Smart Finite State Devices: A Modeling Framework for Demand Response Technologies

Konstantin Turitsyn, Scott Backhaus, Maxim Ananyev and Michael Chertkov

Abstract— **We introduce and analyze Markov Decision Process (MDP) machines to model individual devices which are expected to participate in future demand-response markets on distribution grids. We differentiate devices into the following four types: (a) optional loads that can be shed, e.g. light dimming; (b) deferrable loads that can be delayed, e.g. dishwashers; (c) controllable loads with inertia, e.g. thermostatically-controlled loads, whose task is to maintain an auxiliary characteristic (temperature) within pre-defined margins; and (d) storage devices that can alternate between charging and generating. Our analysis of the devices seeks to find their optimal price-taking control strategy under a given stochastic model of the distribution market.**

I. Introduction

Automated demand response is often used to manage electrical load during critical system peaks[1], [2]. During a typical event as the system approaches peak load, signaling from the utility results in automated customer load curtailment for a given period of time to avoid overstressing the grid. Although this type of load control is useful for maintaining system security, automated demand response must evolve further to meet the coming challenge of integrating time-intermittent renewables such as wind or photovoltaic generation. When these resources achieve high penetration and their temporal fluctuations exceed a level that can be economically mitigated by the remaining flexible traditional generation (e.g. combustion gas turbines), automated demand response will play a large role in maintaining the balance between generation and load. To fill this role, automated demand response must go beyond today's peak-shaving capability

To follow intermittent generation, automated demand response must be bi-directional control, i.e. it should provide for controlled increases and decreases in load. The response must also be predictable and preferably non-hysteretic, otherwise the load-generation imbalance may actually be exacerbated. Predictability would be highly valued by third party companies that aggregate loads into a pool of demand response resources. Finally, whatever control methodology is implemented, it must also be stable and not exhibit temporal oscillations. There are several factors that make achieving these demand response goals challenging: the different options for demand response signal, the uncertainty of the aggregate response to that signal, and the inhomogeneity of the underlying ensemble of loads.

The demand response control signal could take several forms: direct load control where some number of loads could be disabled via a utility-controlled switch[3], [4]; end-use parameter control where an ensemble of loads can be controlled by modifying the set point of the end-use controller, e.g. a thermostat temperature set point[5], [6], [7]; or indirect control via energy pricing in either a price taking (open loop) or auction (closed loop) setting [8], [**?**]. Today's automated demand response for peak-shaving is a form of direct load control which could be adapted and refined for the type of operation we desire, however, it is difficult to assess the impact of demand response on the end user because loads are simply disabled and re-enabled with little concern for the current state of the end use. Direct control is feasible for a relatively small number of large loads because the communication overhead is not extreme. Individual direct control of a large number of small loads would potentially overburden a communication system, however, "ensemble" control using a single parameter for control has been proposed, e.g. set point control for thermostatic loads[5], [6], [7] and connection rate control for electric vehicle charging[9]. However, in these control models, the underlying loads are assumed to be homogeneous (all of the same type), which is advantageous because it allows for a quantifiable measure of the end use impacts and customer discomfort, e.g. increasing all cooling thermostat set points by 1^oF will generate a decrease in load with a known end-use impact.

To control a large ensemble of *inhomogenous* loads with a single demand response signal requires a quantity that applies to all loads, i.e. energy pricing [10]. When given access to energy prices, consumers (or automated controllers acting on their behalf) can make their own local decisions about whether to consume or not. These local decisions open up new possibilities and also create problems. The customer is now enabled to automatically modify and perhaps optimize his consumption of energy to maximize his own welfare, which is a combination of his total energy costs and the completion of the load's end use function. However, without an understanding of how consumers respond to energy prices, the fidelity of the control allowed by the direct or ensemble control schemes described above is lost. Retail-

The work of MC at LANL was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396.

K. Turitsyn is with MIT, Mechanical Engineering, Cambridge, MA 02139 turitsyn@mit.edu

S. Backhaus is with MPA Division at LANL, Los Alamos, NM 87545 backhaus@lanl.gov

M. Ananyev is with New Economic School, Moscow, Russia maksim.ananjev@gmail.com

M. Chertkov is with Theory Division & Center for Nonlinear Studies at LANL, Los Alamos, NM 87545 and also with New Mexico Consortium, Los Alamos, NM 87544 chertkov@lanl.gov

level double auction markets[8], [11], [12] are an effective way of making demand response via pricing a closed-loop control system, however, a logical outcome of these markets would be locational prices potentially driven distribution system constraints making the regulatory implementation troublesome. In contrast, a model where retail customers are price takers may avoid some regulatory issues, however, price taking is in essence a form of open loop control which then requires an understanding of how the aggregate load on the system will respond to price.

Our goal in this initial work is to layout the computational framework for discovering the end-use response to these price-taking "open-loop" control systems. We develop state models for several different loads and subject them to a stochastic price signal that represents how energy prices might behave in an grid with a large amount of time-intermittent generation. We analyze the response of these smart loads using a Markov Decision Process (MDP) to optimize the welfare of the end user. Human owners of the devices have the ability to program the devices in accordance to their strategies and preferences, for instance by adjusting their willingness to sacrifice comfort in exchange for savings on electricity costs. Otherwise, most of the time we assume that the devices operate automatically in accordance to some optimal algorithm that was either preprogrammed by their owners, discovered via adaptive learning[13], or programmed by a third-party aggregator. The resulting load end-use policies can then be turned around to predict the effect of a change in prices on electrical load. Our long term strategic intention is to analyze the aggregated network effect on power flows of many independent customers and design optimal strategies for both consumers and the power operator. However, the prime focus of our first publication on the subject is less ambitious. We focus here on description of different load models and analyze the optimal behavior of individual consumers.

The material in the manuscript is organized as follows. We formulate our main assumptions and introduce the general MDP framework in Section II. Models of four different devices (optional, deferable and control loads and storage devices) are introduced in Section III. Our enabling simulation example of a control load (smart thermostat) is presented in Section IV. We summarize our main results and discuss a path forward in Section V.

II. Setting the Problems

A. Basic Assumptions

Future distribution networks are expected to show complex, collective behavior originating from competitive interaction of individual players of the following three types:

- Market operator, having full or partial control over the signals sent to devices/customers. The most direct signal is energy price. The operator may also provide subsidies and incentives or impose penalties, however in this manuscript, we will mainly focus on direct price control.
- Human customers/owners, who are able to reprogram smart-devices or override their actions.
- Smart devices, capable of making decisions about their operations. The devices are semi-automatic, i.e. pre-programmed to respond to the signal on a short time scale (measured in seconds-to-minutes) in a specific way, however the owner of the device may also choose to change the strategy on a longer time-scale (days or weeks). We model the smart devices as finite state machines using a Markov Decision Process (MDP) framework. At the beginning of each interval, a device decides how to change its state based on the current price. Each change comes with a reward expressing actual transactions between the provider and the consumer and the level of consumer satisfaction with the decision. We assume that smart devices are selfish and not collaborative, each optimizing its own reward.

In this manuscript we restrict our attention to a simple price-taking strategy of consumer behavior, deferring analysis of more elaborate game-theoretic interactions between the operator and the individual customers to further publications.

We model the external states (that include electricity price, weather, and human behavior) as a stochastic, Markov Chain process, $\{s^{(e)}(t)\}$. At the beginning of the time interval, t, the variable describing these factors is set to $s_t^{(e)}$ and changes during the next time step to $s_{t+1}^{(e)}$ with the transition probability $T(s_{t+1}^{(e)}|s_t^{(e)})$. The transition probabilities are assumed to be known to the device and statistically stationary, i.e. independent of t. (The later assumption can be easily relaxed to account for natural cycles and various external factors.) The probability, $p(s^{(e)};t)$, to observe the external state, $s^{(e)}(t)=s^{(e)}$, at the time t, follows the standard Markov chain equation

$$p(s^{(e)};t+1)=\sum_{s_t^{(e)}} T(s^{(e)},s_t^{(e)})p(s_t^{(e)}|t). \tag{1}$$

We also assume that the Markov chain (1) is ergodic and converges after a finite transient to the statistically stationary distribution: $p(s^{(e)};t+1)=p(s^{(e)};t)=p(s^{(e)})$. In the simulation tests that follow we will restrict ourselves to $s^{(e)}$ drawn from a finite set $S^{(e)}$.

B. General Markov Decision Process Framework

Here we adopt the standard (Markov Decision Process) MDP approach [14], [15], [16] to the problem of interest: description of smart devices responding to the external (exogenous) Markov process $\{s^{(e)}(t)\}$. MDPs provide a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of a decision maker. Formally, the MDP is a 4-tuple, $(S,A,P(\cdot,\cdot),R(\cdot,\cdot))$, where

- S is the finite set of states, in our case a direct product of the machine states set $S^{(m)}$, and the externality state set $S^{(e)}$, $S=S^{(m)}\otimes S^{(e)}$.
- A is a finite set of actions. A_s is the finite set of actions available from state $s\in S$. Within our framework we model only the decisions made by the machine, so

the set A consists only of actions associated with the machine, $A = A^{(m)}$.

- $P_a(s,s') = \Pr(s_{t+1} = s' \mid s_t = s, a_t = a)$ is the probability that action a chosen while in state $s = (s^{(m)}, s^{(e)})$ at time t will lead to state s' at time $t+1$. The probabilistic description of the transition allows to account for stochastic nature of the price fluctuations as well as for the randomness in the dynamics of the smart devices.
- $R_a(s,s')$ is the reward associated with the transition $s \to s'$ if the action a was chosen. In our models, the reward will reflect the price paid for electricity consumption associated with the transition as well as the level of discomfort related to the event.

In the most simple setting analyzed in this work, the behavior of the device is modeled via the policy function $\pi(s) : S \to A$ that determines the action chosen by the device for a given state: $a_t = \pi(s_t)$. More general formulations that include randomized decision making process, are not considered in this paper. Our smart device models seek to operate with the policy, $\pi(s)$, that maximizes over actions the expectation value of the total discounted reward, $\sum_{t=0}^{\infty} \gamma^t R_{a_t}(s_t, s_{t+1})$ over the Markov process, $P_a(\cdot,\cdot)$, where $0 < \gamma \leq 1$, is the discount rate. There are numerous algorithms used for optimizing the policies. In our work we use the algorithms implemented in MDP Matlab toolbox [16].

III. Models of Devices

The specifics of our MDP setting are to be described below for four examples of loads. Note that these examples are meant to illustrate the power of the framework and its applicability to "smart grid" problems. In this first paper, we do not aim to make the examples realistic. Instead, we focus on the qualitative features of the loads.The states and actions associated with the devices are illustrated in the diagrams shown in Figs. 1-4. For simplicity, we ignore the external part of the state $s^{(e)}$ in these diagrams. Full diagrams can be produced by taking the Kronecker product of transition graphs associated with the device and the external factors. In our diagrams, the states are marked by squares and actions are marked by dashed circles. Transitions from states to actions and actions to states are marked by dashed and solid arrows, respectively.

A. Optional Loads

A smart device described by an "optional load" pattern can operate in two regimes, at full and limited capacity. An example of such load is a light that can be automatically dimmed if the electricity price becomes too high (see Fig. 1). To simplify the mathematical notations, we denote the states of the machine $s^{(m)}$ by x. The machine can be in either of the two states: $x = 0$ and $x = 1$, shown as *Idle* and *Active* in the diagram (1) respectively. In the $x = 0$ state the machine does not operate (the lights are off). In the $x = 1$ state, the machine is active and the lights are shining at the full brightness, or are dimmed. Actions of the device are $a_0 =$ pass, $a_1 =$ full or $a_2 =$ shed. The a_0 action represents the process of waiting for the external signal of switching on the device. If no external external signal (requesting switching on) appears, the system returns to the $x = 0$ state, otherwise it moves to the $x = 1$ state. When the device is active (in the $x = 1$ state), it has two options: operate at full capacity, corresponding to the action a_1, or shed the load (dim the lights), corresponding to action a_2. Turning the device on or off is an externality dependent on a human. We assume that the external/human action is random, with the probability of turning the device ON and turning the device OFF being ρ_{ON} and ρ_{OFF} respectively. (For simplicity, we assume that the OFF signal may arrive only by the end of the time interval.) Assuming additionally that the transition probabilities do not not depend on the device actions, we arrive to the following expression for the transition kernel:

$$P_{pass}(s,s') = T(c'|c)\left[\rho_{ON}\delta_{x',1} + (1-\rho_{ON})\delta_{x',0}\right], \quad (2)$$

$$P_{full,shed}(s,s') = T(c'|c)\left[\rho_{OFF}\delta_{x',0} + (1-\rho_{OFF})\delta_{x',1}\right], \quad (3)$$

where δ_{x_1,x_2} is the Kronecker symbol: it is unity if $x_1 = x_2$ and zero otherwise.

There is no reward associated with either outcome of the $a_0 =$ pass action, however, the other two actions (a_1 and a_2) result in a reward consisting of two contributions. First is the price paid to the electricity provider, $E_{full,shed}c$, where $c(t)$ is the cost of electricity (considered as a component of $s^{(e)}$) and $E_{full,shed}$ is the amount of energy consumed during the time interval which depends on whether the lights are fully on or dimmed. (Here, $E_{full} > E_{shed} > 0$ and both values do not depend on the resulting state of the device). Second, the reward function accounts for a subjective level of comfort associated with the $a_{1,2}$ actions: $C_{full,shed}$. The discomfort of the light dimming is accounted by choosing $C_{full} > C_{shed}$. Summarizing, the cumulative reward function in this model of the optional load becomes

$$R_{pass}(s,s') = 0, \quad (4)$$

$$R_{full}(s,s') = C_{full} - E_{full}c, \quad (5)$$

$$R_{shed}(s,s') = C_{shed} - E_{shed}c. \quad (6)$$

Obviously, our model of optional loads is an oversimplification because there are a variety of additional effects which may also be important in practice, however, all these can be readily expressed within the MDP framework. For example, one may need to limit the wear and tear on the device, thus encouraging (via a proper reward) minimization of switching. (To account for this effect would require splitting the *Active* state in the model explained above into two states $Active - Full$ and $Active - Shed$.)

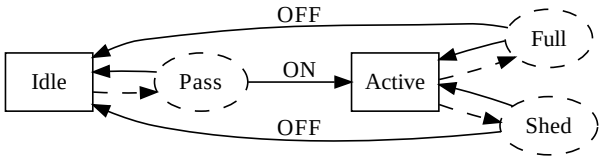

Fig. 1. MDP diagram for the model of optional load. See text for explanations.

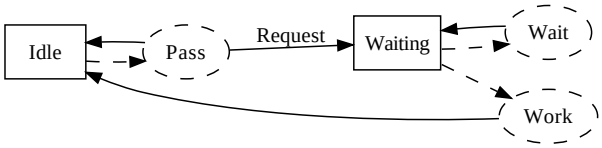

Fig. 2. MDP diagram for the model of deferable load. See text for explanations.

B. Deferable Loads

Our second example model is a deferable load, i.e. a load whose operation can be delayed without causing a major consumer discomfort. Practical examples include dishwashing machines or some maintenance jobs like disk defragmentation on a computer. A simple model of such a device, shown in Fig. 2, has two states: $x = 0$ ($Idle$) when no work is required and $x = 1$ ($Waiting$) when a job has been requested and the machine is waiting for the right moment (optimal in terms of the cost) to execute it. As in the previous model, the only action of the machine in the $Idle$ state is a_0 ($Pass$), however, in the $Waiting$ state, there are two possible actions: $a_1 = Wait$ results in waiting for possible drop of the electricity price and $a_2 = Work$ results in immediate execution of the job. The transition kernel for the model is

$$P_{Pass}(s,s') = T(c'|c)\left[\rho_{ON}\delta_{x',1} + (1-\rho_{ON})\delta_{x',0}\right], \quad (7)$$

$$P_{Wait}(s,s') = T(c'|c)\delta_{x',1}, \quad (8)$$

$$P_{Work}(s,s') = T(c'|c)\delta_{x',0}, \quad (9)$$

where ρ_{ON} is the probability of an exogeneous job request. In this model, there is no reward for choosing the $a_0 = Pass$ action. The reward for the $a_2 = Work$ action is equal to minus the price paid for the electricity, $R_{Work}(s,s') = -E * c$, and the reward for the $a_1 = Wait$ action represents the level of discomfort associated with the delay, $R_{Wait} = C_{delay} < 0$. As in the model of optional loads, E and C_{delay} are constants parameters. The obvious drawback of the presented model is its inability to ensure that the upper bounds on the waiting time. This shortcoming can be fixed by introducing more complicated rules on choosing the actions in the $Waiting$ states.

C. Control Loads

A very important class of devices that will likely play a key role in future demand response technologies are machines tasked to maintain a prescribed level of physical characteristics of some system. For example, thermostats are tasked with keeping the temperature in a building within acceptable bounds. Other examples of the control devices are water heaters, electric ovens, ventilation systems, CPU coolers etc.

In our enabling, proof-of-principle model of the control load, we consider a thermostat responsible for temperature control in a residential home. The state of the device is fully characterized by temperature which can take three possible values: $x = 0,1,2$ corresponding to $Low, Medium, High$ temperatures, respectively. Each temperature is assumed to be

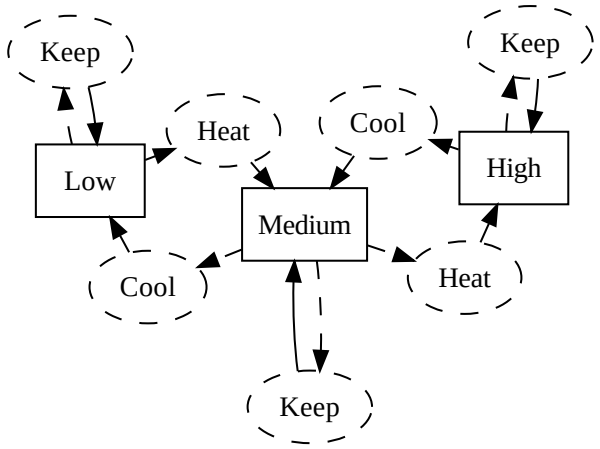

Fig. 3. MDP diagram for the model of controllable load. See text for explanations.

operationally acceptable. For simplicity, we assume that the thermostat uses an electric heater to modify the temperature (i.e. the outside temperature is low). The device can choose between the following three actions. $a_0 = Cool$ leaves the heater idle for the forthcoming interval. Since there is some base consumption associated with the thermostat operation we assume that $E_{Cool} > 0$. The next action, $a_1 = Keep$, maintains the temperature at the current level and requires some energy for heater operation: $E_{Keep} > E_{cool} > 0$. Finally, $a_2 = Heat$ corresponds to intensive heating that raises the temperature and requires the largest amount of energy E_{Heat}, and $E_{Heat} > E_{Keep} > E_{Cool} = 0$. Our thermostat state diagram, shown in Fig. (3), assumes that the dynamics of the thermostat are deterministic, and the resulting state depends only on the action chosen. The transition probabilities of the thermostat MDP is

$$P_{Heat}(s,s') = T(c'|c)\delta_{x',x+1}, \quad (10)$$

$$P_{Keep}(s,s') = T(c'|c)\delta_{x',x}, \quad (11)$$

$$P_{Cool}(s,s') = T(c'|c)\delta_{x',x-1}. \quad (12)$$

Assuming that all levels of temperature are equally comfortable, the reward function depends only on the price and energy consumption associated with the action,

$$R_{Cool,Keep,Heat}(s,s') = -cE_{Cool,Keep,Heat}. \quad (13)$$

For more realistic simulations our basic model should be generalized to account for different comfort levels of different states, the possibility for the owner to override an action, variations of the outside temperature, etc.

D. Storage loads

The number of devices with rechargeable batteries is expected to increase dramatically in the coming years. Currently, these are mostly laptops, uninterruptable power supplies, etc. In addition, a significant number of large-scale batteries will be added to the grid most likely via the anticipated Plug-in Hybrid Electric Vehicles (PHEV)

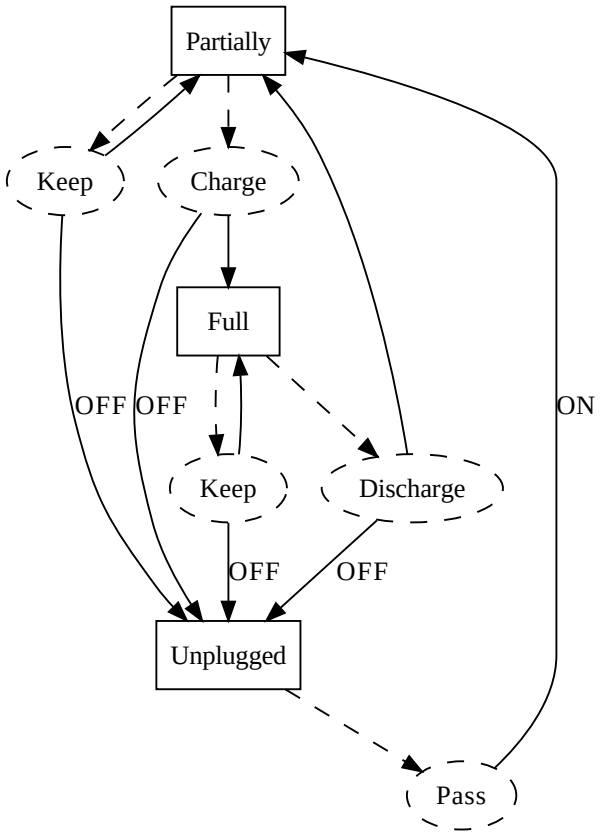

Fig. 4. MDP diagram for the model of storage. See text for explanations.

potentially enabled with Vehicle-to-Grid (V2G) capability. Storage devices, illustrated with the MDP in Fig. (4), share some similarity with the controlled loads discussed in the previous Subsection, but they are also different in two aspects. First, users/owners wants their devices to be charged which leads to a level of discomfort if the devices are not fully charged. Second, and probably most significantly, storage devices such as PHEVs are disconnected from the grid when in use. Having PHEVs in mind, we propose the following model of (mobile) storage. The system can be in either of the three states, the $x = 0 = Unplugged$ state (which is similar to the Idle state in the models of Optional and Deferable loads discussed above), the $x = 1 = Partially$ state where the storage is partially charged, and the $x = 2 = Full$ state where the device is fully charged.

The four available actions are: $a_0 = Pass$ when the device is in the unplugged state, the $a_1 = Keep$ action possible when the initial state is $x = 1 = Partially$ or $x = 2 = Full$, the $a_2 = Charge$ action available from the $x = 1 = Partially$ state which transitions to the $x = 2 = Full$ state, and, finally, the $a_3 = Discharge$ action, that is an inverse of the a_2 one, available from the $x = 2 = Full$ state resulting in the $x = 1 = State$. Except for $a_0 = Pass$, all these actions can be interrupted by transitioning at the end of the time interval to the $x = 0 = Unplugged$ state. As in previous sections, we assume that the unplugging happens at the end of a time interval. Assuming the device can be unplugged with the probability ρ_{OFF} and that it can be reconnected to the grid with the probability ρ_{ON}, we arrive at the following expressions for the transition probability:

$$P_{Pass}(s,s') = T(c'|c)\left[\rho_{ON}\delta_{x',1} + (1-\rho_{ON})\delta_{x',0}\right], \quad (14)$$

$$P_{Keep}(s,s') = T(c'|c)\left[\rho_{OFF}\delta_{x',0} + (1-\rho_{OFF})\delta_{x,x'}\right], \quad (15)$$

$$P_{Charge}(s,s') = T(c'|c)\left[\rho_{OFF}\delta_{x',0} + (1-\rho_{OFF})\delta_{x',2}\right], \quad (16)$$

$$P_{Discharge}(s,s') = T(c'|c)\left[\rho_{OFF}\delta_{x',0} + (1-\rho_{OFF})\delta_{x',1}\right] \quad (17)$$

The reward function accounts for the following effects. First, the $a_1 = Keep$ action has the cost associated with keeping the battery charged, $E_{Keep}(x)$, naturally dependent on the state, $E_{Keep}(2) > E_{Keep}(1) > E_{Keep}(0) = 0$. Second, the $a_2 = Charge$ action requires E_{Charge} of energy while the $a_3 = Discharge$ action generates the $E_{Discharge} < 0$ of energy, both nonzero only if the resulting state is not the $x = 0 = Unplugged$. Therefore, all the "active" actions, $Keep, Charge, Discharge$, contribute the reward function in accordance with the energy price, $c'E_{...}$. Finally, we also assign an additional negative reward, $C_{Unplug} < 0$, accounting for the discomfort (to the human) associated with being in the $x = 0 = Unplugged$ state. The resulting reward function is

$$R_{Pass}(s,s') = 0, \quad (18)$$

$$R_{Keep}(s,s') = C_{Unplug}\delta_{x',0}\delta_{x,1} - cE_{Keep}(x), \quad (19)$$

$$R_{Charge}(s,s') = -cE_{Charge}, \quad (20)$$

$$R_{Discharge}(s,s') = C_{Unplug}\delta_{x',0} - cE_{Discharge}. \quad (21)$$

IV. Simulations

In order to illustrate the capabilities of the proposed framework, we consider a simple model of the control load, describing a smart thermostat, characterized by $N_T = 10$ levels of the temperature parameter T. At every moment of time the thermostat can choose to raise, lower or keep the same temperature. The raise and lower options are not available at the highest and lowest possible temperatures, respectively. The energy consumption associated with the actions is given by $E_{Keep} = 1.0$, $E_{Cool} = 0.1$ and $E_{Heat} = 2.1$, respectively, in some normalized energy units. This choice of energies discourages the system from switching the heater too often: although the combinations $Heat + Cool$ and $Keep + Keep$ lead to the same temperature levels, the latter action is preferable as it consumes less energy.

Variations in price are modeled by a Markov chain of $N_P = 5$ equidistant levels with the minimum and maximum corresponding to 1.0 and 2.0 price units, respectively. At each time interval, the price either increases with probability $T(c+1|c) = 0.5$ by 1 level, decreases with probability $T(c-1|c) = 0.3$ by 1 level, or stays the same. The resulting stationary probability distribution $p(c)$ is shown in the Figure 5. It is skewed towards the higher price, mimicking the effect of intermittent renewable generators that occasionally provide excess power to the grid, thus leading to rapid dips in the price. The reward function (13) is fully determined by the total cost of energy consumed by the thermostat within

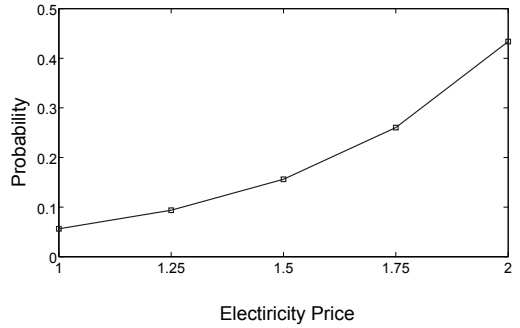

Fig. 5. Probability distribution of electricity price in the model example.

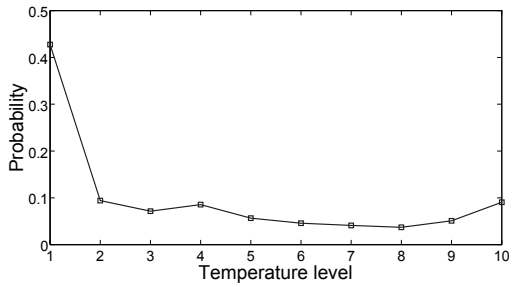

Fig. 7. Probability distribution of temperature levels observed at the optimal policy.

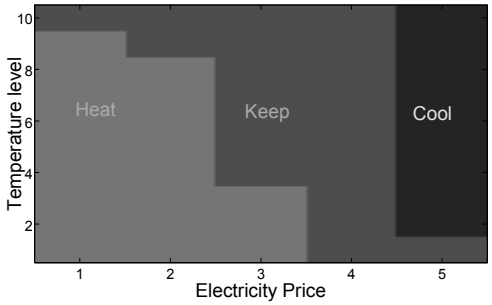

Fig. 6. Visualization of the policy found as a result of optimization.

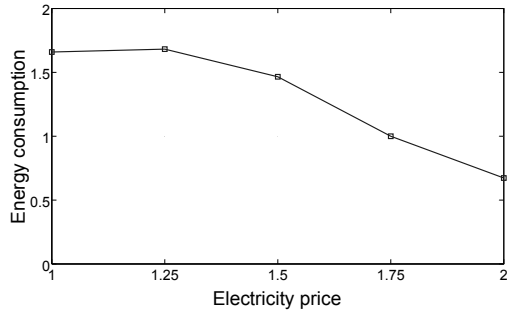

Fig. 8. Demand of the smart thermostats.

the given time-interval. Our MDP model imposes upper and lower bounds on the temperature, and we assume that there is no additional discomfort associated with the variations of temperature between these bounds, i.e. all of the N_T temperature levels are equally comfortable for the consumer.

This system was analyzed with the Matlab MDP package [16] where we used different algorithms to verify the stability of the results. The resulting optimal policy (for the range of parameters tested) is illustrated in 6. As expected, the thermostat chooses the *Heat* action when the price is low and decides to *Cool* when the price is high; a set of actions that lead to the skewed probability distribution of temperatures shown in Figure 7. One finds that the thermostat spends most of the time performing *Keep* in the low temperature state waiting for the price to drop.

Perhaps, the most interesting feature of the MDP model is the relation between consumption and price. We define the expected demand as the average energy demand for a given price

$$\langle E|c\rangle = \frac{\sum_x E_{\pi(x,c)} P_{st}(x,c)}{\sum_x P_{st}(x,c)}, \tag{22}$$

where $P_{st}(x,c)$ is the stationary joint distribution function of the temperature and price at the optimal strategy. Dependence of the consumption on the price for our choice of the parameters is shown in Figure 8, thus illustrating that variations in price indeed produce demand response. An interesting feature is that the demand curve is not monotonic. At low temperatures, the energy consumption shows a slight increase with the price; a surprising behavior related to saturation of the demand. When the electricity price decreases gradually from high to low levels, there is a high probability that the thermostat will reach the highest level of temperature before the price reaches the lowest level. In this case, the demand will be lower at the smallest price levels as there will be no unsatisfied demand left in the system to capitalize on the lowest price. From the economic viewpoint, it is important to note that this non-monotonicity of the demand curve reflects the adaptive nature of the MDP algorithm: the smart devices adjust to fluctuations in price, thus making it more difficult for the electricity providers to exploit the non-monotonic demand curve for making profit.

Remarkably, a non-trivial strategy was observed in our simulations despite the condition $E_{Heat} > E_{Keep} > E_{Cool} = 0$ that naively suggests that the "cool" action is always optimal whenever the temperature is not at the minimal level. The reason for the existence of non-trivial solutions lies in temporal correlations of the dynamics of price. Whenever the price falls below the average level it becomes advantageous for the device to rise the temperature to higher levels, to avoid overpaying for the "Keep" action during the forthcoming period of high price values.

Another interesting result found in our simulations is

an increase in average consumption of the smart (policy optimized) thermostat when compared to its non-smart counterpart, where the latter is defined as the one ignoring price fluctuations and sticking to the *Keep* action. For the set of parameters chosen in the test case, we observed that the average level of consumption in the optimal case is 1.03, i.e. it is 3% higher than in the naive strategy, an effect associated with the additional penalty (in energy) imposed on the *Heat* and *Cool* actions.

It is also instructive to evaluate savings of the consumer. The average value of the reward associated with the optimal policy is equal to -1.67, which should be compared with the reward of -1.73 generated by its non-smart counterpart. Since the reward reflects the customer's cost of electricity, we conclude that the customer saves about 3% on the electricity costs associated with the thermostat. The lower total energy costs for higher energy consumption was also seen in a related "smart-device" demonstration project[8].

V. Discussions, Conclusions and Path Forward

To conclude, we have presented a novel modeling framework to analyze future demand response technologies. The main novel aspect of our approach lies in the capability of the framework to describe behavior of the smart devices under varying/fluctuating electricity prices. To achieve this goal, we modeled the devices as rational agents which seek to maximize a predefined reward function associated with its actions. In general, the reward function includes the price paid for the electricity consumption and the level of owner discomfort associated with the choices made by the device. At the mathematical level, the system can be described via Markov Decision Processes that have been extensively studied over the last 50 years. Utilizing the MDP approach, we showed that a great variety of practical devices can be described within the same framework by simply changing the set of device states, actions and reward functions. Specifically, we identified four main device categories and proposed simple MDP models for each of them. These four categories include optional loads (like light dimming), deferrable loads (like dishwashing), control loads (thermostats and ventilation systems), and finally storage loads (charging of batteries).

To illustrate the approach we experimented with a simple model of a smart heating thermostat. The MDP-optimized policy of the thermostat followed the expected pattern: it chooses to not heat or keep the temperature stationary at high prices and prefers to heat when the price is low. This policy resulted in 3% of savings in the price paid for electricity, but at the same time led to the total of 3% increase in the consumption level due to the energy costs associated with the thermostat actions. The resulting demand curve showed a noticeable amount of elasticity, thus meeting the main objective of the demand response technology.

There are many relevant aspects of the model that we did not discuss in the manuscript. We briefly list some of these and future research challenges and direction.

- *Learning algorithms*. In our model we assumed that smart devices have an accurate model of stochastic dynamics for external factors (such as price for electricity), and use this model to find the optimal policy. In reality, however, this model is not known ab initio and has to be learned from the observations. Moreover, one can expect that the dynamics of external factors will be highly non-stationary (i.e. the transition matrix $T(s^{(e)}_{t+1}|s^{(e)}_{t+1})$ will have an explicit dependence on time). Therefore, the optimal policy has to be constantly adapted to the varying dynamics of the external factors. Of a special practical interest is the generalization of the framework to almost periodic processes, reflecting natural daily/weekly/yearly cycles in the electricity consumption.
- *Price-setting policies*. We did not discuss the price setting policies above, assuming that the policies are given/pre-defined. However, the electricity providers might adjust their policies to consumer response. As the electricity providers pursue their own goals, this setting essentially becomes game-theoretic and as such it requires more sophisticated approaches for analysis. Another extension of the model is to introduce auction-based price-setting schemes, such as in the Olympic Peninsula project [8]. This setting can be naturally incorporated in the same framework, although the modification may require simultaneous modeling of multiple (ensemble of) devices.
- *Time delays*. Another aspect of the real world not incorporated in our analysis concerns the separation of the time scales associated with operations of the device and intervals of the price variations. Multiple time-scale can be naturally incorporated in the framework by introducing additional states of the device. These modifications will certainly affect final answer for the optimal policy, and the resulting demand curve. However, accurate characterization of the multi-scale behavior will be a challenging task, requiring analysis of nonlinear response functions and dynamical description of the underlying non-Markovian processes.

VI. Acknowledgements

We are thankful to the participants of the "Optimization and Control for Smart Grids" LDRD DR project at Los Alamos and Smart Grid Seminar Series at CNLS/LANL for multiple fruitful discussions.

References

[1] N. Motegi, M. A. Piette, W. D., S. Kiliccote, and P. Xu, "Introduction to commercial building control strategies and techniques for demand response," LBNL Report Number 59975, Tech. Rep., 2007. [Online]. Available: http://gaia.lbl.gov/btech/papers/59975.pdf

[2] "U.s. department of energy, "benefits of demand response in electricity markets and recommendations for achieving them"," U.S. DOE, Tech. Report, 2006.

[3] S. S. Oren and S. A. Smith, "Design and management of curtailable electricity service to reduce annual peaks," *OPERATIONS RESEARCH*, vol. 40, no. 2, pp. 213–228, 1992. [Online]. Available: http://or.journal.informs.org/cgi/content/abstract/40/2/213

[4] R. Baldick, S. Kolos, and S. Tompaidis, "Interruptible electricity contracts from an electricity retailer's point of view: Valuation and optimal interruption," *OPERATIONS RESEARCH*, vol. 54, no. 4, pp. 627–642, 2006. [Online]. Available: http://or.journal.informs.org/cgi/content/abstract/54/4/627

[5] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, vol. 50, no. 5, pp. 1389 – 1400, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/B6V2P-4VS9KPY-1/2/32649b4a9a6779a2cea84379a7c1f9a6

[6] D. Callaway and I. Hiskens, "Achieving controllability of electric loads," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 184 –199, 2011.

[7] S. Kundu, N. Sinitsyn, S. Backhaus, and I. Hiskens, "Modeling and control of thermostatically controlled loads," in *17th Power Systems Computation Conference*, 2011, submitted.

[8] D. Hammerstrom and et al, "Pacific northwest gridwise testbed demonstration project:part i. olympic peninsuila project," PNNL-17167, Tech. Rep., 2007. [Online]. Available: http://gridwise.pnl.gov/docs/op_project_final_report_pnnl17167.pdf

[9] K. Turitsyn, N. Sinitsyn, S. Backhaus, and M. Chertkov, "Robust broadcast-communication control of electric vehicle charging," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 2010, pp. 203 –207.

[10] F. Schweppe, R. Tabors, J. Kirtley, H. Outhred, F. Pickel, and A. Cox, "Homeostatic utility control," *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-99, no. 3, pp. 1151 –1163, May 1980.

[11] L. Chen, N. Li, S. Low, and J. Doyle, "Two market models for demand response in power networks," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010, pp. 397–402.

[12] M. Roozbehani, M. Dahleh, and S. Mitter, "Dynamic pricing and stabilization of supply and demand in modern electric power grids," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010, pp. 543–548.

[13] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 2010, pp. 409 –414.

[14] R. E. Bellman, "A markovian decision process," *Jounral of Mathematics and Mechanics*, vol. 6, 1957.

[15] M. L. Putterman, *Markov Decision Processes. Discrete Stochastic. Dynamic Programming.* Wiley-Interscience, 2005.

[16] "Markov decision process (mdp) toolbox for matlab." [Online]. Available: http://www.cs.ubc.ca/~murphyk/Software/MDP/mdp.html