number and nature of sources, we believe it is applicable for heterogeneous building scenarios.

B. Insights on Energy Savings

Next, we zoom into the energy savings insights from areas having high occupancy detection accuracies (open spaces). In this paper, we evaluate loads that have 1-1 relation with area and users (e.g. cubicle lights having $n = 1$). Every user cubicle considered in our experiment has one 98-Watts light. We follow a dynamic policy switching off a load L if the cubicle remains unoccupied for a threshold $\Theta_t = 5minutes$

(Ref. Equation 3). Light is turned on the moment occupancy is detected. This policy is followed by most PIR-based motion sensors in industrial settings too. We compared the usage statistics with the static schedule followed in the office (switch ON lights during working hours and switch OFF at night). We use Average Percentage Energy Savings as a metric for comparison with the static schedule.

For a policy P , if $E(P, T)$ = Energy that is saved by Policy P between $T = t_2 - t_1$ (due to loads being OFF); $E(Static, T)$ = Energy Consumed by Static schedule between $t_1 \rightarrow t_2$, then $APES(P, t_1, t_2|Static) = \frac{E(P, T)}{E(Static, T)}$. Figure 7 plots observed APES against the time of the day, and presents deviations observed across the observation period of 30 days. The energy consumed in running the Softgreen code is accounted for using Joulemeter [9] in our calculations.

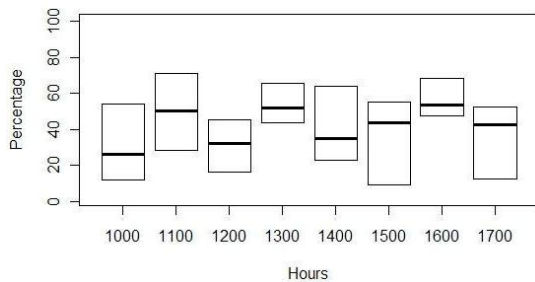


Fig. 7. Temporal distribution of APES between hours 1000 and 1700 for our candidate office, showing the 25th and 75th percentile box plots with the highlighted band indicating median.

We observe that daily variations in savings during the office hours are captured. For example, savings are lower during the early office hours (around 9AM), as the number of meetings are lower at these times, and is the highest at around 13:00 HRS, which is when majority of the employees leave for lunch, leaving their cubicles empty. Deviations are contributed by expected variations in schedules amongst monitored users across 30 days. Most importantly, we observe significant potential for savings in these areas (20% to 50% mean savings), without the use of additional instrumented sensor infrastructure. As a policy, we do not profile energy savings of areas (e.g. some meeting rooms) where SoftGreen Context profiling reports poor accuracies. As a work in progress, we are currently modeling the energy savings potential for other electrical loads with 1 : n relations with users (overhead lights, ACs), which also brings in detection of stray occupants (someone temporarily occupying an user's cubicle).

VI. CONCLUSIONS AND FUTURE WORK

This paper presented an approach of exploiting existing opportunistic context sources to profile occupancy measures of office buildings. Our results on a candidate office over a survey of 1 month indicates that significant areas (specially open spaces like cubicles) can be monitored for occupancy using existing, ubiquitous, commercially installed Soft Sensors. SoftGreen Context and Energy profiling can provide valuable insights on which parts of buildings can be avoided

while instrumenting additional occupancy detection sensors. Hence, it identifies opportunities for reducing deployment and management cost of the new infrastructure. Our models for computation are carefully designed to be generic and applicable across multitude of office environments and loads.

Apart from modeling multi-use electrical loads, we are currently investigating unsupervised techniques to reduce the burden of human tagging to identify situations. We believe our work compliments the literature around energy management of green buildings and recommends novel ways to optimize occupancy sensor deployment and management.

REFERENCES

- [1] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng. Occupancy-driven energy management for smart building automation. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, NY, USA, 2010. ACM.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Mar. 2003.
- [3] D. T. Delaney, G. M. P. O'Hare, and A. G. Ruzzelli. Evaluation of energy-efficiency in lighting systems using sensor networks. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, BuildSys '09, NY, USA, 2009. ACM.
- [4] Dwyer Instruments. <http://www.dwyer-inst.com.au/>.
- [5] V. L. Erickson and A. E. Cerpa. Occupancy based demand response hvac control strategy. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 7–12, New York, NY, USA, 2010. ACM.
- [6] V. L. Erickson, Y. Lin, A. Kamthe, B. Rohini, A. Surana, A. E. Cerpa, M. D. Sohn, and S. Narayanan. Energy efficient building environment control strategies using real-time occupancy measurements. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, BuildSys '09, pages 19–24, New York, NY, USA, 2009. ACM.
- [7] Y. Hua, V. Iofinnes, R. Kraut, and K. M. Powell. Workplace collaborative space layout typology and occupant perception of collaboration environment. In *Environment and Planning B: Planning and Design*, volume 37, pages 429–448, 2010.
- [8] A. Kamthe, L. Jiang, M. Dudys, and A. Cerpa. Scopes: Smart cameras object position estimation system. In *Proceedings of the 6th European Conference on Wireless Sensor Networks*, EWSN '09, pages 279–295, Berlin, Heidelberg, 2009. Springer-Verlag.
- [9] A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya. Virtual machine power metering and provisioning. In *Proceedings of the 1st ACM symposium on Cloud computing*, SoCC '10, pages 39–50, New York, NY, USA, 2010. ACM.
- [10] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse. The smart thermostat: using occupancy sensors to save energy in homes. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, SenSys '10, pages 211–224, New York, NY, USA, 2010. ACM.
- [11] Real time location system, Ubisense. <http://ubisense.net>.
- [12] State-of-the-art review of CO2 demand controlled ventilation technology and application. U.S. Dept. of Commerce, National Institute of Standards and Technology. Gaithersburg, Md. : 2001.
- [13] S. P. Tarzia, R. P. Dick, P. A. Dinda, and G. Memik. Sonar-based measurement of user presence and attention. In *Proceedings of the 11th international conference on Ubiquitous computing*, Ubicomp '09, pages 89–92, NY, USA, 2009. ACM.
- [14] T. Teixeira and A. Savvides. Lightweight people counting and localizing for easily deployable indoors wsns. *Selected Topics in Signal Processing, IEEE Journal of*, 2(4):493–502, aug. 2008.
- [15] US Whole Building Design Guide Sustainable Committee. <http://www.wbdg.org>.
- [16] M. Youssef and A. Agrawala. The horus wlan location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, MobiSys '05, pages 205–218, New York, NY, USA, 2005. ACM.