Towards an Understanding of Campus-Scale Power Consumption

Gowtham Bellala
University of Michigan, Ann Arbor
gowtham@umich.edu

Manish Marwah, Martin Arlitt, Geoff Lyon, Cullen E. Bash
Hewlett Packard Laboratories
firstname.lastname@hp.com

Abstract

Commercial buildings are significant consumers of electricity. In this paper, we collect and analyze six weeks of data from 39 power meters in three buildings of a campus of a large company. We use an unsupervised anomaly detection technique based on a low-dimensional embedding to identify power saving opportunities. Further, to better manage resources such as lighting and HVAC, we develop occupancy models based on readily available port-level network logs. We propose a semi-supervised approach that combines hidden Markov models (HMM) with standard classifiers such as naive Bayes and support vector machines (SVM). This two step approach simplifies the occupancy model while achieving good accuracy. The experimental results over ten office cubicles show that the maximum error is less than 15% with an average error of 9.3%. We demonstrate that using our occupancy models, we can potentially reduce the lighting load on one floor (about 45 kW) by about 9.5%.

Categories and Subject Descriptors

H.4.m [**Information Systems**]: INFORMATION SYSTEMS APPLICATIONS—*Miscellaneous*

Keywords

Commercial buildings, power consumption, occupancy modeling, anomaly detection

1 Introduction

Commercial buildings use a significant amount of energy as part of their day-to-day operations. In 2009, commercial buildings in the United States alone consumed an estimated 1.3 trillion kWh or about 37% of the total electricity generated [13]. Owing to concerns such as increasing energy costs, shrinking operational budgets, and global climate change, there is growing interest in understanding how resources such as electricity are used in commercial buildings, so that steps can be taken to reduce consumption.

In this paper, we collect and analyze real power data from three commercial buildings, totaling 300,000 sq. ft. We have installed 39 meters on the campus power delivery infrastructure, to get a more complete understanding of power use.

The main goal of this study is to provide an understanding of how, when and where power is consumed in a commercial campus. With respect to this goal, we propose an unsupervised technique to identify anomalous usage periods in power consumption time series data. We examine six weeks of data, and show a number of power saving opportunities. We also develop occupancy models based on computer network port-level logs, to help determine more efficient management policies for lighting and HVAC. We propose a novel semi-supervised approach combining a hidden Markov model (HMM) with a classifier. Further, based on our occupancy models, we estimate that modifying the lighting schedule can save about 9.5% of lighting power.

The main contributions of this paper include a (brief) characterization of power use in a commercial campus, the design and use of an unsupervised anomaly detection method, design of semi-supervised occupancy models and their use to help reduce power consumption. These studies also revealed a number of challenges that motivate future work in this area.

The remainder of the paper is organized as follows. Section 2 introduces related work. Section 3 describes the campus and the power delivery and measurement infrastructure. Section 4 provides a brief characterization of power use on the campus. Section 5 discusses our anomaly detection investigation. Section 6 explains how we model the campus occupancy. Section 7 summarizes our work.

2 Related Work

Commercial buildings consume a lot of energy [13]. This motivates research to improve building energy efficiency. Few campus-scale studies of energy use exist. Agarwal et al. [1] examined 6 months of data from the UCSD campus, including aggregate power consumption of four buildings. Our study is complementary, as we examine energy use of a commercial (rather than educational) campus.

Examining data for anomalies is a known approach for identifying abnormal system behavior. Catterson et al. use this approach to monitor old power transformers [2]. Their goal is to proactively search for abnormal behavior that may indicate the transformer is about to fail. Similarly,

BuildSys'11, November 1, 2011, Seattle, WA, USA.

Figure 1. **Aerial view of the campus, showing the six main buildings, three of which are currently instrumented for power monitoring.**

McArthur et al. search for anomalies to detect problems with power generation equipment [9]. Jakkula and Cook compare several outlier detection methods to find which is better at identifying abnormal power consumption [6]. Seem [12] and Li et al. [8] both search for anomalies in building power consumption, to detect abnormal power use. Our work on anomaly detection extends the results of these studies.

Occupants of a building contribute to its energy footprint. Unfortunately, directly tracking the number of people in a building is challenging. To estimate the occupancy, we create models based on data retrieved from periodic scans of the computer network in the campus. This is a similar approach to that of Newsham and Birt [11]. Erickson et al. [4], and Kim et al. [7] model occupancy through a variety of means. Again, our work is complementary.

3 Campus Overview

We use our campus as a testbed to investigate the monitoring and management of resources such as power, gas, water and waste, with initial focus on power. The campus contains six main buildings with a total footprint of 700,000 sq. ft. At present, our efforts are focused on three two-storey buildings (1, 2, 3), as highlighted in Figure 1. These three buildings have a 300,000 sq. ft. footprint, representing 43% of the total campus floor space, hosting about 500 occupants.

3.1 Power Distribution Topology

Figure 2 shows the power distribution topology for buildings 1, 2 and 3. They are all fed by a single utility feed (3-phase 12.5kV). There is an emergency back-up generator (3-phase 480V) to maintain a subset of critical loads in the event of a utility failure. These switch over to generator power via automatic transfer switches (ATS). The main distribution panels in each building (3-phase 480V) branch to about 10 major sub-loads or sub-panels within each building. Building 3 has a 135kW photovoltaic array atop of it, to offset power demand during daylight hours.

3.2 Power Data Collection

Our present monitoring installation has 39 power meters, deployed at the site, building and top-level load distribution panels, as shown in Figure 2. We plan to instrument the second-tier distribution panels within each building, which will triple the number of power meters. This will provide finer grained electrical data monitoring for our future work.

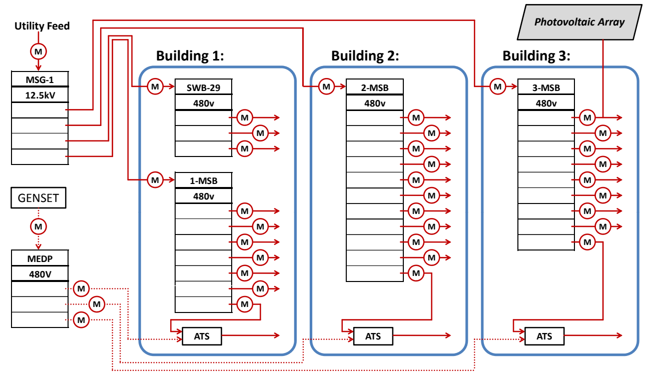


Figure 2. **Power distribution and metering topology.**

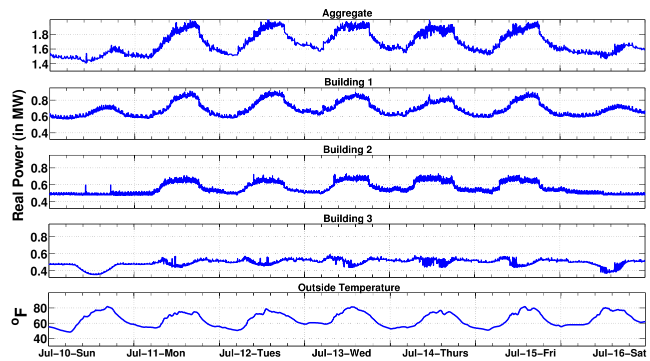


Figure 3. **Campus power use and outside temperature.**

The main challenge in deploying the monitoring infrastructure is accurately deciphering the power distribution topology. The buildings are about 60 years old, and have experienced infrastructural and functional changes over their lifespans, which are not always well documented. While this challenge may not exist in new buildings, we suspect that it will for the many existing (legacy) campuses.

All of our electrical meters are commercial (3-phase) devices from Schneider Electric (`www.schneider-electric.com`). We poll a set of parameters from each meter every 10 seconds using the MODBUS over Ethernet protocol. The monitored parameters include line voltage, real and apparent power, power factor, current and frequency. We maintain a historical log of power related data using PI-Server from OSIsoft (`www.osisoft.com`), which we also use to obtain data from the photovoltaic array installation atop building 3.

4 Characterization of Campus Power Use

This section provides a brief summary of power use on the campus. The top graph in Figure 3 shows the aggregate power demand for buildings 1, 2 and 3 for the period of Sunday, July 10, 2011 through Saturday, July 16, 2011. This graph reveals several key characteristics. First, the demand has both a constant (base) and variable load components. The base load is quite significant, at 1.5 MW. A large IT infrastructure is partially responsible for this. The variable component adds up to 0.5 MW of demand. Second, there is a distinct time of day pattern. Power demand is lowest during the night and early morning, and highest during the late morning and afternoon. Third, there is a pronounced day of week behavior, with weekends (and non-work days in general) consisting primarily of the base load, and work days having the noticeable variable load.

One implication of the significant base load seen in the aggregate power demand is that known disaggregation techniques (e.g., [5]) are not likely to work well. The use of more pervasive power monitoring will help us to determine what contributes to the base load.

The middle three graphs in Figure 3 show the total power demand for each of Buildings 1, 2 and 3, respectively. Building 1 has a base load of about 0.6 MW, and a peak load of nearly 0.9 MW, while Building 2 has a base load of approximately 0.5 MW, and a peak load up to 0.7 MW. It has essentially no variable load on non-work days. However, there were two large spikes (almost 100 KW) on Sunday. This sort of behavior seems anomalous, and helps motivate our work on anomaly detection in Section 5.

Building 3 has a base load of almost 0.5 MW. Unlike Buildings 1 and 2, Building 3's variable load is negative. At first glance, this may seem anomalous; however, this is due to the presence of a 135 KW photo-voltaic array installed on that building. The last graph shows the outside temperature.

5 Anomaly Detection

In this section we use anomaly detection to better understand the campus power consumption. Our primary goal is to detect any abnormal behavior in the power usage time series. Note that an anomaly indicates an irregular usage pattern and may not always correspond to a component failure or faulty operation.

There are two main challenges for performing anomaly detection. The first challenge is the lack of labeled data to train an algorithm for detecting anomalous behavior. Obtaining labeled data is an expensive procedure as it requires a human (usually a building administrator) to meticulously go through the vast amount of power data. In addition, it might also necessitate injecting faults to obtain a good representation of anomalies in the training data. The second challenge is the high dimensionality of the power data.

To deal with these two problems, we propose a novel unsupervised cluster-based algorithm that detects anomalous points via a low-dimensional embedding of the power data. This algorithm takes as input the power time series of a meter over several days, and outputs the probability of a particular day being anomalous. The probability scores can be used to rank the days in terms of anomalousness, providing a building administrator with a prioritized list of data points that require further inspection. The algorithm is described in detail next.

5.1 Methods

First, we need to introduce some convention. We refer to power data measured by a single meter over a 24 hour period (i.e., one day) as one observation or as a single power-time curve. Due to the lack of labeled data, we use an unsupervised approach where we cluster the power-time curves of each meter. The intuition behind this approach is that the data points that exhibit normal behavior will form tight clusters while anomalous points will lie outside these clusters.

To compare two power-time curves, we propose the use of the standard Euclidean distance measure, or the l_2 norm, between the frequency spectrum of the two power-time curves. Note that the frequency spectrum consists of two components - magnitude and phase. However, we restrict our attention to the magnitude of the frequency spectrum as it contains all the necessary information regarding the power consumption behavior.

Our proposed algorithm consists of five steps. The first step is to impute missing values in a power-time curve. These missing values could have been caused either due to a hardware or a software failure. Treating these missing values as zeros will lead to spurious high frequencies in the frequency spectrum. We adopt a weighted global average strategy as this method can be used to impute blocks of missing values, while preserving the local structure. Specifically, let $x[n]$, $n = 1, \cdots, N$ denote a power-time curve where N denotes the number of samples. For any time index $1 \leq m \leq N$ with $x[m]$ missing, we impute its value by $x[m] = \frac{\sum_{k=1}^{N} w[k]x[k]}{\sum_{k=1}^{N} w[k]}$, where the weights $w[k]$ are chosen such that they decrease as a function of their distance from the missing value. For example, the weight function can be chosen to be $w[k] = 1/|m-k|^2$. This imputation strategy can be considered as a temporal smoothing technique. In our data, on average less than 3% values were missing.

Step 2 computes the frequency spectrum of the imputed power-time curve. Given an imputed sequence $x[n]$, for $n = 1, \cdots, N$, its frequency spectrum is computed as

$$X[k] = \sum_{n=1}^{N} x[n] * exp\left(-j2\pi(k-1)\frac{n-1}{N}\right),\ 1 \leq k \leq N.$$

Let $Y[k] = |X[k]|$, $k = 1, \cdots, N$ denote its magnitude.

Given M different observations corresponding to power-time curves on M different days, in step 3 we compute the dissimilarity between the power consumption profiles on any two days using the standard Euclidean distance measure between their frequency spectrums as $\delta_{ij} = \left[\sum_{k=1}^{N}(Y_i[k] - Y_j[k])^2\right]^{\frac{1}{2}}$. The $M \times M$ dissimilarity matrix Δ is obtained by computing the above distance measure for all pairs of observations.

With dissimilarity matrix Δ, step 4 uses a dimensionality reduction algorithm such as MDS (Multi-dimensional scaling) [3] to obtain a low-dimensional Euclidean embedding of the M observations in a $d \ll N$ dimensional Euclidean space (i.e., $\mathbb{R}^d$). Figure 5 demonstrates a low-dimensional embedding of 33 power-time curves where $d = 2$.

Given this low dimensional embedding, the last step is to compute the probability score of each observation being anomalous. We compute these values through a k-NN (nearest neighbor) density estimation algorithm. Note that a low-dimensional embedding of the power data is crucial for this step, as density estimation is known to perform poorly in a high dimensional space due to the curse of dimensionality.

For every point $\mathbf{y} \in \mathbb{R}^d$ in the low dimensional space, the local density at that point can be estimated as

$$\widehat{f}(\mathbf{y}) = \frac{k}{\text{Vol. of smallest hyper-sphere containing } k \text{ NNs of } \mathbf{y}}$$

	AUC
Meter 1	0.87
Meter 2	0.96
Meter 3	0.99

(a)

	Category	# of Anomalies
1	High power usage	17
2	Low power usage	8
3	Irregular Shutdown	4
4	Irregular (time) usage	6
5	Oscillatory behavior	8
6	Abnormal drop/rise	13

(b)

Figure 4. **(a) Accuracy of the proposed algorithm in identifying anomalies as measured by AUC (b) Anomaly types and number.**

where k is chosen to be $O(M^{\frac{1}{d}})$. Given the local densities at each of the M observations, the probability of an observation being an anomaly is computed as

$$\Pr(\mathbf{y}_i \text{ is anomalous}) = 1 - \frac{\widehat{f}(\mathbf{y}_i)}{\max_{j=1,\cdots,M} \widehat{f}(\mathbf{y}_j)}.$$

Intuitively, observations in dense regions are less likely to be anomalous and those in sparse regions are more likely to be anomalous. Finally, this algorithm can be implemented in an on-line fashion, enabling real-time anomaly detection.

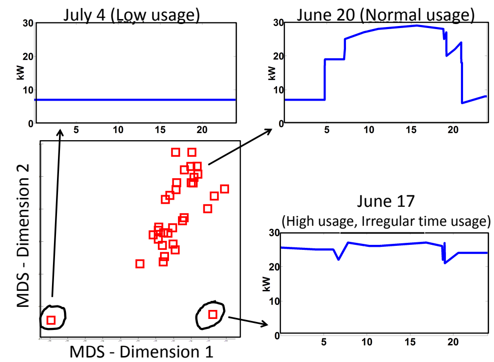
5.2 Experimental Results

We performed anomaly detection on six weeks of data from the 39 power meters. To validate our results, for three meters (the main meters for Buildings 1, 2 and 3), we obtain the ground truth by consulting with the building administrator, who looked at the entire six weeks time series data and marked potentially anomalous regions. As described above, our algorithm assigns a probability score to each day, which can be used to obtain a ranked list of days in decreasing order of them being anomalous.

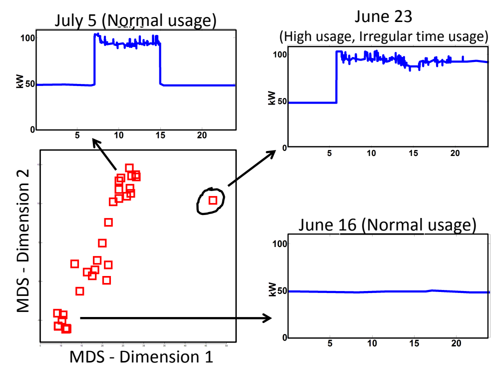
Given this ranked list, a building administrator could choose a threshold k and declare the top k points as anomalies for further inspection, and the remaining as normal, where k could vary from 0 to the maximum number of points in the input data (M). Each choice of k results in a certain number of false positives and false negatives. For example, when $k = 0$, i.e., when all the points are declared as normal, the false positive rate (FPR) is 0 while the false negative rate (FNR) will be 1. On the other hand, when $k = M$, the associated FPR is 1 and FNR is 0. Varying this threshold k results in different values of FPR and FNR, leading to a receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) defines the quality of the obtained ranking. In the ideal case, where all the anomalous points are ranked at the top followed by normal points, the AUC takes the maximum value of 1. On the other hand, a random ranking achieves an AUC value of 0.5. We use AUC as a performance metric for our algorithm. Figure 4(a) demonstrates the performance of our algorithm on 3 meters.

Further, we characterize the anomalies detected in the 39 meters by assigning them categories, as shown in Figure 4(b). Note that a particular anomaly could belong to multiple categories. Some of these categories provide an opportunity for potential energy savings, while others may indicate device malfunction or failures. In Figure 5, we demonstrate only 3 of the 6 categories for reasons of brevity.

The examples shown in Figure 5 offer the potential to save energy. Figure 5(a) corresponds to the overhead lighting load of a floor, while Figure 5(b) corresponds to several



(a) Load corresponding to lighting.



(b) Load corresponding to air handling units.

Figure 5. **Examples of low dimensional embedding of power data.**

air handling units (AHU). The low-dimensional embeddings obtained using MDS in both cases show clusters of normal behavior (June 20th, July 5th, and June 16th), and points that were detected as anomalous (circled).

For the lighting load, two anomalous points are seen: one corresponding to low usage (category 2), which turned out to be July 4th, a holiday; the other (June 17th) corresponding to high and irregular usage (categories 1 and 4), where the lights remained on all night. Detecting and correcting this anomaly could have saved about 180 kWh over this day. The AHU load also has a similar anomaly (June 23rd), where the air handlers were operating until late in the night. Correcting this anomaly could save about 450 kWh. Note that the AHU load has multiple normal modes of operation depending on the utilization level of the air handling capacity.

6 Occupancy Modeling

Occupancy modeling forms another important component for efficient power management in buildings. Many commercial buildings employ either a fixed time L-HVAC (lighting, heating ventilation and cooling) schedule or a fixed temperature set point schedule. This often leads to unnecessary conditioning of the building, especially when the actual occupancy is low. Hence, some recent work has suggested occupancy-based L-HVAC scheduling for efficient power management. However, most of this work assumes the availability of occupancy sensors, whose installation and maintenance may be prohibitive on a large campus. Note that occupancy modeling, irrespective of the methods used, has privacy implications which are outside the scope of this

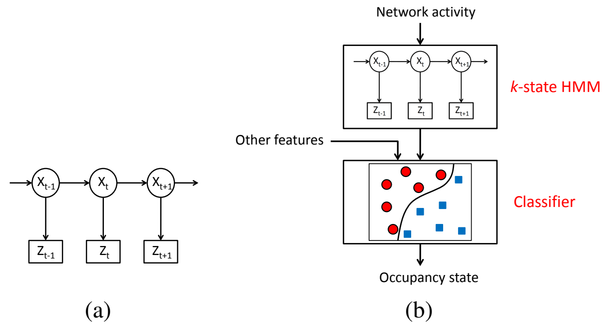


Figure 6. **(a) A hidden Markov model with hidden variable X_t and observed variables Z_t (b) Proposed 2-stage approach.**

paper.

On the other hand, Melfi et al. [10] propose the use of existing network infrastructure to estimate occupancy. In this paper, we propose an implicit occupancy sensing procedure, where we use traffic data associated with network ports in each cube to build occupancy models. Network switches typically maintain per port counters for the amount of in and out flowing traffic. We retrieve this information from the switches in the building every 30 minutes. The estimated occupancies at cube level are then used to estimate occupancy of a zone (e.g., multiple cubes), which can further be used for occupancy-based L-HVAC scheduling of that zone.

However, there are various challenges in modeling occupancy from network activity. One of the main challenges is the lack of labeled data. Obtaining labeled data from each occupant is not only expensive but also raises privacy issues. In addition, note that there are certain intrinsic limitations with the use of network data to estimate occupancy. For example, a cube might be occupied but the occupant may not be using his/her computer. However, our preliminary analysis shows that network activity is still a good feature to estimate occupancy.

We propose two approaches to estimate occupancy from network data. The first approach is an unsupervised approach where we use hidden Markov model (HMM) [3] to estimate occupancy from network data. We then propose a two-stage semi-supervised approach where the first stage involves unsupervised learning using HMM followed by training a classifier using minimal labeled data. The two approaches are described in more detail next.

6.1 Methods

We first propose an unsupervised approach where we model the problem of occupancy estimation from network data as a hidden Markov model with binary occupancy states as the hidden variable X_t and network data as the observed variable Z_t, as shown in Figure 6(a). We consider a simple model where we assume that the transition probabilities and the emission probabilities do not vary with time. As we show in Section 6.2, this model performs fairly well.

However, the above assumptions may not hold. In fact, it seems more appropriate to model these probabilities as a function of other features such as time of the day, day of the week, etc. The dependence of these parameters on such external features is known to significantly increase the complexity of HMMs, to the extent of making them intractable on large datasets. Alternatively, we propose a novel two stage semi-supervised approach that can efficiently incorporate the effect of external features, as shown in Figure 6(b).

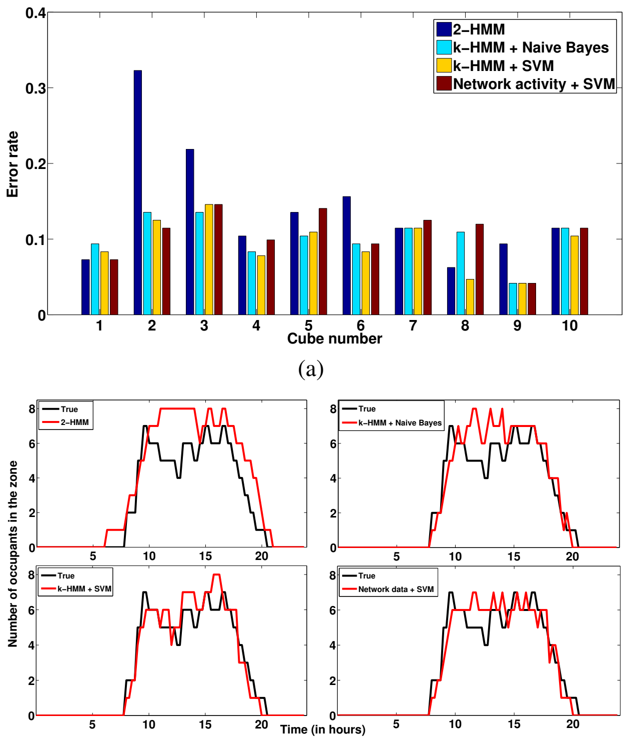
In the first stage of this two stage approach, we model the network data using HMM with k underlying states for the hidden variable, where we choose the value of k that optimizes the log-likelihood function. In the second stage, we train a classifier whose input is a feature vector consisting of the output state of k-HMM along with other external parameters such as time of the day, day of the week, etc. This approach remains tractable even on large datasets while efficiently incorporating the effect of any external parameters on the occupancy.

There are two other key advantages of using this first stage over just a supervised algorithm with network activity as one of the inputs. The first advantage is that it significantly reduces the labeling effort of an occupant during the training phase, where an occupant can now provide binary labels to the k output states of the HMM rather than providing their occupancy logs over time. This also addresses the issue of privacy to some extent. The other advantage of using the k-HMM is that it significantly reduces the size of the input feature space as k is usually very small compared to the total number of possible states for the network data. The result of a smaller feature space is that it requires less training data to efficiently train a classifier.

6.2 Experimental Results

Our first experiment compares the performance of 4 different algorithms. The first algorithm is based on HMM with 2 underlying states, which is a completely unsupervised approach. The second and the third algorithms are semi-supervised approaches using a k-state HMM as described in Section 6.1, followed by a naive Bayes or a support vector machine (SVM) [3] based classifier. The last algorithm is a supervised approach which learns a classifier with input as the network data and other features such as time of day, and output as the occupancy state.

To quantify the estimation accuracy of these algorithms, we collected the ground truth data from 10 different occupants over a period of 16 days. The occupants maintained their occupancy logs by marking their presence in their cube at a time resolution of 30 minutes, where an occupant marks his presence only if he is present for majority of that 30 minute period. The in and out flowing network data is also collected every 30 minutes. The results of the experimental evaluation are shown in Figure 7. Figure 7(a) compares the average error rate of the 4 algorithms for each cube, where the error rate is averaged over 16 different test cases. In each test case, the last 3 algorithms are trained using 15 days of occupancy data and tested on the left out day. Note from this figure that the 2-state HMM in spite of being an unsupervised approach performs well, with an error rate less than 15% in 8 of the 10 cubes. In addition, k-state HMM along with SVM does marginally to significantly better than a completely supervised approach, again in 8 of the 10 cubes. Figure 7(b) compares the true occupancy of a zone comprised of the 10 occupants with that estimated using the 4 different approaches. Note from this figure that k-HMM+SVM performs



(a) (b)

Figure 7. (a) Average error rates of models. (b) Estimated occupancy vs. ground truth.

the best, and 2-HMM, in spite of being an unsupervised approach, does well.

Our next experiment estimates the energy savings that can be obtained by using an occupancy-based lighting schedule. The current lighting schedule in our building is such that the lights are switched on as occupants arrive in the morning and the lights of all zones on one floor (about 159 cubicles) are set to turn off automatically at 9 pm. One of the 39 meters captures the lighting load on this floor. We demonstrate the potential energy savings by implementing an occupancy-based schedule for switching off the lights at night, where the occupancy is estimated using 2-HMM. The use of this completely unsupervised approach is justified since 2-HMM usually over-estimates the leaving time of an occupant, thus providing a conservative estimate of the time when all the occupants in a zone would have left. Figure 8 demonstrates the average energy savings for 13 different zones in one of the buildings, using two approaches. The first is static scheduling where for each zone, the lights are re-scheduled to turn off based on the worst case scenario observed over all days for that zone. This approach does not offer much savings as there could be a rare incident where an occupant of a zone stays till 9 pm or later. The second approach is to dynamically schedule the switch-off time for lights in a zone based on the estimated occupancy of that zone. That is, the lights in a zone are turned off when the estimated occupancy of that zone becomes zero. This time may vary each day. Note from Figure 8 that this approach provides significant savings in the lighting energy. Overall, the proposed approach provides around 9.53% in savings for the building in consideration whose peak lighting load is 45 kW.

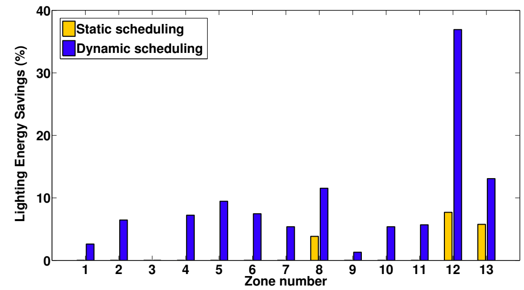


Figure 8. Percentage energy savings by using an occupancy-based lighting schedule.

7 Conclusions

Energy efficiency in commercial buildings is important since they are significant consumers of power, e.g., about 37% of the total electricity generated in the US. With the broad goal of understanding the power consumption of a campus, we collect real data and propose techniques for (1) anomaly detection, to identify energy saving opportunities, and (2) occupancy modeling, to better manage lighting and HVAC. The initial results are promising. Several instances of unusual power consumption were detected, many of which could result in savings. Further, we discovered that an occupancy based lighting schedule on a floor could provide average power savings of about 9.5%. Our future plans include deploying a more extensive metering infrastructure, to enable a more complete understanding of how power (and other resources) are used on our campus.

8 References

[1] Y. Agarwal, T. Weng, and R. Gupta. The energy dashboard: improving the visibility of energy consumption at a campus-wide scale. In *BuildSys*, Berkeley, CA, November 2009.

[2] V. Catterson, S. McArthur, and G. Moss. Online conditional anomaly detection in multivariate data for transformer monitoring. *IEEE Transactions on Power Delivery*, 25(4):2556–2564, 2010.

[3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, New York, 2. edition, 2001.

[4] V. Erickson and A. Cerpa. Occupancy based demand response HVAC control strategy. In *BuildSys*, Zurich, Switzerland, November 2010.

[5] G. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(2):1870–1891, 1995.

[6] V. Jakkula and D. Cook. Outlier detection in smart environment structured power datasets. In *IEEE Intelligent Systems*, London, UK, July 2010.

[7] Y. Kim, R. Balani, H. Zhao, and M. Srivastava. Granger causality analysis on ip traffic and circuit-level energy monitoring. In *BuildSys*, Zurich, Switzerland, November 2010.

[8] X. Li, C. Bowers, and T. Schnier. Classification of energy consumption of a building with outlier detection. *IEEE Transactions on Industrial Electronics*, 57(11):3639–3644, 2010.

[9] S. McArthur, C. Booth, J. McDonald, and I. McFadyen. An agent-based anomaly detection architecture for condition monitoring. *IEEE Transactions on Power Systems*, 20(4):1675–1682, 2005.

[10] R. Melfi, , B. Rosenblum, B. Nordman, and K. Christensen. Measuring building occupancy using existing network infrastructure. *International Green Computing Conference*, 2011.

[11] G. Newsham and B. Birt. Building-level occupancy data to improve ARIMA-based electricity use forecasts. In *BuildSys*, Zurich, Switzerland, November 2010.

[12] J. Seem. Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy and Buildings*, 29(1):52–58, 2007.

[13] United States Energy Information Administration. Annual energy review 2009. `http://www.eia.gov/aer`, 2010.