SAVES: A Sustainable Multiagent Application to Conserve Building Energy Considering Occupants

Jun-young Kwak, Pradeep Varakantham*, Rajiv Maheswaran, Milind Tambe, Farrokh Jazizadeh, Geoffrey Kavulya, Laura Klein, Burcin Becerik-Gerber, Timothy Hayes, Wendy Wood
University of Southern California, Los Angeles, CA, 90089
*Singapore Management University, Singapore, 178902
{junyounk,maheswar,tambe,jazizade,kavulya,lauraakl,becerik,hayest,wendy.wood}@usc.edu,
*pradeepv@smu.edu.sg

ABSTRACT

This paper describes an innovative multiagent system called SAVES with the goal of conserving energy in commercial buildings. We specifically focus on an application to be deployed in an existing university building that provides several key novelties: (i) jointly performed with the university facility management team, SAVES is based on actual occupant preferences and schedules, actual energy consumption and loss data, real sensors and hand-held devices, etc.; (ii) it addresses novel scenarios that require negotiations with groups of building occupants to conserve energy; (iii) it focuses on a non-residential building, where human occupants do not have a direct financial incentive in saving energy and thus requires a different mechanism to effectively motivate occupants; and (iv) SAVES uses a novel algorithm for generating optimal MDP policies that explicitly consider multiple criteria optimization (energy and personal comfort) as well as uncertainty over occupant preferences when negotiating energy reduction – this combination of challenges has not been considered in previous MDP algorithms. In a validated simulation testbed, we show that SAVES substantially reduces the overall energy consumption compared to the existing control method while achieving comparable average satisfaction levels for occupants. As a real-world test, we provide results of a trial study where SAVES is shown to lead occupants to conserve energy in real buildings.

Categories and Subject Descriptors

I.2.11 [**ARTIFICIAL INTELLIGENCE**]: Distributed Artificial Intelligence

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Innovative Applications, Energy, Sustainable Multiagent Building Application, Multi-objective Optimization

1. INTRODUCTION

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems – Innovative Applications Track (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

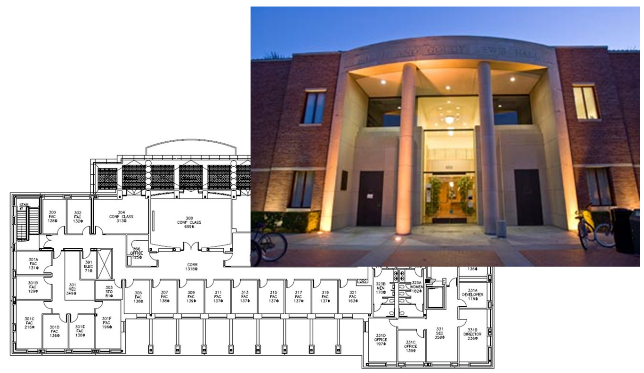


Figure 1: The actual research testbed (RGL) at the University of Southern California

Limited availability of energy sources has led to the need to develop efficient measures of conserving energy. Motivated by this need, researchers at AAMAS have been developing multiagent systems to conserve energy, both for deployment in smart grids and in buildings, with a particular focus on residential buildings [10, 16, 17, 21].

Inspired by this prior work, we describe an innovative multiagent system called SAVES (Sustainable multi-Agent building application for optimizing Various objectives including Energy and Satisfaction), where agents communicate and negotiate with human occupants to conserve energy. SAVES focuses on energy conservation in commercial (including office and educational) buildings given their significant burden on energy consumption, e.g., in 2008 buildings in the U.S. consumed 18.5 QBtu, representing 46.2% of building energy consumption and 18.4% of U.S. energy consumption [1]. To this end, this paper specifically focuses on an application to be deployed at Ralph & Goldy Lewis Hall (RGL) at the University of Southern California (shown in Figure 1).

SAVES provides the following key novelties. First, jointly performed with the university facility management team, our research is based on actual occupant preferences and schedules, actual energy consumption and loss data, real sensors and hand-held devices, etc. Second, SAVES addresses novel scenarios that require agents to negotiate with groups of building occupants to conserve energy; previous work has typically focused on agents' negotiation with individual occupants [3, 12]. Third, it focuses on non-residential buildings, where human occupants do not have a direct financial incentive in saving energy. Furthermore, commercial buildings offer new opportunities for energy conservation, since occupants may follow a more regular schedule, allowing SAVES to plan ahead for energy conservation. Finally, SAVES uses a novel algorithm for generating optimal *BM-MDP* policies that explicitly considers multiple criteria optimization (energy and personal comfort) as well as uncertainty over occupant preferences when negotiating for energy reduction – this combination of challenges has not been considered in previous MDP algorithms [5, 6, 8, 13].

We provide three sets of evaluations of SAVES. First, we constructed a detailed simulation testbed, with details all the way down to individual electrical outlets in our targeted building and variations in solar gain per day; and then validated this simulation. Within this simulation testbed, we show that SAVES substantially reduces the overall energy consumption compared to existing control methods while achieving comparable satisfaction level of occupants. Second, we show the benefits of BM-MDPs by showing that it gives a well-balanced solution while considering multiple criteria. Third, as a real-world test, we provide results of a human subject study where SAVES is shown to lead human occupants to significantly reduce their energy consumption in real buildings.

In Section 2, we describe our testbeds. This includes both the real educational building where SAVES is to be deployed, and our simulation testbed, which we validate by comparing with real building data. Next, in Section 3, we describe the SAVES multiagent system, and the novel BM-MDP algorithm at the heart of SAVES. Section 4 provides evaluations discussed above.

2. TESTBEDS

2.1 Educational Building Testbed

SAVES is to be deployed in an actual educational building. Figure 1 shows the real testbed building (RGL) and the floor plan of 3^{rd} floor. It is a multi-functional building that has been designed with a building management system, and it provides a good environment to test various control strategies to mitigate energy consumption. In particular, this campus building has three floors in total and is composed of different types of spaces including classrooms, offices for faculty and staff, and conference rooms for meetings. Each floor has a large number of rooms and zones (a set of rooms that is controlled by specific piece of equipment) with various physical properties including different building devices, orientation, window size, room size and lighting specifications. For instance, the 3^{rd} floor has 24 zones and 39 rooms.

Within this building, components and equipment include HVAC (Heating, Ventilating, and Air Conditioning) systems, lighting systems, office electronic devices such as computers and AV equipment, and different types of sensors and energy meters. Human occupants of the building are divided into two main categories: permanent and temporary. Permanent occupants include office users such as faculty, staff, researchers and laboratory residents. Temporary occupants include scheduled occupants like students or faculty attending classes or meetings and unscheduled occupants who are students or faculty using common lounges or dining spaces.

In this domain, there are two types of energy-related occupant behaviors that SAVES can influence to conserve energy use: individual behaviors and group behaviors. Individual behaviors only affect an environment where the individual is located. They include adjusting light sources and temperature in individual offices and turning on/off computers and other electronics. Group behaviors lead to changes in shared spaces and require negotiation with a group of occupants in the building. For instance, SAVES may negotiate with a group of occupants to adjust the lighting level and temperature in their shared office or to relocate a meeting to a smaller office. As we will show later, energy savings by considering such group negotiations together are significant.

The desired goal in this educational building is to optimize multiple criteria, i.e., achieve maximum energy savings without trading off the comfort level of occupants. The research on this testbed building is intended to be generalized to other building types, where we can observe many different types of energy-use and the behavioral patterns of occupants in the buildings.

2.2 Simulation Testbed

As an important first step in deploying SAVES in the actual building described in the previous section, we test SAVES in a realistic simulation environment using real building data. To that end, we have constructed a simulation testbed based on the open-source project OpenSteer (http://opensteer.sourceforge.net/), which provides a 2D, OpenGL environment. It can be used for efficient statistical analysis of different control strategies in buildings before deploying the system.

Building Components: Our simulation considers three building component categories: HVAC devices, lighting devices, and appliances. The HVAC components control the temperature of the assigned zone. The lighting devices control the lighting level of the room. The appliances in our simulation are either desktop or laptop computers. These components have two possible actions: "on" and "standby". When the lighting or appliance devices are on, they consume a fixed amount of energy. We attempt to very accurately reflect the energy consumed by each of the three component categories in the simulation. The energy consumption of HVACs is calculated based on changes in air temperature and airflow speeds, and gains from natural light source and appliances in the space. To calculate the energy consumption of the lighting and appliance devices, we collected actual energy consumption data in the testbed building. For the appliances, a desktop computer spends 0.150 *kW/h* and 0.010 *kW/h* when it is on and standby, respectively. A laptop computer spends 0.050 *kW/h* when it is on and 0.005 *kW/h* when it is on standby.[1]

Human Occupants: We built two types of human occupants in our simulation using the agent behavior framework presented in [20]. Permanent occupants stay in their offices or follow their regular schedules. Temporary occupants stay in the building for classes and leave once classes end. Each occupant has access to a subset of the six available behaviors according to her/his type — wander, attend class, go to meeting, teach, study, and perform research — any one of which may be active at a given time, where the behavior is selected based on class and meeting schedules. Occupants also have a satisfaction level based on the current environment, modeled as a percentage between 0 and 100 (0 is fully dissatisfied, 100 is fully satisfied).

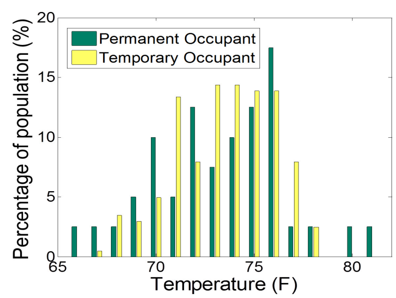


Figure 2: Actual Temp. Preference

To model the satisfaction level in this simulation, we use a Gaus-

[1]The detailed equations to compute the energy consumption and actual parameter values are presented here: http://teamcore.usc.edu/junyounk/energy/AAMAS12-SAVES-supplementary.pdf

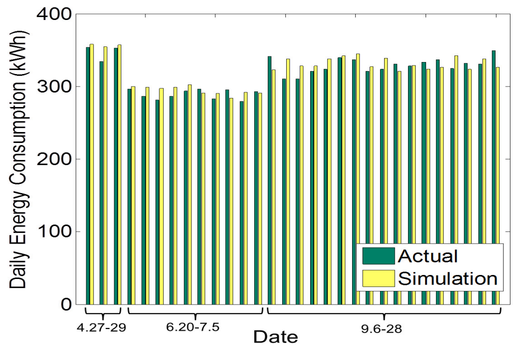


Figure 3: Energy Consumption Validation

sian distribution $N(\mu, \sigma)$ for each occupant. The mean (μ) of each individual Gaussian is drawn from actual occupant preference data shown in Figure 2 (e.g., for 18% of permanent occupants, μ=76°F). This data was gathered from 40 permanent occupants and 202 temporary occupants in RGL over two weeks in the spring of 2011. We use this actual data instead of the ASHRAE standard, which fails to account for individual preferences. The standard deviation (σ) of each Gaussian is selected uniformly randomly from a range of 3–5°F [11]. Based on the constructed Gaussian model for each occupant, the satisfaction level is computed as follows:

$$S(t) = \begin{cases} 100, & \text{if } t = \mu \\ \frac{f(t)}{f(\mu)} \times 100, & \text{if } t \neq \mu \end{cases} \quad (1)$$

where $S(t)$ is the satisfaction function, $f(x)$ is the probability density function of $N(\mu, \sigma)$, and t is the current temperature.

Validation: Before testing SAVES in simulation, we first validate the simulation testbed. Specifically, we compare the energy consumption calculated in the simulation testbed with actual energy meter data using the 3^{rd} floor of the actual testbed building.

Figure 3 shows that daily energy use comparison data (y-axis) measured for 30 sample weekdays throughout different seasons (x-axis; 3 weekdays in 2011 Spring, 10 weekdays in 2011 Summer, 17 weekdays in 2011 Fall). The energy consumption includes the amount consumed by HVACs, lighting devices and appliances. We use measured parameter values such as solar gain and outdoor temperature and real parameter values for the building obtained from the facility management system. We set the starting indoor temperature using real data. The likelihood value for human occupants to "turn off" lights and appliances when they leave their offices is 76%, based on a survey of the testbed building. Students follow 2010 Fall, 2011 Spring and 2011 Fall class schedules, and faculty, staff and students follow their meeting schedules.

As shown in the figure, the difference between actual energy meter data and energy use from the simulation testbed was between 0.17% – 8.71% (mean difference: 3.37%), which strongly supports our claim that the simulation testbed is realistic.

3. SAVES

In this section, we describe the key components of SAVES and how to optimally plan negotiations with groups of occupants to conserve energy in our application.

3.1 Agents in SAVES

At the heart of SAVES are two types of agents: room agents and proxy agents (Figure 4). There is a dedicated room agent per office and conference room, in charge of reducing energy consumption in that room. It can access sensors to retrieve information such as the

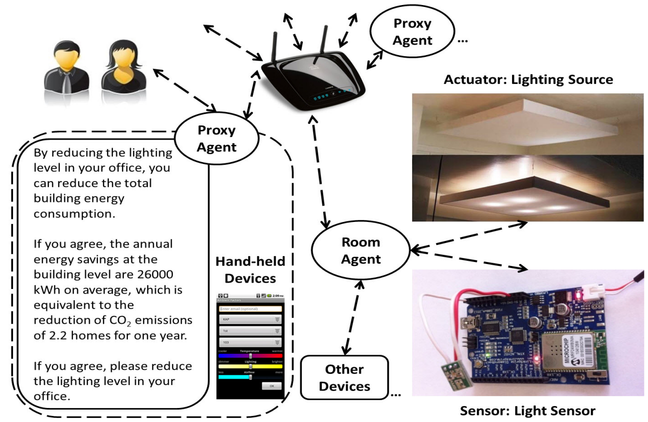


Figure 4: Agents & Communication Equipment in SAVES. An agent in SAVES sends feedback including energy use to occupants.

current lighting level and temperature and energy use at different levels (building-level, floor-level, zone-level, and room-level) and impact the operation of actuators. A proxy agent [18] is on an individual occupant's hand-held device and it has the corresponding occupant's preference and behavior models. Proxy agents communicate on behalf of an occupant to the room agent. Such proxy agents' adjustable autonomy – when to interrupt a user and when to act autonomously – is recognized as a major research issue [18, 19], but since it is not our focus, we use preset rules instead. Room agents may directly communicate with occupants without proxy agents if needed. Finally, different room agents coordinate among themselves via proxy agents, e.g., if two separate conference room agents wish to move a meeting to one occupant's office, the proxy of that occupant allows one of the room agents to proceed, blocking the other's request (see Figure 4).

Room agent reasoning is based on a new model called *Bounded parameter Multi-objective MDPs* (BM-MDPs), which is one of the contributions of this research. BM-MDPs are a hybrid of MO-MDPs [5, 13] and BMDPs [8]. BM-MDPs are responsible for planning simple and complex tasks. Simple tasks include turning on the HVAC before a class or a meeting, and do not need the full power of the BM-MDPs. Complex tasks were why BM-MDPs were created; these include negotiating with groups of individuals to relocate meetings to smaller rooms to save energy, negotiating with multiple occupants of a shared office to reduce energy usage in the form of lights or HVACs, and others. Before describing BM-MDPs in depth, we motivate their use by elaborating on the meeting relocation negotiation scenario.

Group Meeting Relocation Negotiation Example Consider a meeting that has been scheduled with two attendees (P_1 and P_2) in a large conference room that has more light sources and appliances than smaller offices. Since the meeting has few attendees, the room agent can negotiate with attendees to relocate the meeting to nearby small, sunlit offices, which can lead to significant energy savings. The room agent handles this negotiation based on BM-MDPs. There are three objectives (i.e., three separate reward functions) that the room agent needs to consider during this negotiation: i) energy saving (R_1), ii) P_1's comfort level change (R_2), and iii) P_2's comfort level change (R_3). The room agent first checks the available offices. Assuming there are two available offices A and B, the room agent asks each attendee if she or he will agree to relocate the meeting to one of the available offices. In asking an attendee, the room agent must consider the uncertainty of whether an attendee is likely to accept its offer to relocate the meeting. Since asking incurs a cost (e.g., cost caused by interrupting people), the room agent needs to reason about which option is preferable considering P_1 and P_2's likelihood to accept each option (A or B) and the reward functions for each option to reduce the required cost and maximize benefits. Assuming A is preferable, the optimal policy of the agent is "ask P_1 first about A"–"if P_1 accepts, ask P_2 about A"–"if P_1 does not reply, ask P_1 about A again"–"repeat the process with B"–"if both agree, relocate the meeting"–"if both disagree, find other available options." While this is a simplified example, in practice the problem is more difficult, as there may be more than two attendees in a meeting. The room agent must also first communicate with the proxies of the owners of offices A and B and there may be uncertainty in their agreement to have a meeting in their office; further adding to the challenge of sequential decision making under uncertainty. In addition, the agent must decide if it should ask P_1 first and use that result to influence P_2, etc.

Thus, BM-MDPs must reason with multiple objectives, but simultaneously must reason with the uncertainty in the domain. In fact, in a complex domain such as ours, the probabilities of attendees' or others' acceptance of the room agent's offer, or the probabilities of other outcomes may not be precisely known — we may only have a reasonable upper and lower bound over such probabilities. Indeed, precisely knowing the model is very challenging, and we ended up building BM-MDPs to address both these challenges and requirements. However, before explaining BM-MDPs, we first explain MO-MDPs on which BM-MDPs are built.

3.2 Multi-objective MDPs

The negotiation scenarios described earlier require SAVES to consider multiple objectives simultaneously: energy consumption and satisfaction level of multiple individuals. To handle such multiple objectives, MDPs have been extended to take into account multiple criteria assuming no *model uncertainty*. *Multi-Objective MDPs* (MO-MDPs) [5, 13] are defined as an MDP where the reward function has been replaced by a vector of rewards. Specifically, MO-MDPs are described by a tuple $\langle S, A, T, \{R_i\}, p\rangle$, where R_i is the reward function for objective i and p denotes the starting state distribution ($p(s) \geq 0$). In the *meeting relocation example* shown in Section 3.1, specifically, the multiple reward functions, $\{R_i\}$, include energy consumption (which is the reduction in energy usage in moving from a conference room to a smaller office), and comfort level defined separately for each individual (based on data related to their temperature comfort zones).

The key takeaway from MO-MDPs towards BM-MDPs is an understanding of how to generate a policy in the presence of such multiple objectives that are not aggregated into one single value. The key principle we rely on, given the current domain of non-residential buildings is one of fairness; we wish to reduce energy usage, but we cannot sacrifice any one individual's comfort entirely in service of this goal. To meet this requirement, we focus on minimizing the maximum *regret* instead of maximizing the reward value based on a min-max optimization technique [14] to get a well-balanced solution.

To minimize the maximum regret, we first need to compute the optimal value for each objective using the MDP framework relying on the following standard formulation:

$$\min V^*(s) \tag{2}$$

$$\text{s.t.} \quad V^*(s) \geq R(s,a) + \gamma \sum_{s' \in S} T(s,a,s') \cdot V^*(s'), \tag{3}$$

$$0 \leq \gamma < 1 \tag{4}$$

where V^* is an optimal value, and γ is a discount factor.

We define the regret in MO-MDPs as following:

Definition Let $H_i^\alpha(s)$ be the *regret* with respect to a policy α for objective i and state s. Formally,

$$H_i^\alpha(s) = V_i^{\alpha_i^*}(s) - V_i^\alpha(s), \tag{5}$$

where $V_i^{\alpha_i^*}(s)$ is the value of the *optimal* policy, α_i^*, and $V_i^\alpha(s)$ is the value of the policy α for objective i and state s.

Therefore, we can minimize the maximum regret in MO-MDPs using the following optimization problem:

$$\min D \tag{6}$$

$$\text{s.t.} \quad D \geq \sum_{s \in S} p(s) \cdot \left[V_i^*(s) - V_i(s)\right], \forall i \in I, \tag{7}$$

$$V_i(s) = \sum_{a \in A} \alpha(s,a) \left[R_i(s,a) + \gamma \sum_{s' \in S} T(s,a,s') \cdot V_i(s')\right], \tag{8}$$

$$\sum_{a \in A} \alpha(s,a) = 1, \forall s \in S, \qquad 0 \leq \gamma < 1 \tag{9}$$

where V_i^* is the constant value pre-calculated by (2) of the MDP formulation using the reward function for objective i, R_i, and I is a set of objectives.

Unfortunately, in BM-MDPs, we have an upper and lower bound on transition probabilities and rewards, and thus this optimization problem cannot be directly used. Nonetheless, it helps us understand the key difference in minimizing max regret between MO-MDPs and BM-MDPs — specifically in addressing such upper and lower bounds in BM-MDPs, we end up with different transition probabilities T_i for each objective i, as discussed below, and hence rely on a different approach to compute regret.

3.3 BM-MDPs

We now extend MO-MDPs, using ideas from BMDPs [8], to create BM-MDPs. BMDP (represented by tuple $\langle S, A, \hat{T}, \hat{R}, p\rangle$) is an extension to the standard MDP, where upper and lower bounds on transition probabilities and rewards are provided as closed real intervals. In addition to representation of uncertainty over transition probabilities and rewards, a key takeaway for BM-MDPs from BMDPs is the algorithm to generate policies. This algorithm is based on the notion of *Order-Maximizing* MDPs [8], which selects transition probabilities from the given intervals. Order-maximizing MDPs crucially take the order of states as an input – this order is ascending if we are to return a pessimistic policy (based on lower bound values), and it is descending for an optimistic policy (based on upper bound values). More specifically, using this order as an input, order-maximizing MDPs construct the transition function, and generate a policy as an output relying on value iteration. We rely on order-maximizing MDPs to generate policies in our BM-MDPs as well (but manipulate the order of states input). To provide some intuition behind the operations of the order-maximizing MDPs, we provide a simple example to show how transition values are assigned from their intervals using the given order in the following example. For more details, please refer to [8].

Example of Order Maximizing MDPs Consider a BMDP with two states: s_1 and s_2. The transition ranges are $T(s_1, a, s_1)$ = [0.5, 0.9], $T(s_1, a, s_2)$ = [0.2, 0.6]. Let us assume that the upper bound of value is $V_{ub}(s_1)$ = 3 and $V_{ub}(s_2)$ = 2 at a certain iteration of order-maximizing MDP value iteration. In BMDP, the intuition is that for calculating the optimistic value, we require movement to s_1 as much as possible within the given range of transition probability (since it has a higher upper bound value). Therefore to create an optimistic policy, the input to the order-maximizing MDPs is to sort the states in a descending order based on the upper bounds. Given this input, the transition probabilities in the order-maximizing MDP for calculating optimistic value would be $T'(s_1, a, s_1)$ = 0.8 because $T'(s_1, a, s_2)$ should be at least 0.2, and $T'(s_1, a, s_2)$ = (1 - 0.8). Based on these transition probabilities, we obtain a new set of expected values via value iteration, generate a new descending order, and iterate until convergence.

Similar to BMDPs, the transition and reward functions in BM-MDPs have closed real intervals. Whereas BMDPs are limited to optimizing a single objective case (i.e., the BMDP model requires one unified reward function), BM-MDPs can i) optimize over multiple objectives (i.e., a vector of reward functions) with ii) different degrees of model uncertainty. Specifically, BM-MDPs are described by a tuple $\langle S, A, \hat{T}, \{\hat{R}_i\}, p\rangle$, where $\hat{R}_i$ represents the reward function for objective i.

Algorithm 1 SOLVEBMMDP()

1: **for** $i = 1 \in I$ **do**
2: $\quad \left\langle \mathbf{V}^*_{i,lb}, \mathbf{V}^*_{i,ub} \right\rangle \leftarrow$ SolveBMDP($BMDP_i$)
3: $\{\mathbf{V}'_{i,lb}\} \leftarrow \infty$; $\{\mathbf{V}_{i,lb}\} \leftarrow \mathbf{0}$
4: **while** $|\{\mathbf{V}'_{i,lb}\} - \{\mathbf{V}_{i,lb}\}| > \epsilon$ **do**
5: $\quad \{\mathbf{V}_{i,lb}\} \leftarrow \{\mathbf{V}'_{i,lb}\}$
6: $\quad$ **for** $i = 1 \in I$ **do**
7: $\quad\quad \mathbf{O}_i \leftarrow$ SortIncreasingOrder($\{V_{i,lb}\}$)
8: $\quad\quad \mathbf{M}_i \leftarrow$ ConstructOrderMaximizingMDP($\mathbf{O}_i$);
9: $\quad \{\mathbf{V}'_{i,lb}\} \leftarrow$ SolveMOMDPPessimistic($\{\mathbf{V}_{i,lb}\}, \{\mathbf{V}^*_{i,lb}\}, \{\mathbf{M}_i\}$)
10: $\alpha_{pes} \leftarrow$ ObtainPessimisticPolicy($\{\mathbf{V}_{i,lb}\}$)
11: $\{\mathbf{V}'_{i,ub}\} \leftarrow \infty$; $\{\mathbf{V}_{i,ub}\} \leftarrow \mathbf{0}$
12: **while** $|\{\mathbf{V}'_{i,ub}\} - \{\mathbf{V}_{i,ub}\}| > \epsilon$ **do**
13: $\quad \{\mathbf{V}_{i,ub}\} \leftarrow \{\mathbf{V}'_{i,ub}\}$
14: $\quad$ **for** $i = 1 \in I$ **do**
15: $\quad\quad \mathbf{O}_i \leftarrow$ SortDecreasingOrder($\{V_{i,ub}\}$)
16: $\quad\quad \mathbf{M}_i \leftarrow$ ConstructOrderMaximizingMDP($\mathbf{O}_i$);
17: $\quad \{\mathbf{V}'_{i,ub}\} \leftarrow$ SolveMOMDPOptimistic($\{\mathbf{V}_{i,ub}\}, \{\mathbf{V}^*_{i,ub}\}, \{\mathbf{M}_i\}$)
18: $\alpha_{opt} \leftarrow$ ObtainOptimisticPolicy($\{\mathbf{V}_{i,ub}\}$)
19: **return** $\{\langle \alpha_{pes}, \alpha_{opt} \rangle\}$

To solve BM-MDPs, we introduce a novel algorithm that is a hybrid of BMDPs and MO-MDPs. Specifically, our algorithm marries the minimization of max regret idea from MO-MDPs with that of order maximizing MDPs to handle uncertainty over transition function and rewards. The overall flow is described in Algorithm 1. At a higher level, we have three stages: (i) computing the optimal value bounds $\left\langle \mathbf{V}^*_{i,lb}, \mathbf{V}^*_{i,ub} \right\rangle$ for each objective i using BMDPs (lines 1–2), (ii) using the MO-MDP idea to optimize multiple objectives based on a min-max formulation (lines 3–9 & 11–17), and (iii) obtaining a policy α based on the final value functions $\langle\{\mathbf{V}_{i,lb}\}, \{\mathbf{V}_{i,ub}\}\rangle$ (lines 10 & 18). The output of this algorithm is in the form of two policies (pessimistic and optimistic), and we leave it to the user to determine which one is used.

We now describe the computation of the pessimistic policy (lines 3–10). The optimistic policy (lines 11–18) is similarly computed. The pessimistic policy minimizes the maximum regret with respect to the optimal lower bound values of all objectives ($\{V^*_{i,lb}\}$) over all states; this computation is iteratively performed in line 9. For each objective i, we first get an ascending order of states using the current lower bound values $V_{i,lb}$ (line 7) to construct the order-maximizing MDP (line 8). This set of order-maximizing MDPs, $\{\mathbf{M}_i\}$, is an input to the function SolveMOMDPPessimistic() to optimize multiple objectives by directly computing regret on line 9. This computation is performed by Eq. (6) with a different transition probability function T_i in the given $\mathbf{M}_i$ instead of T. This in turn influences the sorting order of states, and the process continues until the expected values $\{V_{i,lb}\}$ converge.

4. EVALUATION OF SAVES

In this section, we provide three sets of evaluations: two sets of results tested in the simulation testbed and a set of results tested in the real-world.

4.1 Simulation: Overall Evaluation

We evaluate the performance of SAVES using both 2^{nd} and 3^{rd} floors of RGL in the simulation environment. We test BM-MDPs using a pessimistic setting and compare it with two other control heuristics discussed below.

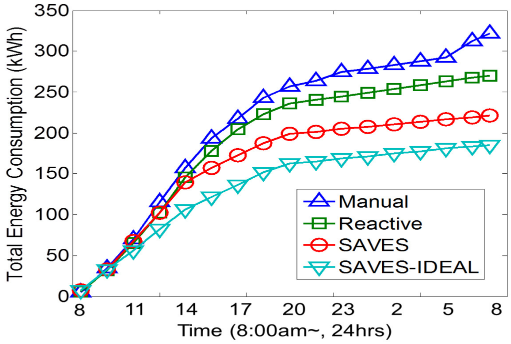
Manual Control: The manual control strategy is the baseline system that represents the current strategy operated by the facility management team in the real testbed building (RGL). In this strategy, temperature is regulated by the facility managers according to two set ranges for occupied (70°F–75°F) and unoccupied periods (50°F–90°F) of the day. In this control setting, HVACs always attempt to reach the pre-set temperature regardless of the presence of occupants and their preferences in terms of temperature. Lighting and appliance devices are controlled by human occupants. The same likelihood value for human occupants to "turn off" lights and appliances was used as in Section 2.2.

Reactive Control: We consider the reactive control heuristic for comparison purposes since it can be easily implemented using cheap sensors in the real building, and recently, some buildings have already started adopting this simple heuristic to reduce energy use. The lighting and appliance devices are now automatically controlled and turned on and off according to the presence of people. Additionally, as in [9], appropriate temperature set points of HVACs are computed based on the average preference of human occupants. HVACs automatically turn on and off according to the presence of people and temperature set points.

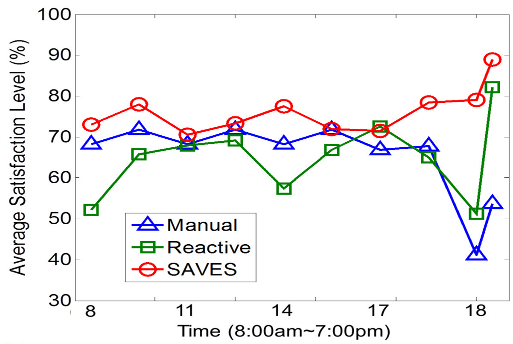
We focus on measuring two different criteria — total energy consumption (*kWh*) and average satisfaction level of occupants (%). The experiments were run on Intel Core2 Duo 2.53GHz CPU with 4GB main memory. All techniques were evaluated for 100 independent trials throughout this section. We report the average values.

4.1.1 Result: Total Energy Consumption

We compared the cumulative total energy consumption measured during 24 hours for all control strategies. Figure 5(a) shows the cumulative total energy consumption on the y-axis in *kWh* and time on the x-axis. We report the average total energy consumption measured over the same 30 weekdays used in Figure 3. As shown in the figure, the manual control strategy showed the worst result since it does not take into account behaviors or schedules of human occupants and building components simply follow the predefined policies. The reactive control strategies showed lower energy consumption than the manual setting by 16.06%. SAVES showed the best results compared to other control heuristics and statistically significant improvements (t-test; $p < 0.01$) in terms of energy used in the testbed building. Specifically, our algorithm with the ideal compliance rate (i.e., SAVES-IDEAL: occupants *always* accept the suggestions provided by the SAVES room agents to conserve energy) reduced the energy consumption by 42.45% when compared to the manual control strategy. If we use the compliance rate (68.18%) of human subjects shown in Table 3 (as measured



(a) Total Energy Consumption



(b) Average Satisfaction Level

Figure 5: Performance Evaluation of SAVES

in our real-world experiments), SAVES achieved energy savings by 31.27% (40% of the savings due to SAVES came out of group tasks, such as reducing energy consumption in shared offices, relocating meetings, and others) as compared to the manual setup. This is double the rate of the reactive approach.

4.1.2 Result: Average Satisfaction Level

Here, we compare the average satisfaction level of human occupants under different control strategies in the simulation testbed. Figure 5(b) shows the average satisfaction level in percentage on the y-axis and time on the x-axis. As shown in the results, the manual setting and our novel algorithm showed the best results. This is because the manual setting makes HVACs attempt to reach the desired temperature set point as soon as possible while disregarding the resulting energy consumption, and our method plans ahead of the schedules; thus, these two can achieve the desired comfort level faster than the reactive control strategy.

The manual strategy, however, is very sensitive to the given temperature range. In our experiment, the temperature set point was set by the facility management team (e.g., 70–75°F) based on the average preference model, thus it achieved high comfort level in the testbed. However, if the actual preferred temperature in the building is different from the average model, it fails to meet the occupant's desired level. This phenomenon can be seen when occupants stay during the unoccupied time (after typical working hours). As we can see at 18 on the x-axis (i.e., 6pm) in the figure, the average comfort level drops significantly. Due to the delayed effects in temperature change, the reactive control strategy showed significantly lower satisfaction results than other methods. For instance, it has a satisfaction level below 60% at 14 on the x-axis (i.e., 2pm). Thus, SAVES not only provides superior energy savings, but also avoids the reduction in comfort level that a reactive strategy may cause.

Table 1: Average Maximum Regret Comparison

Problem Set	MDPs	BM-MDPs	Difference
m_1	168.62	4.72	163.90
m_2	359.44	164.17	195.27
m_3	448.15	164.97	283.18
m_4	291.27	138.59	152.68
m_5	143.32	95.88	47.44

Table 2: Example of the *Meeting Relocation Negotiation*

	Max. Regret	
Objective	MDPs	BM-MDPs
Energy Savings	443.54	162.83
P_1's Comfort Lv. Change	15.34	162.84
P_2's Comfort Lv. Change	15.34	97.58

4.2 Simulation: Multi-objective Optimization

In this section, we perform more analysis on our novel algorithm. Table 1 shows the average maximum regret comparison tested in 5 different problem sets between the standard MDP with a unified reward based on the weighted sum method [22] and BM-MDPs (in this case, we assume no transition or reward uncertainty). The uniform weight distribution was applied to the weighted sum method. Our goal is to show that BM-MDPs give lower maximum regrets, which indicates well-balanced solutions as discussed earlier.

Each problem is an instance of the *meeting relocation negotiation* task, having its own reward structure but the same transition function. The problem instances are divided into five groups (problem sets m_1–m_5) based on the percentage of objectives that have positive rewards in all objectives. Recall that in the meeting relocation scenario, the different objectives include energy reduction and change in comfort level of individual participants. Specifically, in problem set m_1, relocating a meeting leads to positive rewards in over 75% of objectives (76–100%) and negative rewards in the rest of objectives, problem set m_2 has 51–75% of objectives with positive rewards, and similarly for the remaining sets, so that in problem set m_5, all objectives have negative rewards if the meeting is relocated. Each problem set has 100 independent problem instances. We then measured the average maximum regret of each method in each problem set. As shown in Table 1, BM-MDPs always showed lower maximum regrets (column 3) compared to the MDP with uniform weight (column 2), which suggests that our method gives well-balanced solutions regardless of reward characteristics.

The next question is what the well-balanced solution means in our energy domain. Let us take the *meeting relocation example* with two attendees (P_1 and P_2) discussed in Section 3.1. In Table 2, column 1 shows three objectives (energy savings and two attendees' comfort level change) and columns 2–3 indicate the maximum regret from MDPs and BM-MDPs, respectively. As shown in the table, MDPs generated a policy that almost entirely disregards energy-savings, leading to significantly large regrets (row 3, column 2). BM-MDPs, on the other hand, were able to achieve small regrets over all objectives (rows 3–5, column 3).

Lastly, we test our BM-MDP algorithm considering different degrees of model uncertainty. Figure 6 shows the average maximum regret tested over 100 different problem instances on the y-axis. We choose 1 problem from each problem set (m_1, m_2, $\cdots$, m_5) from the previous test. The noise of each model is proportional (20%) to the mean reward value and transition probability. MDPs and MO-MDPs generate policies ignoring the model's uncertainty and BM-MDPs generate two types of policies (BM-MDP-Pes: pessimistic, BM-MDP-Opt: optimistic) that explicitly account for the uncertainty. We then randomly generate 20 different instances within the range for each problem (e.g., for m_1, we generate $m_{1,1}, \cdots, m_{1,20}$). Each generated policy is evaluated over those 20 problem instances and the average maximum regret is computed for each algorithm. For the other 4 problems ($m_2, \cdots, m_5$), we repeat the same procedure and report the overall average value. As shown in the figure, BM-MDPs have the best performance (i.e., lowest average maximum regret), which means BM-MDPs are capable of generating more robust and well-balanced solutions compared to previous work when there is model uncertainty. However, the solution quality between the pessimistic and optimistic BM-MDPs was not significantly different and their performance is domain dependent. *Note that the results shown in Figure 6 are average maximum regrets over all problem instances, and in some particular instances, MO-MDP might outperform either BM-MDP-Pes or BM-MDP-Opt (but not both even in this case).* We leave this issue for future investigation.

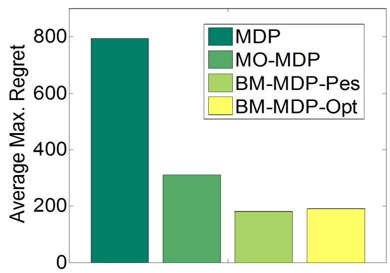


Figure 6: Performance of BM-MDPs

4.3 Real-world Test: Human Experiments

As a real-world test, we design and conduct a validation experiment on a pilot sample of participants (staff on campus). We conduct this investigation: i) to verify if SAVES can lead to changes in occupants' behaviors and to reduce energy consumption in commercial buildings, ii) to validate the parameter values used during the negotiation process such as the acceptance/compliance rate for the suggestion and iii) to understand what types of feedback are most effective to affect occupants' energy-related decisions.

In this study, we consider two test conditions: i) feedback without motivation (Test Group I) (e.g., please reduce the lighting level in your office), and ii) feedback with motivation including participant's own energy use, and environmental motives (Test Group II) (e.g., if you reduce your lighting for working hours, the annual energy savings at the building level are $26000kWh$ on average, which is equivalent to the reduction of CO_2 emissions of 2.2 homes for one year). From this experiment, we answer the following question by comparing change in energy behavior patterns and possible estimated energy consumption between test groups I and II.

HYPOTHESIS 1. *More informed feedback (provided to subjects in Test Group II) will be more effective to conserve energy than feedback without motivation (Test Group I).*

We tested the hypothesis above as follows: we first recruited 22 staff from 7 buildings at the University of Southern California who are over 18 years old. Subjects were tested under two different conditions, and each test group had 11 individuals respectively, each of whom has her/his own office. Since we tested using a simple lighting negotiation scenario, each participant must be able to adjust the lighting levels in her/his office. With participants' agreements, we installed lighting sensors (Figure 4) in their offices. During the experiment, participants were supposed to stay in their own offices and do their regular work. We then measured the baseline energy behavior and energy consumption, and SAVES provided feedback via emails based on sensed lighting level (two times per day, at 11am and 2pm, for three consecutive weekdays). In each message, participants received a simple suggestion for lighting level with a certain type of feedback (e.g., please reduce the lighting level in your office). We systematically observed and logged their energy behavior during the entire experiment using the light sensors. At the end of the experiment, each participant was required to take a short survey (i.e., the reasons why they agree or disagree with a provided suggestion). We conducted this study for two weeks in the fall of 2011 and collected data from human subjects using multiple sensors and routers.

Table 3: Lighting Negotiation Results (*: p < 0.05)

	Avg. Accep. Rate (%)	User Rating (Max: 5.0)
Group I	28.79 (11.03)	3.82 (0.26)
Group II	68.18 (9.65)	4.18 (0.18)
Mean Diff.	39.39*	0.36

In Table 3, column 2 displays the average acceptance rate in percentage (0–100%) of two test groups, and column 3 represents the average user rating of the provided feedback during the experiment. The range of ratings is between 0 and 5, and 0 indicates that the feedback was not helpful at all, and 5 means that the feedback was extremely helpful. In both columns, values in parentheses indicate the standard errors. The last row shows a mean difference between two groups for each value.

Table 3 shows that when we provided more informed feedback including environmental motives (Group II), occupants showed statistically significantly higher compliance acceptance rate (68.18%), which provides strong evidence for the above hypothesis (t-test; $p < 0.05$). In addition, human subjects in Group II felt that the provided feedback was more helpful during the negotiation process. However, the difference in user ratings between two groups was not significant, and thus we took a quick survey from participants at the end of the experiment to further analyze their decisions. In contrast with Group I, in Group II, the main reason why participants who agreed to reduce the lighting level in their offices (over 80% of conformers in Group II) was because the feedback significantly improved awareness of energy use. In addition, more than half of all participants strongly believed that this study will be very helpful by encouraging occupants to think about energy usage. This discrepancy in average user ratings and acceptance rates remains an issue for future work.

In this trial study, we have learned that although occupants in commercial buildings do not have a direct financial incentive in saving energy, proper motivations can achieve a higher compliance rate for the energy-related suggestion. This study specifically gives us the insights that there is a significant potential to conserve energy by investigating effective and tailored methods to improve occupants' motivation to conserve energy.

5. RELATED WORK

In discussing related work, a key point we wish to emphasize is the uniqueness of our work in combining research on multiagent systems, specifically our multi-objective MDP algorithm that handles uncertainty, and negotiations with human subjects, in an innovative application for energy savings. It is this specific combination of attributes that sets SAVES apart from previous research.

Multiagent Energy Systems: Multiagent systems have been considered to provide sustainable energy for smart grid management. Voice *et al.* [21] provided a game-theoretic framework for modeling storage devices in large-scale systems where each storage device is owned by a self-interested agent that aims to maximize its monetary profit. In addition, [10] addressed research challenges to integrate plug-in Electric Vehicles (EVs) into the smart grid.

To model and optimize building energy consumption, Ramchurn *et al.* [16] considered more complex deferrable loads and managing comfort in the residential buildings. Rogers *et al.* [17] addressed the challenge of adaptively controlling a home heating system in order to minimize cost and carbon emissions within a smart grid using Gaussian processes to predict the environmental parameters. Our domain is different in focusing on energy savings in commercial buildings, and the representation and approaches are also different from previous work by allowing consumers (i.e., occupants) to play a part in optimizing the operation in the building instead of managing the optimal demand on buildings.

Energy Literacy via Feedback: Abrahmase *et al.* [2] reviewed 38 interventions aimed to reduce household energy consumption, and they concluded that normative feedback about energy use is the most promising strategy for reducing and maintaining low consumption. However, it focused on residential environments, which is different from our work. In a recent study, Carrico and Riemer [4] provided monthly normative feedback via email to occupants of a commercial building about their own buildings' energy use in comparison with and other, similar buildings. Unfortunately, the study relied on self-reporting to assess the behaviors. Instead, our work relies on real sensors to observe their energy behavior in real-time. Faruqui *et al.* [7] reviewed past experiments and pilot projects to evaluate the effect of in-home displays (IHDs) on energy consumption. Our work is different because we simultaneously consider multiple criteria including energy consumption *and* occupant comfort level.

Multi-objective Optimization Techniques: There has been a significant amount of work done on multi-objective optimization. The most common approaches to multi-objective optimization are to find Pareto optimal solutions [15], use the weighted sum method to aggregate multiple objectives using a prior preference [22], or consider the weighted min-max (or *Tchebycheff*) formulation that provides a nice theoretical property in terms of sufficient/necessary conditions for Pareto optimality [14].

Recently, Chatterjee *et al.* [5] considered MDPs with multiple discounted reward objectives. They theoretically analyzed the complexity of the proposed approach and showed that the Pareto curve can be approximated in polynomial time. Ogryczak *et al.* [13] focused on finding a compromise solution in multi-objective MDPs for a well-balanced solution. They compared their approach relying on the Tchebycheff scalarizing function to the weighted sum method. On the other hand, there has been some significant advances to handle model uncertainty on standard MDPs including [6, 8]. Our work is different from them as we assume model uncertainty while simultaneously optimizing multiple criteria in MDPs.

6. CONCLUSION

In this work, we presented an innovative multiagent system called SAVES with the goal of conserving energy in commercial buildings. There are several key novelties in SAVES: (i) SAVES is based on a real building and uses actual building data, including energy data, occupant preferences and schedules; (ii) it investigates both individual and group negotiations to save energy in smaller offices and shared rooms; (iii) it focuses on a commercial building, which requires a different mechanism to effectively motivate occupants since they do not have a direct financial incentive in conserving energy; and (iv) SAVES's reasoning is based on a novel BM-MDP algorithm for generating optimal policies that explicitly considers multiple criteria optimization as well as uncertainty over occupant preferences. We justified SAVES in a validated simulation testbed and showed that it can provide solutions to significantly reduce energy consumption while achieving comparable satisfaction levels of building occupants. As a real-world test, we provided results of a human subject study where SAVES is shown to lead occupants to conserve energy in real buildings.

7. ACKNOWLEDGMENTS

We thank Perceptronics Solutions, Inc. for their support of this research.

8. REFERENCES

[1] *Buildings Energy Data Book*. U.S. Dept. of Energy, 2010.

[2] W. Abrahmase, L. Steg, C. Vlek, and T. Rothengatter. A review of intervention studies aimed at household energy conservation. *J Environ. Psychol.*, 25:273–291, 2005.

[3] S. Abras, S. Ploix, S. Pesty, and M. Jacomino. A multi-agent home automation system for power management. *Informatics in Control Automation and Robotics*, 15:59–68, 2008.

[4] A. Carrico and M. Riemer. Motivating energy conservation in the workplace: An evaluation of the use of group-level feedback and peer education. *J Environ. Psychol.*, 31, 2011.

[5] K. Chatterjee, R. Majumdar, and T. A. Henzinger. Markov decision processes with multiple objectives. In *STACS*, 2006.

[6] K. V. Delgado, S. Sanner, L. N. de Barros, and F. G. Cozman. Efficient solutions to factored MDPs with imprecise transition probabilities. In *AAAI*, 2009.

[7] A. Faruqui, S. Sergici, and A. Sharif. The impact of informational feedback on energy consumption - a survey of the experimental evidence. *Energy*, 35, 2010.

[8] R. Givan, S. Leach, and T. Dean. Bounded-parameter Markov decision processes. *Artificial Intelligence*, 2000.

[9] F. Jazizadeh, G. Kavulya, L. Klein, and B. Becerik-Gerber. Continuous sensing of occupant perception of indoor ambient factors. In *ASCE International Workshop on Computing in Civil Engineering*, 2011.

[10] S. Kamboj, W. Kempton, and K. S. Decker. Deploying power grid-integrated electric vehicles as a multi-agent system. In *AAMAS*, 2011.

[11] H. E. Khalifa, C. Isik, and J. F. I. Dannenhoffer. Energy efficiency of distributed environmental control systems. Technical Report DOE-ER63694-1, Syracuse Univ., 2006.

[12] Z. Mo and A. Mahdavi. An agent-based simulation-assisted approach to bi-lateral building systems control. In *IBPSA*, 2003.

[13] W. Ogryczak, P. Perny, and P. Weng. A compromise programming approach to multiobjective Markov decision processes. In *MCDM*, 2011.

[14] A. Osyczka. An approach to multicriterion optimization problems for engineering design. *Comput. Methods Appl. Mech. Eng.*, 15:309–333, 1978.

[15] V. Pareto. *Manuale di Economica Politica*. Societa Editrice Libraria, 1906.

[16] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings. Agent-based control for decentralised demand side management in the smart grid. In *AAMAS*, 2011.

[17] A. Rogers, S. Maleki, S. Ghosh, and N. Jennings. Adaptive home heating control through Gaussian process prediction and mathematical programming. In *International Workshop on Agent Technology for Energy Systems (ATES)*, 2011.

[18] P. Scerri, D. V. Pynadath, and M. Tambe. Towards adjustable autonomy for the real world. *JAIR*, 17:171–228, 2002.

[19] N. Schurr, J. Marecki, and M. Tambe. Improving adjustable autonomy strategies for time-critical domains. In *AAMAS*, 2009.

[20] J. Tsai, N. Fridman, E. Bowring, M. Brown, S. Epstein, G. Kaminka, S. Marsella, A. Ogden, I. Rika, A. Sheel, M. Taylor, X. Wang, A. Zilka, and M. Tambe. ESCAPES - Evacuation Simulation with Children, Authorities, Parents, Emotions, and Social comparison. In *AAMAS*, 2011.

[21] T. Voice, P. Vytelingum, S. Ramchurn, A. Rogers, and N. Jennings. Decentralised control of micro-storage in the smart grid. In *AAAI*, 2011.

[22] K. Yoon and C.-L. Hwang. *Multiple Attribute Decision Making, An Introduction*. Sage Publications, 1995.