EnergyTrack : Sensor-Driven Energy Use Analysis System

Deokwoo Jung[1], Varun Badrinath Krishna[1], Ngo Quang Minh Khiem[1], Hoang Hai Nguyen[1], David K. Y. Yau[1,2]
[1]Advanced Digital Sciences Center, Illinois at Singapore
[2]Singapore University of Technology and Design, Singapore

ABSTRACT

Demand side management (DSM) has emerged as a promising way to balance the electrical grid's demand and supply in an economical and environmentally friendly manner. For successful DSM, it is crucial to automate the analysis of building energy usage with respect to important factors that drive it, such as occupancy. In this paper, we present a sensor-driven energy use analysis system, *EnergyTrack*, that continuously analyzes, evaluates, and interprets building energy use in real-time. We develop an energy usage model in *EnergyTrack* that simultaneously considers device-specific energy consumption, occupancy changes, and occupant utility. We also design a low-cost occupancy estimation algorithm with a lightweight training requirement. The *EnergyTrack* testbed is implemented in a commercial building office space. Through this testbed, we demonstrate the performance of our occupancy estimation algorithm and the application of *EnergyTrack* in energy use analysis.

Categories and Subject Descriptors

H.1 [**Information Systems**]: Models and Principles; C.3 [**Special-Purpose and Application-Based Systems**]: Real-time and Embedded Systems

General Terms

Design,Algorithms,Performance,Measurement

Keywords

Smart Grid, Demand Side Management,Measurement & Verification, Wireless Sensor Network

1. INTRODUCTION

In recent years, both industry and academia have been jointly striving to make electrical grids more efficient with two-way communication of information and control between electricity suppliers and consumers. Such *smart grids* facilitate demand side management (DSM), which uses monetary incentives to encourage consumers to adapt their consumption patterns. DSM has emerged as a promising way to balance the grid's electricity demand and supply without increasing the grid's generation capacity. It may lead to both economical and environmentally friendly use of electricity.

In order to facilitate DSM, it is crucial to have reliable, accurate, and verifiable methods to continuously perform Measurement & Verification (M&V), which is the process of using measurements to assess the energy consumption for pre- and post-DSM periods. An international standard [4] suggests several best-practice guidelines for M&V and correlates energy use to relevant driving factors such as occupancy and weather conditions. In particular, it emphasizes the importance of *Energy Monitoring and Targeting* (M&T), which is the procedure of continuously performing the M&V process to provide energy managers with constant feedback to help improve the control of energy use. It is reported that M&T can typically achieve an additional 7%-12% of energy savings compared to other one-time M&V methods, by persistent energy saving [4]. Despite its advantage, it is difficult to seamlessly and correctly practice M&T due to the lack of low-cost solutions to automate such M&T processes with a good level of accuracy.

In this paper, we attempt to address the problem by proposing a middleware architecture for a sensor-driven energy use analysis system, *EnergyTrack*, that continuously analyzes, evaluates, and interprets energy usage in buildings using real-time sensor data. In order to evaluate the efficiency in energy use, we develop an analytical energy usage model that simultaneously considers device-specific energy consumption, occupancy, and occupant utility (e.g. comfort level). Our model allows users to systematically quantify the useful and wasted energy use by jointly considering occupancy and occupant utility. It deems more energy to be wasted when occupancy or occupant utility is low. There are two major advantages that this model has over existing models [4], which only consider static baseline consumption to quantify energy savings. Firstly, our model naturally accounts for the trade-off between energy saving and the impact it has on occupant utility. Secondly, it also favors the energy spent during periods of high occupancy. These two features quantitatively incorporate the intuition that energy usage efficiency of a load is maximized when the maximum number of concerned occupants experience the highest utility from its operation.

We also develop an occupancy estimation algorithm from motion sensors and CO2 sensors with knowledge of only the maximum number of occupants. Compared to exist-

BuildSys '13, November 14-15 2013, Roma, Italy.

ing training-based techniques, the proposed algorithm has a major advantage in many practical settings as it relaxes the requirement of expensive training data sets of occupancy ground truth. Our proposed middleware architecture and energy use analysis framework is employed for designing *EnergyTrack* through seamless integration of four sub-system modules: sensor network, database, data analytics engine, and user interface. The *EnergyTrack* testbed is implemented in a commercial building office space with 90 occupants. In this testbed, we evaluate the performance of our occupancy estimation algorithm and demonstrate the application of *EnergyTrack* in energy use analysis.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 proposes an analysis framework to quantify energy savings. Section 4 develops algorithms for occupancy estimation. Section 5 describes the system architecture of *EnergyTrack* and presents our testbed implementation. Section 6 evaluates the performance of the proposed occupancy estimation algorithm and analyzes empirical results from *EnergyTrack*. Section 7 concludes.

2. RELATED WORK

In recent years, researchers in the wireless sensor network (WSN) community have developed large-scale energy monitoring systems [8, 7, 3] and analysis models for building energy consumption [5]. Our work differs from these studies in that we provide a general framework that jointly evaluates occupant utility and energy consumption, as opposed to evaluating them in isolation. To the best of our knowledge, we present a first analytical framework and system architecture that systematically quantifies energy savings using real-time sensor data.

In [8], the authors develop a wireless sensor platform, *ACme*, for large-scale energy monitoring and actuation of plug loads in buildings. In [7, 3], *ACme* is deployed for the empirical study of a large and long-lived appliance energy monitoring system. In [5], the authors propose a new method to automatically detect abnormal energy usages from power meter measurements using Empirical Mode Decomposition (EMD). They identify the anomalous behavior of devices in an unsupervised manner by exploring the correlations between devices at different time scales. They do not consider occupancy in their analysis. We use conditional probabilities to flag out such anomalies, while considering the dependency between device energy consumption and occupancy.

There are several industry standards related to analysis frameworks for building energy savings. The International Performance Measurement and Verification protocol [4], which we refer to as the M&V protocol, provides a general guideline to determine the baseline energy consumption and savings potential due to energy conservation measures in buildings. The predicted percentage dissatisfied (PPD) metric [6] is an ISO standard that provides an analytical model to quantify occupant thermal comfort from parameters, such as temperature, humidity, clothing and metabolic rate. We incorporate these guidelines proposed into our analytical framework and estimate key parameters, such as occupancy and thermal comfort, using real-time sensor data.

The occupancy level in buildings is one of the most important parameters for accurately quantifying energy savings. It has been shown that real-time occupancy information can greatly reduce energy consumption [1]. Therefore, occupancy estimation using WSN has been an active area of research [11, 12]. In [11], the authors show that PIR sensors alone cannot support reliable occupancy estimation. This led them to employ cameras deployed in public hallways along with PIR sensors [12]. Our approach is similar to theirs in that we also use complementary sensors (i.e CO2 sensors) with PIR sensors. However, our approach does not require training or a calibration process, as our algorithm (cf. Algorithm 1) simultaneously estimates both model parameters and occupancy given sensor measurements. Hence, our solution is potentially less sensitive to the locations of sensors in the building, which can reduce WSN deployment efforts.

3. ENERGY USAGE ANALYSIS

3.1 Taxonomy and Assumptions

Generally, energy consumption is governed by complex interactions between occupants, (electrical) end-loads and zones. A zone is defined as a logical space that supports a certain type of occupant activity. For example, a pantry, an office cubicle, or an aisle can be considered to be an individual zone. Let us consider a well-defined closed space of a building (e.g. office space), and refer to it as a root zone. This root zone has a fixed number of sub-zones and end-loads. It is also associated with a maximum occupancy that denotes the maximum number of people who are simultaneously present in the zone. We only consider energy that is consumed within this root zone and ignore any consumption that originates from outside of it.

We assume that all end-loads provide a certain type of benefit for building occupants, such as entertainment, work productivity, comfort, etc. We refer to these benefits collectively as a *service*. We use a metric, *utility*, to quantify how effectively the end-load's energy is spent to provide such service. This utility metric is essential to analyzing operational energy efficiency of various loads with respect to occupants. For example, the energy wasted by air conditioners is largely dependent on the occupants' subjective comfort level (e.g. too cold or too warm). This is because the cost of maintaining this comfort needs to be simultaneously considered.

End-loads may service a single zone or multiple zones when they operate. We define the *utility coverage* of an end-load as the spacial range within which an occupant can enjoy its service. Some end-loads, such as refrigerators and computer servers, can provide service regardless of occupant locations and thus their utility coverage extends to all zones, whether inside or outside of the root zone. Other end-loads, such as desktop computers, require occupants to be in their proximity and thus their utility coverage typically extends to only the sub-zone to which they belong.

3.2 Model Formulation

We now formalize the building energy usage model as a function of occupants, end-loads, and sub-zones. Let $\mathcal{G}$ denote the two dimensional space of a root zone. We denote the number of end-loads, sub-zones, and the maximum number of occupants by N, L, and M respectively. Let $\mathbb{V} = (v_0, v_1 \cdots v_L)$ be a mutually exclusive set of zones where $v_{i>0} \in \mathcal{G}$ are sub-zones and $v_0 \notin \mathcal{G}$ for any locations outside the root zone. Let $\mathbb{U} = (u_1, \cdots u_N)$ be the set of end-load locations where $u_i \in v_{\{1 \leq j \leq L\}}$.

3.2.1 Mean utility

We assume that each occupant stays in a zone for a random amount of time, then makes a random transition to a different zone. Let $a_i(t)$ be the location of occupant i at time t where $a_i(t) \in \mathbb{V}$. Let $I_A(x)$ denote the indicator function of A where $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise. We define occupant m's occupancy of zone v_l between time t_1 and t_2 as follows:

$$a_{ml}(t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} I_{v_l}(a_i(t))\, \mathrm{d}t, \tag{1}$$

where $0 \le a_{ml}(t_1, t_2) \le 1$.

Let $q_{nl}(t) \in \mathbb{R}_{[0\ 1]}$ denote the utility at zone l delivered by end-load n at time t. Let $\tilde{q}_{nm}(t_1, t_2)$ denote the utility provided by end-load n and delivered to occupant m during the time (t_1, t_2), expressed as

$$\tilde{q}_{nm}(t_1, t_2) = \frac{1}{t_2 - t_1} \sum_{l=0}^{L} \int_{t \in T_{ml}(t_1, t_2)} q_{nl}(t)\, \mathrm{d}t,$$

where $T_{ml}(t_1, t_2)$ is the time period for which occupant m stays in zone l during the time (t_1, t_2).

Let $\bar{q}_n(t_1, t_2)$ denote the weighted mean of utility per occupant that we refer to as *mean utility*. Then the mean utility can be defined as:

$$\bar{q}_n(t_1, t_2) = \sum_{m=1}^{M} w_{nm} \tilde{q}_{nm}(t_1, t_2) \tag{2}$$

where w_{nm} is the weight coefficient for the average such that $w_{nm} \in \mathbb{R}_{[0\ 1]}$ and $\sum_{m=1}^{M} w_{nm} = 1$. It can be easily verified that $0 \le \bar{q}_n(t_1, t_2) \le 1$. The weight coefficient w_{nm} is determined by the occupant m's ownership of end-load n.

If the end-load n is solely owned by occupant m, then $w_{nm} = 1$ and $w_{nk} = 0$ for $k \ne m$. Similarly, we have $w_{nm} = \frac{1}{M}$ if the end-load n is equally shared by all occupants. In plain terms, the metric $\bar{q}_n(t_1, t_2)$ is the ratio of the utility delivered to its intended occupants, to the total utility provided by end-load n between times t_1 and t_2. It can be easily seen that $\bar{q}_n(t_1, t_2) = 0$ if there are no occupants within the utility coverage of end-load n, which implies that all energy might be wasted during the time. Conversely, it is a strong indication of more efficient energy use of end-load n if $\bar{q}_n(t_1, t_2)$ is close to 1.

3.2.2 Analysis of useful energy usage

Let $E_n(t_1, t_2)$ denote the energy consumption (in kWh) of end-load n between times t_1 and t_2. Let us assume that an end-load constantly consumes a certain amount of energy, which is deemed to be mandatory for the functionality and/or safety of its operation. We refer to this consumption as *static consumption*. We assume that the rest of the consumption is controllable and can be further optimized based on utility and occupancy. We refer to this consumption as *dynamic consumption*. Let us consider the following breakdown of consumption, $e_n(t_1, t_2) = e_n^s(t_1, t_2) + e_n^d(t_1, t_2)$ where e_n^s and e_n^d denote static and dynamic consumption, respectively. We propose an energy usage analysis model for an end-load, given its utility and occupancy, as follows:

$$\begin{aligned} e_n(t_1, t_2) \;=\; & \underbrace{e_n^s(t_1, t_2)}_{\text{Static Consumption, } e_n^s} + \underbrace{\bar{q}_n(t_1, t_2) e_n^d(t_1, t_2)}_{\text{Useful Consumption, } e_n^u} \\ & + \underbrace{(1 - \bar{q}_n(t_1, t_2)) e_n^d(t_1, t_2)}_{\text{Wasted Consumption, } e_n^w}. \end{aligned} \tag{3}$$

The formula in (3) naturally embeds the cause (i.e. occupancy) and the effect (i.e. utility) of energy consumption into the usage analysis model. Note that the wasted consumption E_n^w, is an aggressive estimate and can be thought of as the maximum energy savings that can be realized by employing DSM. The model can be written in matrix form where all the measurements are grouped into discrete time intervals over a single time reference, thereby avoiding large overheads in retrieving and processing data. Let us consider a time interval of $\Delta T = 15$min by default, for 24 hours of a day. We denote the discrete time reference $T_{ref} = \{k\Delta T \mid k = 0, \cdots, \frac{1440}{\Delta T}\}$. Let us define the following matrix at a discrete time k: *a)* $[\mathbf{O}_t]_{ml} = a_{ml}(t_k, t_k + \Delta T)$; *b)* $[\mathbf{Q}_t]_{nl} = q_{nl}(t_k, t_k + \Delta T)$; and *c)* $[\mathbf{W}]_{nm} = w_{nm}$. Let $\bar{\mathbf{q}}_k$ denote a vector of the mean utilities of n end-loads where $[\bar{\mathbf{q}}_k]_i = \bar{q}_i(t_k, t_k + \Delta T)$ for end-load i. Then it can be easily verified that a vector of mean utility of N end-loads at time index k is $[\bar{\mathbf{q}}_k]_i = [\mathbf{O}_k \mathbf{Q}_k \mathbf{W}]_{ii}$.

Let $\mathbf{e}_k$ be an $N \times 1$ vector of the energy consumption of n end-loads where $[\mathbf{e}_k]_n = e_n(t_k, t_k + \Delta T)$. The energy usage analysis in (3) can be rewritten in matrix form as follows:

$$\mathbf{e}_k = \mathbf{e}_k^s + diag(\bar{\mathbf{q}}_k)\mathbf{e}_k^d + (\mathbf{I} - diag(\bar{\mathbf{q}}_k))\mathbf{e}_k^d, \tag{4}$$

where $\mathbf{e}_k^s$ and $\mathbf{e}_k^d$ are vectors of static and dynamic consumptions of N end-loads, respectively.

4. OCCUPANCY ESTIMATION

Let $a_l(t_1, t_2)$ and $\bar{a}_l(t_1, t_2)$ denote the occupancy and the average occupancy level, respectively, at zone v_l by M number of occupants between time t_1 and t_2. Their definitions are shown in (5).

$$a_l(t_1, t_2) = \sum_{m=1}^{M} a_{ml}(t_1, t_2); \quad \bar{a}_l(t_1, t_2) = \frac{1}{M} a_l(t_1, t_2), \tag{5}$$

where M is the maximum number of occupants in zone l. We develop an algorithm to estimate the occupancy in (5) using CO2 and PIR motion sensors, given lower and upper bound information of the occupancy $a_l(t_1, t_2)$. It can be easily seen that the initial bound is $0 \le a_{ml}(t_1, t_2) \le M$ by default. Our algorithm computes the occupancy estimates independently using measurements from each CO2 and PIR sensor, and then heuristically combines them for the final estimate. The main advantage of our algorithm in practical settings is that it can robustly estimate the occupancy given only the maximum occupancy M.

For discrete times $t = 1, \cdots, n$ with a time interval ΔT, let $\mathbf{x}^c = (x_1^c, \cdots, x_n^c)$ and $\mathbf{x}^p = (x_1^p, \cdots, x_n^p)$ denote a vector of the average of CO2 sensor measurements and the total number of PIR sensor triggering events, respectively. Similarly, let $\hat{\mathbf{a}}^c = (\hat{a}_1^c, \cdots, \hat{a}_n^c)$ and $\hat{\mathbf{a}}^p = (\hat{a}_1^p, \cdots, \hat{a}_n^p)$ denote a vector of the total occupancy estimates from CO2 sensors and PIR sensors, respectively. Note that we use the notation a_t for the total occupancy at time index t in place of a zone l in a_l to simplify notations.

We use the autoregressive moving average model (ARMA) for our estimation function, mapping $\mathbf{x}^* \mapsto \hat{\mathbf{a}}^*$ as follows:

$$\hat{a}_t^c = \alpha_c + \sum_{i=0}^{\tau_c} \alpha_i x_{t-i}^c, \quad \hat{a}_t^p = \beta_p + \sum_{i=0}^{\tau_p} \beta_i x_{t-i}^p \qquad (6)$$

where the model orders of ARMA, τ_p and τ_c, depend on the time interval ΔT. Note that the ARMA model of a CO2 sensor for $\tau_c = 1$ has been experimentally verified [9].

The model coefficients α and β can be found by linear regression from training data sets $(\mathbf{a}^g, \mathbf{x}^p)$ and $(\mathbf{a}^g, \mathbf{x}^p)$, respectively, where $\mathbf{a}^g$ gives the ground truth measurements of occupancy $\mathbf{a}$. However, the exact $\mathbf{a}^g$ is generally not available or expensive to obtain. Instead, it is more realistic to assume that only rough lower or upper bounds of $\mathbf{a}^g$ are available. Let $\mathbf{a}^l$ and $\mathbf{a}^u$ denote the lower and upper bounds of $\mathbf{a}^g$ such that $0 \leq \mathbf{a}^l \leq \mathbf{a}^g \leq \mathbf{a}^u \leq M$. Our algorithm iteratively computes optimal estimates of the coefficients α and β from the data sets $(\mathbf{a}^l, \mathbf{a}^u, \mathbf{x}^c)$ and $(\mathbf{a}^l, \mathbf{a}^u, \mathbf{x}^p)$, respectively. Our iterative solution is derived based on the general algorithmic framework of Expectation Maximization (EM) for linear regression with incomplete data [10].

We present the detailed solution of an EM-based estimator in Algorithm 1. Let us consider the observed data of $\mathbf{y}$ for dependable variables and $\mathbf{x}$ for explanatory variables, and the coefficient vector β. We would like to estimate β that best explains the linear model $\mathbf{y} = \mathbf{x}\beta + \mathbf{e}$, where $\mathbf{e}$ is a model error with the Normal distribution, and

$$\mathbf{x} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1\tau} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & \cdots & x_{n\tau} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_\tau \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Without loss of generality, we assume that the first m samples of $\mathbf{y}$ are observed, but for the rest of the samples only their bounds are given such that $-\infty \leq y_i^l \leq y_i \leq y_i^u \leq \infty$ for $i = m+1, \cdots, n$. Let $\mathbf{y}^m$,$\mathbf{y}^l$. Let $\mathbf{y}^u$ denote a vector of the observed $(y_i)_{i=1,\cdots,m}$, the lower bounds $(y_i^l)_{i=m+1,\cdots,n}$, and the upper bounds $(y_i^u)_{i=m+1,\cdots,n}$. In general, no closed form solutions exist for the optimal maximum likelihood estimator of β. Instead, a solution can be found by iteratively updating estimates $\hat{\beta}$ and $\hat{\mathbf{y}}$ by the EM algorithm.

Algorithm 1 Estimate β from $(\mathbf{x}, \mathbf{y}^m, \mathbf{y}^l, \mathbf{y}^u)$

---

1. Initialization: $\hat{\sigma}^{(0)}, \beta^{(0)}, \mu^{(0)}$
2. Expectation step: update $\mathbf{y}^{(r)}$
   $y_i^{(r)} = y_i$ for $i = 1, \cdots, m$
   $y_i^{(r)} = E_{[y_i^l, y_i^u]}(y_i)$ for $i = m+1, \cdots, n$
3. Maximization step: update $\beta^{(r)}, \mu^{(r)}, \hat{\sigma}^{2(r)}$
   $\beta^{(r)} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}^{(r)}$
   $\mu_i^{(r)} = \beta_i^{(r)}\mathbf{x}$
   $\hat{\sigma}^{2(r)} = \frac{1}{n}\left\{\sum_{i=1}^{m}(y_i^{(r)} - \mu_i^{(r)})^2 + \sum_{i=m+1}^{\tau} Var_{[y_i^l, y_i^u]}(y_i)\right\}$
4. Evaluation: convergence test for $\hat{\sigma}^{(r)}$
   Repeat 2,3,4 until $\hat{\sigma}^{(r)}$ converges.

The EM algorithm is shown in Algorithm 1. For brevity, we omit its derivation. In Algorithm 1, we use shorthand notations for the conditional expectation and variance of y_i: $E_{[y_i^l, y_i^u]}(y_i) = E[y_i | y_i^l \leq y_i \leq y_i^u]$ and $Var_{[y_i^l, y_i^u]}(y_i) = Var[y_i \mid y_i^l \leq y_i \leq y_i^u]$. Given that the random variable $\{Y_i | y_i^l \leq Y_i \leq y_i^u\}$ follows a truncated Normal distribution, it can be easily verified that closed form formulas for $E_{[y_i^l, y_i^u]}(y_i)$ and $Var_{[y_i^l, y_i^u]}(y_i)$ are

$$\begin{aligned} E_{[y_i^l, y_i^u]}(y_i) =& \mu_i^{(r)} + \hat{\sigma}^{(r)} \frac{\phi(z_i^{l,(r)}) - \phi(z_i^{u,(r)})}{\Phi(z_i^{u,(r)}) - \Phi(z_i^{l,(r)})} \\ Var_{[y_i^l, y_i^u]}(y_i) =& \hat{\sigma}^{2(r)} + \hat{\sigma}^{2(r)} \frac{z_i^{l,(r)}\phi(z_i^{l,(r)}) - z_i^{u,(r)}\phi(z_i^{u,(r)})}{\Phi(z_i^{u,(r)}) - \Phi(z_i^{l,(r)})} \\ & - \hat{\sigma}^{2(r)} \left(\frac{\phi(z_i^{l,(r)}) - \phi(z_i^{u,(r)})}{\Phi(z_i^{u,(r)}) - \Phi(z_i^{l,(r)})}\right)^2 \end{aligned} \qquad (7)$$

where $z_i^{l,(r)} = \frac{y_i^l - \mu_i^{(r)}}{\hat{\sigma}^{(r)}}$ and $z_i^{u,(r)} = \frac{y_i^u - \mu_i^{(r)}}{\hat{\sigma}^{(r)}}$, and ϕ and $\Phi(x)$ denote the probability density function and cumulative distribution function of the standard normal distribution, respectively.

The final occupancy estimate at time t is the weighted average of the two estimates shown below:

$$\hat{a}_t = \frac{w_t^c \hat{a}_t^c + w_t^p \hat{a}_t^p}{w_t^c + w_t^p} - \hat{a}^0, \ w_t^c = \frac{|\hat{a}_t^p|}{M}, \ w_t^p = \frac{|M - \hat{a}_t^c|}{M} \qquad (8)$$

where $\hat{a}^0$ is the occupancy offset constant found by $\hat{a}^0 = \min(\{\hat{a}_i | \hat{a}_i < 0\})$. The weight coefficient w_t^c penalizes the estimate by CO2 sensors $\hat{a}_t^c$ as the PIR sensor detects low occupancy, i.e. $\hat{a}_t^p \to 0$. Similarly, we can interpret w_t^p as a penalty coefficient for $\hat{a}_t^p$ given $\hat{a}_t^c$.

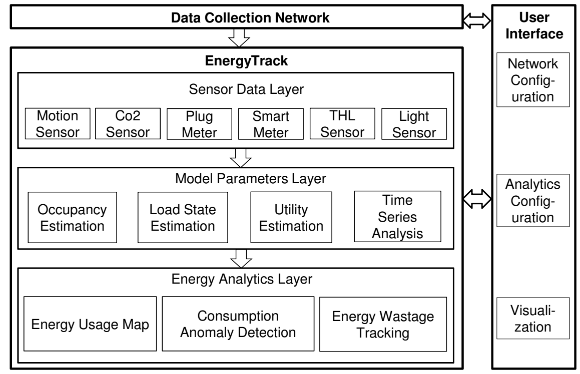
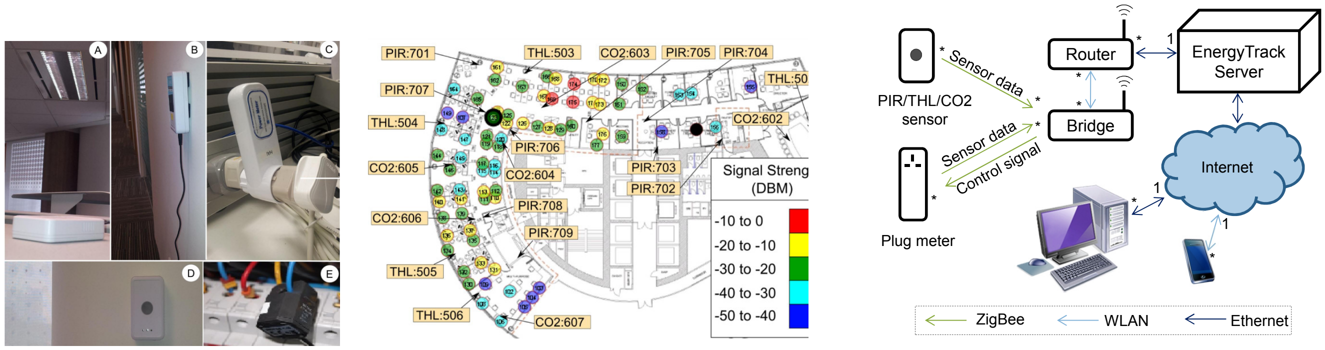


Figure 1: System Architecture of EnergyTrack

5. SYSTEM DESIGN

5.1 Overview

EnergyTrack is designed with a multi-layered and modular architecture to provide flexibility and scalability. Figure 1 shows the overall system architecture of *EnergyTrack*. It consists of three main layers: sensor data layer (SDL), model parameter layer (MPL), and energy analytics layer (EAL). It is supported by the data collection network (DCN) and a user interface layer (UIL). The DCN gives a network infrastructure abstraction to deliver sensor data (e.g., temperature or power measurements) to the SDL. The UIL allows users to easily modify configuration settings of *EnergyTrack* and access the various analytics results. The SDL provides



(a) (b) (c)

Figure 2: Testbed illustration: (a) Sensors types in deployment: A:CO2, B:PIR, C:Plug Meter, D:Temp./Humd./Lux., E:Current Transducers at sub-branch, (b) Wireless sensor deployment map, (c) Data collection network.

a long-term repository for raw data collected from the DCN. The MPL consists of functions to estimate the following key model parameters: occupancy level, mean utility, and end-load states. It also performs a time series analysis that finds temporal correlations of end-load energy consumption with parameters such as occupancy level. In particular, it finds the best linear or quadratic fitting function of energy consumption for the estimated occupancy levels.

The EAL uses parameters from the MPL to provide three key features: *a*) Energy wastage tracking (EWT); *b*) Consumption anomaly detection (CAD); and *c*) Energy usage map (EUM). The EWT evaluates energy wastage (or equivalently, potential energy savings opportunities) in real-time using the energy usage analysis model in (4). The CAD finds abnormal consumptions that do not follow the inferred usage patterns, i.e., fitted functions of consumption, given occupancy rate provided by the MPL. The EUM hierarchically maps energy consumption over directed tree graphs of end-use loads, occupants, and zones.

5.2 EnergyTrack Testbed

EnergyTrack is implemented in an office space in a commercial building, where about 90 occupants regularly work during the business hours of 8am-6pm on weekdays.

5.2.1 Data Collection Network

The testbed system consists of a wireless sensor network of 80 plug meters, 5 THL sensors, 5 CO2 sensors, 9 PIR sensors, and a panel-level power monitoring system, as shown in Figure. 2(a). The wireless sensor network consists of motes running TinyOS with a TI MSP430F1611 MCU (8MHz clock rate and 10KB RAM) and a Chipcon CC2420 Zigbee radio. The power panel monitoring system samples and reports power consumption of electrical branches at the main switch board (MSB) of our office testbed, as well as its 42 sub-branches, every 1.8 seconds. The sub-branches are grouped into plug, lighting, and server use. They allow us to analyze consumption by load-types. The plug meters are mostly installed at individual desktop computers reporting power consumption every second by default. The locations of the THL, CO2, and PIR sensors are carefully chosen for the deployment according to their sensing requirements. Deployment details of the wireless sensor network are shown in Figures 2(b) and (c).

5.2.2 Sensor Data Layer

The SDL is implemented using MySQL 5.1.49. From all the sensors combined, we collect and store approximately 6.7 million data points every day. Different database designs were tested and ultimately designs favoring fast read times were chosen over those that increase modularity. We maintain a DB table for each sensor node to reduce the SQL query execution time. We also take averages of sensor data every 15 minutes and store these in separate tables. This caching method greatly reduces the execution time for search queries with acceptable storage redundancy.

In our current testbed, the consumption data of the HVAC system is not available as the entire building management system (BMS) is confidentially managed by a private building management company. Instead, we use data generated from EnergyPlus simulations, a method that is described in great detail in [2]. To drive realistic simulations we create models in EnergyPlus using detailed building specifications such as thermal envelope parameters, floor plan measurements, and air handler unit (AHU) specifications obtained from the building management company. Our simulations reveal that the cooling capacity of our HVAC system is oversized by about three times of what it needs to be. This is common in commercial buildings, because these systems are designed to handle worst case cooling loads. As a consequence of this oversizing, the HVAC system operates inefficiently and draws a large, constant power regardless of occupancy and weather variations.

5.2.3 Model Parameter and Energy Analytics Layer

Our system considers only a root zone (i.e., no sub-zones) since a single AHU serves our office. The hourly occupancy level is estimated for the root zone. Fig. 3(a) shows our energy analytics interface that displays hourly mean utility for the HVAC and lighting systems of the testbed, using thermal and visual comfort metrics. Note that the mean utility is set to 1 for all plug loads. For thermal comfort, we adopt the predicted percentage of dissatisfied (PPD) standard [6]. The PPD metric is a function that predicts the percentage of occupants who are dissatisfied with their thermal comfort, given various environment and human conditions such as air temperature, humidity, metabolic rate, and clothing insulation. Our system updates the PPD value hourly, based on temperature and humidity data from the SDL. Other vari-

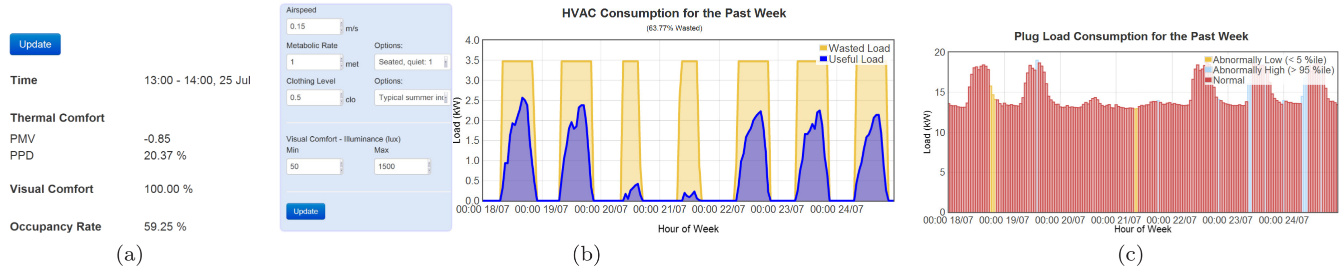


Figure 3: EnergyTrack Implementation: (a) Visualization of real-time parameter estimates: thermal comfort, visual comfort, and occupancy level, (b) Energy wastage tracking for HVAC with thermal and visual comfort parameters that can be adjusted. The wasted energy is the difference between the actual and the useful consumption, (c) Consumption anomaly detection for plug loads with anomalous consumptions marked in blue and yellow.

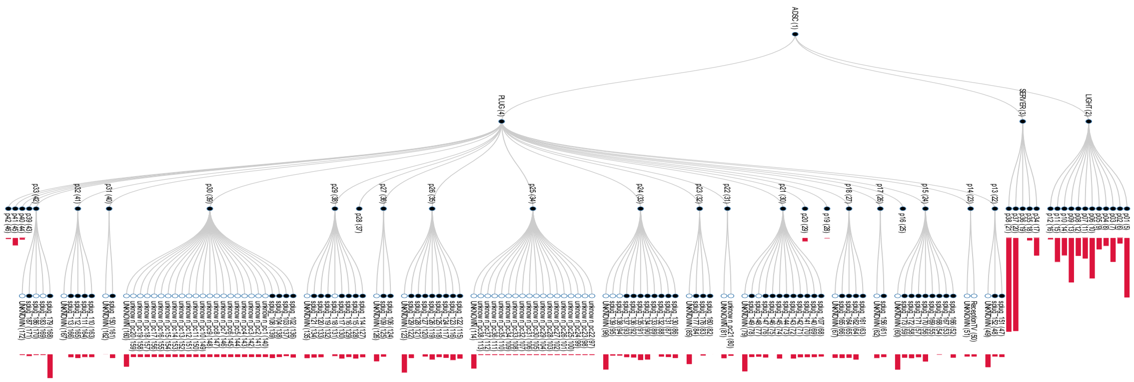


Figure 4: Snapshot of energy usage map by end-loads from the *EnergyTrack* testbed at 10am, 07/18/2013.

ables can be set by users via the UI shown in Figure. 3(a). The visual comfort is evaluated using a logistic function of illuminance (lux) measurements, where its mean utility becomes 1 (or 0) beyond (or below) a certain threshold set by the user. Fig. 3(b) illustrates the HVAC system wastage in our office. This is calculated using occupancy levels as well as thermal comfort parameters on the left panel. Similarly, visual comfort is used to assess lighting load wastage. Figure 3 illustrates the consumption of plug loads, with certain hours exhibiting anomalous consumption. These anomalies could raise alerts, or simply be marked for further investigation. Figure 4 shows a snapshot of the EUM of our testbed. In the figure, the length of the red bar on a leaf node indicates its consumption.

6. EVALUATION

6.1 Occupancy Estimation

We evaluate our occupancy estimation algorithm presented in Section 4. In our evaluation, we estimate the occupancy level in our current *EnergyTrack* testbed every 15 minutes on a weekday (Thursday). This gives us a total of 96 discrete samples for corresponding sensor measurements. The results are compared to ground truth data, which was obtained by manually counting occupants in the testbed. The locations of the CO2 sensors and PIR sensors used for the estimation can be found in Fig. 2(a). For $\Delta T = 15$ minutes, we set $\tau_c = \tau_p = 2$ for the orders of the ARMA model in (6). We set the lower and upper bounds of occupancy to 0 and 90, respectively. The upper bound represents the total number of staff members in our current living lab testbed.

Data set	Max.Bound			Training data
Sensors	PIR	CO2	PIR+CO2	PIR+CO2
Occupancy Level Error	6.81 %	9.47 %	4.06%	3.80%

Table 1: Occupancy level error comparison.

Given the bound information, we run our estimation algorithm for each of the following sensor combinations: PIR sensors alone (cf. $\hat{a}_t^p$ in (6)), CO2 sensors alone (cf. $\hat{a}_t^c$ in (6)), and both CO2 and PIR sensors together (cf. $\hat{a}_t$ in (8)). We also use the ground truth data as a training data set to obtain the best estimation performance given the data set. As a performance metric, we use the absolute difference in occupancy level (%) between the estimate and the ground truth. The occupancy level error at time t is formally defined by $100 \times |\frac{\hat{a}_t}{M} - \frac{a_t^g}{M}|$.

Figure 5 shows the CO2 and PIR measurements in the top plot, which correspond to x_t^c and x_t^p in (6), and the estimates by the algorithm in the bottom plot, for the different combinations of CO2 and PIR measurements with or without training data. Figure 6 shows the occupancy level errors for each of the 96 sample points during a day. It shows that the estimation experiences a large offset error during low-occupancy periods for CO2 sensors alone. There is also high instability during high-occupancy periods for PIR sensors alone. Hence, the two types of sensors during low and high-occupancy periods have complementary performance with each other.

As predicted, estimates by our algorithm using $PIR +$

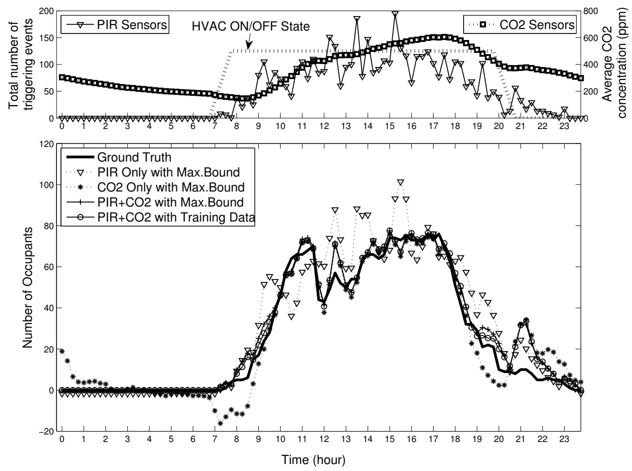


Figure 5: Sensor measurements and HVAC ON/OFF states (top), and estimated total occupancy for various settings vs. the ground truth (bottom).

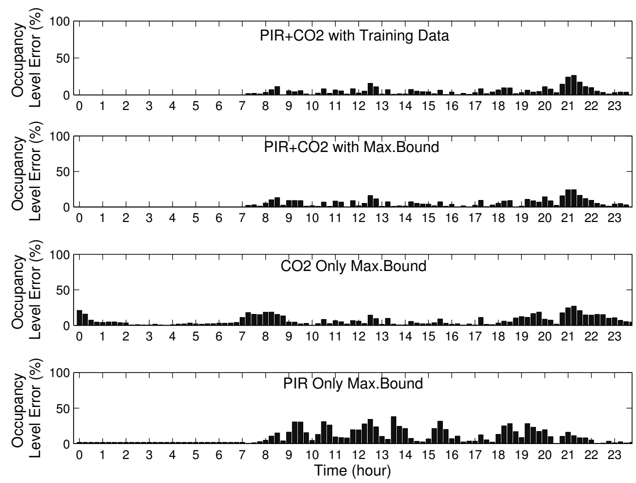


Figure 6: Occupancy level errors during a day for various combinations of CO2 sensors and PIR measurements with and without training data.

$CO2$ with Max.Bound closely follow the ground truth for most of the day, except between 8-10 pm, as show in Figure 5. The performance degradation during these hours occurs because of a sudden drop in the ventilation rate as the HVAC system turns off at 8pm. The occupancy levels for different settings are compared in Table 1. The table shows that our algorithm $PIR+CO2$ with Max.Bound can achieve an error of 4.06%, which is close to the best estimation performance of 3.80%.

6.2 Analysis

We evaluate our EnergyTrack EWT method against a baseline heuristic method. The baseline method corresponds to reducing the estimated wastage without the use of a control system. For lights, the wasted energy is given by the amount of energy consumed when the occupancy is zero. This can be avoided if the last person to leave the office were to switch off the lights. For HVAC, wastage is given by the amount of energy that can be saved if the temperature setpoint were set so that at least 90% of the persons in the office are satisfied per the PPD metric. This saving can be achieved by simply increasing the temperature setpoint to the optimal value, which is 26.7°C for our office. This saving is calculated by running multiple EnergyPlus simulations to determine the AHU load for various indoor temperature setpoints, as illustrated in Figure 7. For computers, we define wastage as the amount of energy consumed when $\bar{X} > \bar{X}_t$ and $\sigma^2 > {\sigma^2}_t$, where $\bar{X}$ and σ^2 are the sample mean and variance, respectively, for half-hour intervals. Setting the thresholds in our office to $\bar{X}_t = 10W$ and ${\sigma^2}_t = 1W^2$ captures the periods when the computers are on but not performing any processing tasks. In these situations, they ought to be switched to standby mode.

Load	Heuristic Method	EnergyTrack
Lighting	458kWh (15%)	2009kWh (66%)
PC	442kWh (14%)	1391kWh (44%)
HVAC	400kWh (32%)	770kWh (62%)

Table 2: Energy Wastage Estimates for One Month

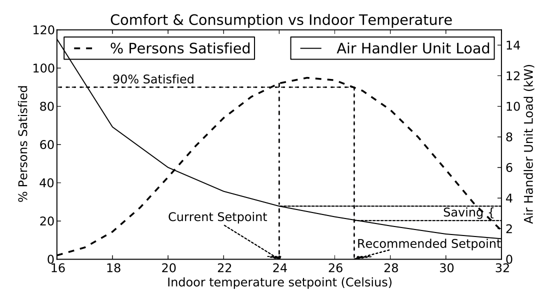


Figure 7: Plotting AHU power consumption as estimated using EnergyPlus, and thermal comfort as expressed as percentage of persons satisfied with the temperature setpoint.

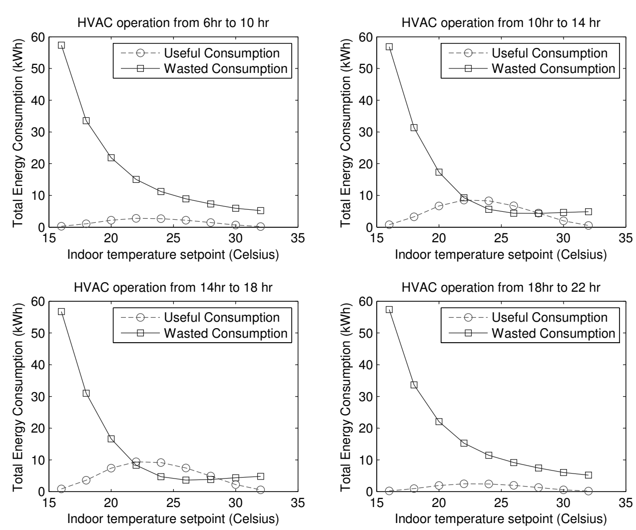


Figure 8: Energy use analysis of HVAC for different temperature setpoints over operation periods.

In addition to the comfort-cost trade-off depicted in Figure 7, *EnergyTrack* incorporates the ground truth occupancy level in Figure 5 and applies Equation (4) to analyze the useful and wasted HVAC consumption for different temperature setpoints. These consumptions are analyzed for each 4-hour-period from 6:00hrs to 22:00hrs, as shown in Figure 8. During the high-occupancy period (10-18:00 hrs), energy is most efficiently used at 24°C where the useful consumption is greater than the wasted consumption. However during the low-occupancy period (6-10:00 hrs and 18-22:00 hrs), no satisfactory temperature setpoint exists since the useful consumption is less than the wastage for all settings. This result is consistent with our intuition, since HVAC systems will be most efficient when the maximum number of occupants experience optimal thermal comfort. For plug loads alone, $\mathbf{e}_k^s$ in Equation (4) is non-zero, and it is estimated by calculating the base load beyond which consumption cannot be reduced.

The two methods are compared using the same set of data for a period of one month. The office staff were not informed about this energy wastage investigation, and thus their consumption behaviors would not have changed during this period. The wastage estimates of *EnergyTrack* are significantly greater than those of the heuristic method as shown in Table 2. The key reason for this difference is that *EnergyTrack* accounts for dynamic changes in occupancy and comfort whereas the heuristic method does not. The table shows the wasted energy for a period of one month, in absolute terms and as a percentage of the total energy consumed by each appliance.

For the one-month period, the heuristic method estimates the total energy wasted to be 1300kWh while *EnergyTrack* estimates this to be 4170kWh. At an electricity tariff of 20c/kWh and no demand rates, the advantage realized by the *EnergyTrack* method over the heuristic method is 574 dollars per month. This effectively quantifies how much a facility manager would stand to gain by installing an automated control system that can realize the full savings potential estimated by *EnergyTrack*, as opposed to relying on manual control based on heuristic methods.

7. CONCLUSION AND FUTURE WORK

In this paper, we present *EnegyTrack*, a system that analyzes and interprets energy consumption patterns in buildings. We propose an analysis model for energy usage that jointly considers occupancy levels and the utility provided by end-loads. Our occupancy estimation algorithm uses PIR and CO2 sensors, and has a lightweight training requirement. Finally, we demonstrate the application of *EnergyTrack* for energy use analysis in a real testbed system. The database and analysis tools of our current *EnergyTrack* testbed are freely available to the scientific community through the Internet[1]. In future work, we intend to deploy *EnergyTrack* in other testbeds to gain insights into consumption patterns in different settings.

8. ACKNOWLEDGMENTS

This work was supported by Singapore's Agency for Science, Technology, and Research (A∗STAR), through a research grant for the Human Sixth Sense Programme at the Advanced Digital Sciences Center, and by the Korea Micro Energy Grid (KMEG) of the Office of Strategic R&D Planning (OSP), Korea government Ministry of Knowledge Economy (No. 2011T100100024).

[1]www.illinois.adsc.com.sg/livinglab/

9. REFERENCES

[1] Y. Agarwal, B. Balaji, S. Dutta, R. Gupta, and T. Weng. Duty-cycling buildings aggressively: The next frontier in hvac control. In *Proceedings of IPSN 2011*, pages 246–257, April.

[2] A. Boyano, P. Hernandez, and O. Wolf. Energy demands and potential savings in european office buildings: Case studies based on energyplus simulations. *Energy and Buildings*, 65(0):19 – 28, 2013.

[3] S. Dawson-Haggerty, S. Lanzisera, J. Taneja, R. Brown, and D. Culler. @scale: insights from a large, long-lived appliance energy wsn. In *Proceedings of IPSN '12*, pages 37–48, New York, NY, USA, 2012. ACM.

[4] Efficiency Valuation Organization. International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings Vol.1, 2012.

[5] R. Fontugne, J. Ortiz, N. Tremblay, P. Borgnat, P. Flandrin, K. Fukuda, D. Culler, and H. Esaki. Strip, bind, and search: a method for identifying abnormal energy consumption in buildings. In *Proceedings of IPSN '13*, IPSN '13, 2013.

[6] ISO 7730:2005. Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria, 2005.

[7] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. Culler. Design and implementation of a high-fidelity ac metering network. In *Proceedings of IPSN '09*, pages 253–264, Washington, DC, USA, 2009. IEEE Computer Society.

[8] X. Jiang, M. Van Ly, J. Taneja, P. Dutta, and D. Culler. Experiences with a high-fidelity wireless building energy auditing network. In *Proceedings of SenSys '09*, pages 113–126, New York, NY, USA, 2009. ACM.

[9] S. A. Mumma and J. Jeong. Field experience controlling a dedicated outdoor air system. *ASHRAE Transactions*, 111(2):433, 2005.

[10] M. Stewart. On least squares estimation when the dependent variable is grouped. *Review of Economic Studies*, 50(4):737–53, 1983.

[11] A. E. C. Varick Erickson, Miguel A. Carreira-Perpinan. Observe: Occupancy-based system for efficient reduction of hvac energy. In *Proceedings of IPSN '11*, IPSN '11, 2011.

[12] A. E. C. Varick Erickson, Stefan Achleitner. Poem: Power-efficient occupancy-based energy management system. In *Proceedings of IPSN '13*, IPSN '13, 2013.