Provision of Regulation Service by Smart Buildings

Enes Bilgin, *Member, IEEE*, Michael C. Caramanis, *Senior Member, IEEE*,
Ioannis Ch. Paschalidis, *Fellow, IEEE*, and Christos G. Cassandras, *Fellow, IEEE*

***Abstract*—Regulation service (RS) reserves, a critical type of bi-directional capacity reserves, are provided today by expensive and environmentally unfriendly centralized fossil fuel generators. This paper investigates provision of RS reserves by the demand side. We consider a smart building operator that is capable of modulating the aggregate consumption of the building loads via price signals in response to an unanticipated RS signal that an independent system operator broadcasts. We first model the RS signal and load behavior, and formulate the related stochastic dynamic programming (DP) problem. Then, in order to deal with the complexity of the DP problem resulting from the uncountably infinite allowable price set, we characterize certain key properties of the DP dynamics, solve the DP problem for a discretized price policy to observe the structure of the optimal policy and re-capture the continuous price policy in an analytic approximate policy iteration (API) algorithm using the above properties and structure. We finally provide numerical evidence that the novel API algorithm converges to a continuous dynamic price policy that outperforms optimal discretized price policies in both computational effort and average cost.**

***Index Terms*—Distributed demand management, regulation service provision, real-time stochastic control, hour-ahead and real-time market response.**

I. Introduction

THE INTEGRATION of renewable energy into the power grid is progressing at an increasing rate ([1]). The intermittency and volatility of renewable generation, however, results in a commensurate increase in the reserves that Independent System Operators (ISOs) must secure ([2], [3]). In [4], a rough rule of thumb suggests an additional 1 MW RS reserve requirement per each 100 MW increase in the nameplate capacity of wind power in a system. Among various types of reserves, Regulation Service (RS) reserve is considered to be the most valuable one since an RS reserve provider is required to respond to an RS signal, which is updated every few seconds, by adjusting its output up and down. Not surprisingly, RS is also the most expensive reserve type on the market and its cost is comparable to or more expensive than the cost of energy. As an example, in the first quarter of 2014, the average price of electricity (Locational Marginal Price) was around $\$40-55$/MWh in the region managed by the Independent System Operator of the New England (ISONE). During the same time period, the average final hourly RS clearing price was \$44.4/MWh. In addition, 17% of the time, the RS price was above \$70/MWh, with a maximum of \$1407/MWh on January 28, 2014 at 8 a.m ([5]).

Today, RS reserves are provided by expensive and environmentally unfriendly centralized fossil fuel generators. This study investigates provision of RS reserves by the demand side. In particular, we consider a Smart Building Operator (SBO) that is capable of influencing the aggregate consumption of a population of flexible loads, through a dynamic control mechanism, in response to the RS signal that the ISO broadcasts to preserve the energy balance in the power system. The SBO of interest, we assume, anticipates an average aggregate consumption rate of A kW for the hour of operation and it purchases At kWh of energy in the forward power market (hour ahead market), where $t = 1$h. In addition, the SBO sells R kW of RS reserve in the hour ahead market for the same one-hour time horizon. As we denote the hour ahead energy and RS market clearing prices by Π_E and Π_R, respectively, the SBO is charged $tA\Pi_E - R\Pi_R$. Notice that the SBO is credited by $R\Pi_R$ for providing RS reserve. However, as an RS reserve provider, the SBO promises to modulate its consumption rate, $P(t)$, in the range $[A-R, A+R]$ during the hour of operation in order to track the RS signal, $y(t) \in [-1, 1]$, which is updated at $\Delta t = 4$-second time intervals as in the current practice at the ISONE. Perfect tracking requires that $P(t)$ reaches $A + Ry(t)$ by the next signal update, at $t + \Delta t$. Otherwise, the SBO is charged a penalty proportional to the tracking error $|P(t+\Delta t)-(A+Ry(t))|$. The SBO's optimal dynamic control policy must balance *(i)* the tracking error penalty minus the utility realized by flexible appliance users during the period Δt, against *(ii)* future costs affected by $P(t + \Delta t)$. This study focuses on developing such an optimal dynamic pricing policy for the use of the SBO to modulate its aggregate consumption in response to the RS signal for given A and R values. It is important to emphasize that the RS signal is determined by the ISO as a result of the energy imbalance in a much larger area compared to the size of the smart building of interest. Therefore, the RS signal is beyond the control of the SBO; and the effect of the SBO actions to

Manuscript received November 16, 2014; revised May 3, 2015 and October 2, 2015; accepted November 6, 2015. Date of publication December 2, 2015; date of current version April 19, 2016. This work was supported by the National Science Foundation through the Division of Electrical, Communications, and Cyber Systems under Grant EFRI-1038230. Paper no. TSG-01132-2014.

E. Bilgin is with IT Global Operations, Advanced Micro Devices, Inc., Austin, TX 78735 USA (e-mail: enes@bu.edu).

M. C. Caramanis is with the Department of Mechanical Engineering and the Division of Systems Engineering, Boston University, Brookline, MA 02446 USA (e-mail: mcaraman@bu.edu).

I. C. Paschalidis and C. G. Cassandras are with the Department of Electrical and Computer Engineering, and the Division of Systems Engineering, Boston University, Boston, MA 02215 USA (e-mail: yannisp@bu.edu; cgc@bu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2015.2501428

the overall energy balance is insignificant. Also note that the discussion about how to chose A and R in the hour ahead market is beyond the scope of this paper, for which we refer the reader to [6].

Obtaining a dynamic control policy for the SBO requires *(i)* modeling the dynamics of the RS signal, which is essentially an exogenous random variable, and *(ii)* modeling the energy consumption behavior of the flexible loads together with their interaction with the price control. This paper provides a Markov chain representation of the RS signal dynamics. In order to model the latter, we consider a generic flexible appliance whose operation and requirements are similar to autonomous heating or cooling appliances, space air renewal ventilators, refrigerators, electric water heaters, and in general loads that demand a fixed amount of energy over a certain time period. Rather than controlling the appliance consumption directly, the SBO broadcasts a price signal. Then, each appliance makes the decision -in an automated manner- of whether to consume energy or not by comparing the price with its energy need at discrete time intervals. This type of control approach is known as "indirect control" in power systems literature. In [7], the author proposes a hierarchical control strategy for the power system and introduces the idea of price-based demand response. In [8], the use of real-time prices to assist in the control of frequency and tie-line deviations in electric power systems is discussed. In [9], the authors develop a state-queuing model to analyze the price response of aggregated loads consisting of thermostatically controlled appliances (TCAs). Finally, in [10], a model predictive control strategy is proposed for tracking an RS signal by groups of flexible loads such as plug-in hybrid electric vehicles (PHEVs) and TCAs.

This paper is an extension and generalization of past work [11] and [12]. In particular, the model investigated here differs relative to [11] in the following ways: *(i)* RS signals are uncontrollable by the SBO as they depend on a proportional integral filter of the Area Control Error and System Frequency excursions, *(ii)* a Two-Dimensional Markov chain model whose parameters are estimated using actual historical RS signal dynamics is used to represent the statistical behavior of $y(t)$ available to the SBO, and *(iii)* a dynamic control strategy is obtained in place of a static asymptotic control. Furthermore, we derive properties of the average cost resulting from the optimally controlled RS provision and show that these properties can assist the SBO in bidding optimally for Energy and RS in the Hour Ahead Market. We also extend the results reported in [12] as they pertain to the utility of smart building appliance users and provide a novel Approximate Policy Iteration (API) solution algorithm and extensive numerical experience. The rest of the paper is organized as follows: In Section II, we present a nomenclature that summarizes the definitions in the paper. In Section III, we model the RS signal stochastic dynamics, define the flexible load response to SBO control actions, and formulate the resulting Stochastic Dynamic Programming problem. In Section IV, we compare dynamic and static control policies, and show how the average appliance user utility decreases under dynamic control policies. In Section V, we develop a Linear Programming based solution methodology for a discretized allowable control space approximation, make observations on the structure of the associated optimal policy, and use it to develop an analytic approximate policy iteration (API) approach that can be calibrated to relax the control space discretization approximation and provide near optimal control. We present numerical results in Section VI and conclude in Section VII.

II. Nomenclature

Π_E	Energy clearing price (\$/kWh).
Π_R	RS reserve clearing price (\$/kW).
A	Anticipated average energy consumption rate for the SBO given Π_E (kW).
R	Amount of RS reserve the SBO sells in the hour ahead power market (kW).
$P(t)$	SBO energy consumption rate at time t (kW).
T	Time horizon of the DP problem (min).
$y(t)$	Value of the RS signal at time t.
Δy	Change in the RS signal between two updates.
$\Delta\bar{y}$	Maximum value allowed for $\lvert\Delta y\rvert$.
$d(t)$	RS signal direction at time t, $d(t) \in \{-1, +1\}$.
N	Number of appliances subject to SBO control.
$n(t)$	Number of active appliances at time t.
n_m	Lower bound imposed on $n(t)$.
n_M	Upper bound imposed on $n(t)$.
Δn	Change in the number of active appliances between sequential time steps.
r	Energy consumption rate per appliance (kW).
$u(t)$	Price signal the SBO broadcasts at time t (\$).
λ_a	Rate at which an appliance monitors the price control $u(t)$ (1/min).
λ_M	Maximum aggregate connection rate (1/min).
$\lambda(t)$	Aggregate connection rate at time t (1/min).
$\bar{\lambda}$	Average aggregate connection rate (1/min).
μ	Active appliance disconnection rate (1/min).
U_M	Maximum value that $u(t)$ and $\phi_i(t)$ can take(\$).
Δu	Discretization increment for $u \in [0, U_M]$ (\$).
$\bar{u}$	Price average over $[0, T]$ (\$).
σ_u^2	Variance of the price $u(t)$ over $[0, T]$.
$\phi_i(t)$	Utility of appliance i at time t (\$).
$\Phi(\phi(t))$	Probability density function of $\phi(t)$.
$\Phi_{[t_1,t_2]}$	Utility realized over $[t_1, t_2]$ (\$).
$C_{[t_1,t_2]}$	Tracking cost incurred over $[t_1, t_2]$ (\$).
κ	Tracking cost coefficient ($\$/kW^2s$).
$T_i(t)$	Temperature in zone i at time t (°F).
$T_{min,i}$	Desired minimum temperature for zone i (°F).
$T_{max,i}$	Desired maximum temperature for zone i (°F).
$h(x)$	Differential cost of the system state x (\$).
J	Average cost of the DP problem (\$).
$p_{\boldsymbol{\theta}}(x)$	Linear Programming dual variable associated with state x with parameter vector $\boldsymbol{\theta}$.
θ_i	i^{th} element of the parameter vector $\boldsymbol{\theta}$.
$\theta_i(k)$	Value of θ_i at the k^{th} iteration of the API.
$u_\theta(x)$	Price policy under $\boldsymbol{\theta}$ for state x.
$\Delta\theta$	Upper bound imposed on $\lvert\theta_i(k+1) - \theta_i, (k)\rvert$.
ρ	API algorithm parameter to scale $\Delta\theta$.

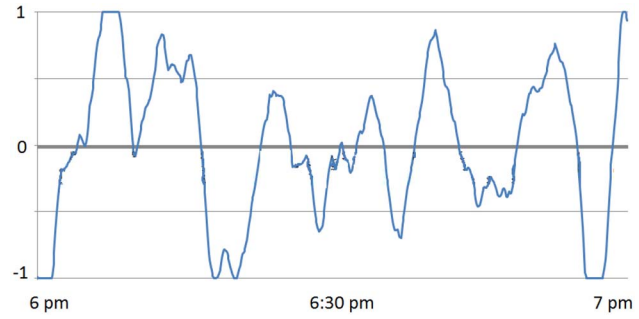


Fig. 1. Sample RS Signal Data from PJM.

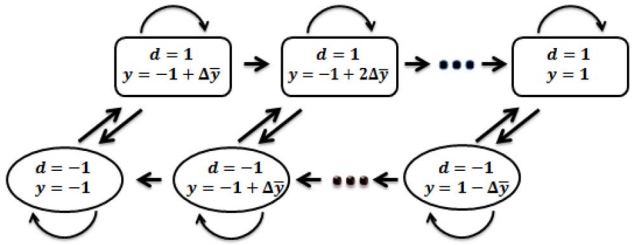


Fig. 2. Markov Chain Representation of RS Signal Dynamics, $y \in \{-1, -1+\Delta\bar{y}, \ldots, 1-\Delta\bar{y}, 1\}$, $d \in \{-1, 1\}$.

III. Problem Formulation

A. RS Signal Dynamics and Tracking Costs

ISOs manage the RS reserves, which are procured from multiple suppliers in the Hour Ahead Market, in real-time by broadcasting a single RS signal $y(t) \in [-1, 1]$ that is updated in Δt second intervals. Although Δt may vary among different ISOs, we follow the ISONE practice by setting $\Delta t = 4$ seconds. The RS signal is non-deterministic as ISOs calculate the signal value through a proportional-integral filter of system wide frequency deviation.

Assumption 1: The RS signal the ISO broadcasts is Energy Neutral over an hour, i.e., $\sum_{k=0}^{K-1} y(k\Delta t) = 0$, where $K = 900$, i.e., $K\Delta t = 1$h for $\Delta t = 4$ sec.

Assumption 1 implies that if the SBO tracks the regulation signal without error, its total consumption over the hour of operation will be equal to At kWh, $t = 1$h, the quantity it purchased in the hour ahead market.

Figure 1 displays a sample of actual PJM RS Signal data over an hour. The figure shows that RS signal dynamics exhibit monotonically increasing or decreasing sub-trajectories. We represent these dynamics by a Markov chain model whose state space is two dimensional, where the first dimension is the value of the RS signal, $y(t)$, and the second dimension is the "trend" or the "direction" of the RS signal, $d(t)$. This two-dimensional Markov chain (TDMC) is shown in Figure 2. The RS signal, $y(t)$, may take any value in $[-1, 1]$ subject to a maximal ramp rate $\Delta\bar{y}$ between two updates. However, it is adequate to represent its probabilistic change by $\Delta y = y(t+\Delta t) - y(t)$, where Δy is a discrete random variable taking values in the set $\{-\Delta\bar{y}, 0, \Delta\bar{y}\}$. The probability mass function of Δy is calibrated appropriately conditional upon $y(t)$ and $d(t)$. The dynamics of $d(t)$ are completely specified by the dynamics of $y(t)$ as follows:

$$d(t+\Delta t) = \begin{cases} S\{\Delta y\}, & \text{if } \Delta y \neq 0 \\ d(t), & \text{if } \Delta y = 0, \end{cases} \tag{1}$$

where S is the sign function. Note that $d(t) \in \{-1, +1\}$, so that there are only two possible states for the direction, namely "down" and "up." Based on the value Δy takes, the direction either stays the same or it is reversed. When $d(t) = +1$, and $y(t)$ is small, then $Pr[\Delta y \geq 0 | d(t) = +1] \geq Pr[\Delta y = -\Delta\bar{y} | d(t) = +1]$, and vice versa for $d(t) = -1$. In short, the transition probabilities of $y(t)$ depend on the value of $d(t)$ as well as on the magnitude of $y(t)$ that takes values in the discrete state space, i.e., $y(t) \in \{-1, -1+\Delta\bar{y}, \ldots, 0, \ldots, 1-\Delta\bar{y}, 1\}$. As $y(t)$ approaches 1, its upward transition probability decreases to 0. Similarly downward transition probabilities decrease to 0 as $y(t)$ approaches -1. In order to calibrate the probability matrix of the TDMC, one needs to derive the state frequencies of the RS signal using historical ISO data, and count the number of upward and downward transitions in each state.

B. Consumption Dynamics and Utility

Assume that the *Smart Building (SB)* of interest has N flexible appliances (e.g., heating or cooling zones, electric vehicles (EV) etc.) whose aggregate consumption is to be modulated in response to the RS signal. $n(t)$ of N appliances are "active", i.e., they consume electricity at time t each at a rate of r kW. At the same time, $N-n(t)$ of N are "idle" and they do not consume electricity at time t. We define the transition of an appliance from idle to active as a "connection" and the transition from active to idle as a "disconnection". We assume that each appliance has a smart controller that is designed to ensure that it remains active for an exponentially distributed time with parameter μ, so that the average connection time is $1/\mu$. In addition, to avoid immediate switching back to an active state, the controller is designed to reconsider connecting after an exponentially distributed time with parameter λ_a. Probabilistic monitoring and disconnections of the appliances are on purpose and to avoid undesirable chattering, synchronization, and exceedingly short active status duration ([9]).

The SBO can influence the rate at which idle appliances connect by broadcasting a "price" $u(t)$. When an idle appliance i considers connecting at time t, it proceeds to connect if and only if its "utility" exceeds $u(t)$. The utility level of appliance i at time t, $\phi_i(t)$, represents the occupant's "value for connecting", i.e., the value measured in dollars for connecting and consuming at the rate r for an exponentially distributed time period with average $1/\mu$. It is reasonable to consider that $\phi_i(t)$ will be a function of the appliance i heating zone temperature at time t, $T_i(t)$, a point in the occupant's comfort interval $[T_{min,i}, T_{max,i}]$. When $T_i(t)$ is closer to $T_{min,i}$ during the cooling season, $\phi_i(t)$ should be low, and when it is closer to T_{max} $\phi_i(t)$ should be high. Here, we assume that the appliance utility is always nonnegative. Moreover $U_M > 0$ denotes the maximum utility for appliance i for the case $T_i(t) \geq T_{max,i}$; therefore we note $\phi_i(t) \in [0, U_M]$. As a consequence, it is appropriate to limit the values that $u(t)$ can take to the same

interval, i.e., $u(t) \in [0, U_M]$. Note that U_M indicates the price level beyond which no connections occur; and this is also how the SBO could obtain an estimate of U_M from the historical consumption data.

We assume that the SBO updates $u(t)$ just after the ISO updates $y(t)$, namely in Δt second intervals. Since $N - n(t)$ gives the number of idle appliances at time t, the total rate at which the idle appliance population monitors the price at time t is given by $(N - n(t))\lambda_a$. Consider now the utilities of the idle appliances monitoring the SBO price, $\{\phi_1(t), \phi_2(t), \ldots, \phi_i(t), \ldots, \phi_{N-n(t)}(t)\}$. Given that each idle appliance is equally likely to monitor the SBO price at time t due to the identical, independent and exponentially distributed inter-monitoring times, these utilities are statistically equivalent to a random sample of size Z selected from $N - n(t)$ idle appliances at time t. We can construct the frequency histogram of utility samples grouped in small intervals Δu spanning the segment $[0, U_M]$. For large $N - n(t)$ the histogram normalized by Z converges to a p.d.f. $\Phi(\phi(t))$ as Δu approaches 0. $\Phi(\phi(t))$ characterizes a random process $\phi(t)$ representing the utility of a randomly selected idle appliance at time t. The rate of idle appliances connecting when the SBO broadcasts price $u(t)$ can be now expressed as $\lambda_a(N - n(t))Pr\{\phi(t) \geq u(t)\}$, where $Pr\{\phi(t) \geq u(t)\}$ is given by $\Phi(\phi(t))$. Use of $\Phi(\phi(t))$ simplifies immensely the description of the dynamics of $n(t)$ which would otherwise require that the state includes all of the individual idle appliance utilities $\phi_i(t), \ \forall \ i$. Given the relationship of $\phi_i(t)$ and individual idle appliance space conditioning zone temperature $T_i(t)$, $\phi(t)$ varies over time in a manner that is similar to that of the p.d.f. of $T_i(t)$ over all appliances i. We describe below conditions under which $\phi(t)$ can be approximated reasonably by a time invariant p.d.f. that is constant over the random variable's range, hence uniform, over $[0, U_M]$.

Assumption 2: $n(t)$ takes values in a bounded range such that $n_m \leq n(t) \leq n_M$, where n_m and n_M are positive constants, $n_M - n_m = 2R/r$ and $n_M - n_m \ll N - n_M$.

When $n(t)$ takes values in, $n_m \leq n(t) \leq n_M$, where $n_M - n_m \ll N - n_M$ and $(N - n_M)/(N - n_M) \approx 1$, it is reasonable to approximate $(N - n(t))\lambda_a$ by a time invariant constant $\lambda_M = (N - (n_m + n_M)/2)\lambda_a$, and $\Phi(\phi(t))$ by a time invariant uniform distribution over $[0, U_M]$. The assumption above holds under weather and time invariant building heat loss conditions that result in duty cycle appliances being active a small fraction of the time rendering $n(t) \ll N$, and, in addition, hour ahead market bids of RS levels that are relatively small, resulting in $2R/r = n_M - n_m \ll N - n_M$. We have verified the veracity of the assumption with extensive simulation, where we have observed that the aggregate consumption stay in the $A \pm 3R/2$ range almost all the time. Although the time invariant uniformity assumption is a sacrifice in the model accuracy for the sake of simplicity, it is adopted both in communication networks and power systems literature together when the user pool is considered to be sufficiently big (e.g., [11] and [13]).

We summarize the consequence of the discussion as follows.

Assumption 3: At any given time t, the utility of an idle appliance is determined by drawing a random number that is uniformly distributed in $[0, U_M]$.

Therefore, we simulate a random variable ϕ from a time invariant uniform probability distribution over the interval $[0, U_M]$.

Poisson Connections and Disconnections During Δt: Given Assumption 3, the probability that an idle appliance connects when it observes the price $u(t)$ at time t is $(1 - u(t)/U_M)$. Moreover, given Assumption 2, the aggregate connection rate at time t is $\lambda(t) = \lambda_M(1 - u(t)/U_M)$, which limits the connection rate to the set $\lambda(t) \in [0, \lambda_M]$.

For the dynamic connection rate $\lambda(t)$ and the constant disconnection rate per active appliance μ, the controlled dynamics of $n(t)$ coincide to the dynamics of an $M/M/\infty$ queue. Note that the connection rate is constant between price updates. Moreover, if a constant price $u \in [0, U_M]$ is broadcast over consecutive time intervals, the connection rate is $\lambda = \lambda_M(1 - u/U_M)$, i.e., time invariant, and the steady state distribution of $n(t)$, as $t \to \infty$, is Poisson distributed with rate λ/μ.

C. Objective Function

Recall that the SBO is required to modulate its aggregate consumption rate as it observes the RS signal $y(t)$ at time t so as to reach $A + Ry(t)$ kW by the time $t + \Delta t$. In other words, it promises to achieve connected appliances $n(t + \Delta t)$, each consuming at the same rate of r kW, so that $(n(t+\Delta t))r$ is close to $A + Ry(t)$. To this end, the SBO's task is to influence the idle appliance connection rate by broadcasting an appropriate price $u(t)$ over the period from t to $t+\Delta t$. To the extent that it fails to track the RS signal, the SBO is assessed the following tracking cost,

$$C_{[t,t+\Delta t]} = \Delta t \kappa \big[(n(t + \Delta t))r - (A + Ry(t))\big]^2 \qquad (2)$$

where $\kappa \geq 0$ is a fixed cost coefficient. We henceforth refer to $(n(t + \Delta t))\ r - (A + Ry(t))$ as the "tracking error". Note that the quadratic cost function aims to capture two types of costs that the SBO incurs in case of imperfect tracking: *(i)* Loss in the revenue of RS reserve provision, $R\Pi_R$, *(ii)* the possible cost of losing RS provision license, which happens in the case of exceedingly poor tracking ([14]) of the RS signal. Therefore, severe tracking errors are penalized more due to the increased possibility of the latter case.

As mentioned above, each appliance realizes a utility when it connects to the system. We consider this utility as a contribution to smart building's social welfare. Given Assumption 3, the expected utility realized by an appliance that connects at time t is $\mathbb{E}[\phi(t)|\phi(t) \geq u(t)] = (u(t) + U_M)/2$. Therefore, the expected utility realized over $[t, t + \Delta t]$ is

$$\Phi_{[t,t+\Delta t]} = \Delta t \lambda_M (1 - u(t)/U_M)(u(t) + U_M)/2. \qquad (3)$$

An optimal SBO price control policy must trade off optimally tracking cost against appliance user utility gained every time an idle appliance connects. Therefore, the SBO's objective is

$$\min \quad \sum_{k=0}^{(K-1)\Delta t} C_{[k\Delta t,(k+1)\Delta t]} - \Phi_{[k\Delta t,(k+1)\Delta t]}, \qquad (4)$$

where $\Delta t = 4$ seconds and $K = 900$, so that $K\Delta t = 1$ hour.

D. Stochastic Dynamic Programming Model

As mentioned before, the SBO makes price decision every Δt seconds, which is significantly shorter than the one-hour problem horizon. It is also apparent that SBO makes the price decisions sequentially, which implies Markov Decision Process given the system model. Therefore, the problem is modeled as discrete-time, finite-state, average cost infinite horizon Dynamic Programming (DP) problem, which is characterized by the following Bellman equation.

$$h(n,y,d)+J = \min_{u\in[0,U_M]}\left\{\mathop{\mathbb{E}}_{\Delta n,\Delta y,d'|u,n,y,d}\big[\Delta t g(u,n,y,d) + h\big(n+\Delta n, y+\Delta y, d'\big)\big]\right\} \quad (5)$$

where

- d' is the new direction, whose stochastic dynamics are described in Equation 1,
- $\Delta y = y(t+\Delta t) - y(t)$ is an exogenous random variable independent from u,
- $\Delta n = n(t+\Delta t) - n(t)$ is a random variable that depends on u,
- J is the average cost per Δt,
- $h(n,y,d)$ is the differential cost function ([15]),
- $g(u,n,y,d)$ is the one step cost function, i.e.,

$$g(u,n,y,d) = \kappa((n+\Delta n)r - (A+Ry))^2 - \lambda_M(1-u/U_M)(u+U_M)/2.$$

The optimal price policy, $u^*(n,y,d) \in [0,U_M]$, is a function of the state of the system, n,y,d. Once such a policy is obtained, the SBO could observe n,y and d values every Δt seconds and broadcast the price the optimal policy recommends in an automated manner. Given Assumption 2, a single price update will only be observed by a small fraction of the appliance population, which will prevent big fluctuations in the aggregate consumption that might have been caused by the frequent changes in the price.

IV. Impact of Dynamic Price Policies on Average Consumption and Utility

In this section, we discuss the implications of having a dynamic price policy and how it affects the aggregate consumption and utility of the appliances. For this reason, we define a discrete time dynamic policy $u(k) \in [0,U_M]$, so that $u(t) = u(k)$, $t \in [k\Delta t, (k+1)\Delta t]$, where $k = 0,1,\ldots,K-1$, with $K\Delta t = T$. In addition, we define $u_c \in [0,U_M]$ to represent the static price policy of broadcasting a constant price, i.e., $u(t) = u_c$, $t \in [0,T]$. Note that related proofs for the propositions and the corollary in this section are included in the Appendix.

A. Effect of Dynamic Prices on Average Consumption

Proposition 1: The expected electricity consumption of all N appliances over the horizon $[0,T]$ is invariant to dynamic pricing policies $u(k)$ that satisfy $\frac{1}{T}\sum_{k=0}^{K-1} u(k)\Delta t = u_c$, where $T = K\Delta t$.

Note that if the SBO were not an RS reserve provider, it would directly reflect the hour ahead energy clearing price in its price policy, which would be a constant one since the SBO would not have to modulate the aggregate consumption; therefore it would have $u_c = \Pi_E$. Also recall that A is the anticipated average energy consumption rate under Π_E. On the other hand, the SBO is an RS reserve provider and it has to employ a dynamic price policy so that $P(t)$ tracks $A+Ry(t)$. By Assumption 1, when the RS signal is tracked closely, the average energy consumption rate will be equal to A. Now, Proposition 1 implies that a dynamic price policy that leads to an average consumption rate A must have a time average equal to Π_E, i.e., $\bar{u} = \frac{1}{T}\int_0^T u(t)dt = \Pi_E$. Defining $\bar{\lambda} = \lambda_M(1-\bar{u}/U_M)$, we derive the following corollary.

Corollary 1: For sufficiently large λ_M and κ, the following equation holds:

$$A = \bar{\lambda}^* r/\mu \quad (6)$$

where

- $\bar{\lambda}^* = \lambda_M(1-\bar{u}^*/U_M)$,
- u^* is the optimal price policy,
- $\bar{u}^* = \frac{1}{T}\int_0^T u^*(t)dt$,
- T is the problem horizon.

In Section VI-B, Corollary 1 is verified numerically.

B. Effect of Dynamic Price Variation on the Time-Average of Utility

Proposition 2: The Time-Average of the utility rate realized over the period $[0,T]$ is smaller under a dynamic pricing policy $u(k)$ than under the static price policy $u_c = \frac{1}{T}\sum_{k=0}^{K-1} u(k)\Delta t$, for $T = K\Delta t$. Moreover, the difference in the average utility rate is equal to $\lambda_M\sigma_u^2/2U_M$.

This is an important result on two counts: First, provision of Regulation Service reserves results in utility loss on the part of building appliance users. Hence, even if the SBO is able to track the RS signal in real-time and avoid tracking error penalties, it must take this prospective utility loss into consideration while bidding for energy and reserves in the hour ahead market. Second, responding to RS signals aggressively will certainly result in a higher utility loss. The clear trade-off among tracking error cost and utility, justifies the effort to determine an optimal price policy to which we turn next.

V. Determining the Optimal Dynamic Policy

In this section, we propose and implement two distinct solution approaches to the DP problem formulated in Section III-D: *(i)* A well known Linear Programming (LP) based method [15], which requires discretization of the allowable control space, and *(ii)* an innovative computationally tractable Approximate Policy Iteration (API) algorithm which bypasses the need to discretize the allowable control space through reliance on an analytic functional approximation of the control policy inspired by optimal policy structural properties revealed in numerical experience with the LP approach.

A. Linear Programming Based DP Solution Approach

Denoting for notational simplicity the discrete state (n, y, d) by x and the set of all possible states by X, i.e., $x \in X \subset \mathcal{Z}$, and discretizing the control set $u \in \mathcal{U} = \{0, \Delta u, 2\Delta u, \ldots, m\Delta u\}$, where $\Delta u = U_M/m$, allow us to write the optimality-conditions-representing Bellman equation as the following Linear Program ([15])

$$\begin{aligned} &\underset{J,\mathbf{h}}{Maximize} \quad J \\ &J\mathbf{v} + \mathbf{h} \leq \mathbf{g}(u) + P(u)\mathbf{h}, \qquad \forall u \in \mathcal{U} \end{aligned} \tag{7}$$

where

- $\mathbf{h}$ and $\mathbf{g}(u)$ are vector notations for $h(x)$ and $g(x, u)$ respectively for $\forall x \in X$,
- $h(x)$ and J are the differential cost and average cost rate as defined in Section III-D,
- $g(x, u) = \underset{z|x}{\mathbb{E}}[\Delta t\kappa((n + \Delta n)r - (A + Ry))^2 - \Delta t\lambda_M(1 - u/U_M)(u + U_M)/2]$ is the expected period cost for control u,
- $x = (n, y, d)$ and $z = (n + \Delta n, y + \Delta y, d')$ represent the current and the next system state respectively,
- $P_{xz}(u)$ is the probability of a transition from state x to z when the price control is u, and $P(u)$ is the corresponding transition matrix,
- $\mathbf{v}$ is vector of ones.

The optimal control policy u^* is the mapping of a state x to a policy $u^*(x)$ obtained by identifying the specific u of the $m + 1$ LP constraints considered for each state i, which is binding in the optimal solution yielding a nonzero dual variable $p(x)$. The dual variables, $p(x)$, are useful LP solution output as they represent the steady state probability that the controlled system visits a particular state x using the associated optimal policy $u^*(x)$ ([15]). This allows us to estimate the expected utility that the system will realize under optimal dynamic prices, the expected variance of the dynamic price, and the expected utility loss as discussed in Section IV-B.

Numerical solutions to various instances of the discretized DP problem reveal interesting properties of the optimal policy, indicating that the tracking error, $nr - (A + Ry)$, is a useful feature of the state space. More specifically, when $nr - (A + Ry) \approx 0$, the optimal price tends to be close to u_s that satisfies

$$\begin{aligned} A + Ry &= \lambda_M(1 - u_s/U_M)r/\mu \\ u_s &= U_M\left(1 - \frac{\mu}{\lambda_M r}(A + Ry)\right). \end{aligned} \tag{8}$$

In light of previous discussion on the distribution of n under a constant price control, note that u_s is the price level that is asymptotically related to $(A + Ry)/r$ active appliances in expectation. Indeed, recall that the number of active appliances behaves as an $M/M/\infty$ queue, which is Poisson distributed in steady state with mean λ/μ. Moreover, numerical results indicate that, for fixed y and d, the optimal prices are non-decreasing in the tracking error, $nr - (A + Ry)$, approaching 0 and U_M asymptotically. Consequently, for fixed y and d, the optimal price policy tends to have a sigmoid shape as a function of the tracking error. As discussed further in Section VI-B, the prevalence of the sigmoid property is strengthened by the observation that changes in y and d shifts the sigmoid along the axis that measures the tracking error.

Motivated by the discretized optimal policy properties discussed above, we propose a continuous optimal price policy approximation by optimizing the parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4]$ in the following sigmoid function

$$U^*(x) \approx u_{\boldsymbol{\theta}}(n, y, d) = \frac{U_M}{1 + e^{\theta_1(nr - (A+Ry)) + \theta_2 y + \theta_3 d + \theta_4}} \tag{9}$$

The functional approximation reduces the computational effort to the optimal selection of the parameters in $\boldsymbol{\theta}$, a task undertaken in the next section where we develop an approximate policy iteration (API) algorithm to this end.

B. Novel Approximate Policy Iteration DP Solution Algorithm

Denote the cardinality of the state space by $|X|$. Note that, given $n_m \leq n(t) \leq n_M$, $t \geq 0$ by Assumption 2, we have $|X| = 2(n_M - n_m + 1)(1 + (2/\Delta\bar{y})$. Also recall that the size of the discretized allowable control set is $m + 1$. A price policy is represented in general by a lookup table with $|X|$ number of rows mapping each state x to a price $u(x)$. The LP based DP solution approach requires solving a linear problem with $|X| + 1$ decision variables and $(m + 1)|X|$ constraints. In contrast, a policy iteration algorithm consists of consecutive estimates of *(i)* the average and differential cost for a policy mapping states to prices, $u(x)$, which requires the solution of a linear system with $|X|$ number of equations, followed by *(ii)* a policy improvement which requires $(m + 1)$ evaluations for each state x to yield a new table $u'(x)$ [15].

Based on the optimal policy structure discussed above, we replace the role of table $u(x)$ mapping a state to a price by the functional approximation of Equation (9), which, in fact, reduces the table $u(x)$ to the four parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4]$. Denoting a policy by the parameter vector $\boldsymbol{\theta}$, our policy iteration algorithm consists of consecutive estimates of *(i)* the average and differential costs $J_{\boldsymbol{\theta}}$ and $h_{\boldsymbol{\theta}}(x)$ for all $x \in X$, that satisfy the stationarity condition for policy $\boldsymbol{\theta}$, followed by *(ii)* a policy improvement step that revises the old policy $\boldsymbol{\theta}$ to a new policy $\boldsymbol{\theta}'$.

Step *(i)* is accomplished by solving a linear system or equivalently the following LP with $|X| + 1$ decision variables and $|X|$ constraints, where the subscript $\boldsymbol{\theta}$ denotes the policy as specified by Equation (9)

$$\begin{aligned} &\underset{J_{\boldsymbol{\theta}},\mathbf{h}_{\boldsymbol{\theta}}}{Maximize} \quad J_{\boldsymbol{\theta}} \\ &J_{\boldsymbol{\theta}}\mathbf{v} + \mathbf{h}_{\boldsymbol{\theta}} \leq \mathbf{g}(u_{\boldsymbol{\theta}}) + P(u_{\boldsymbol{\theta}})\mathbf{h}_{\boldsymbol{\theta}}. \end{aligned}$$

The LP solution for a given $\boldsymbol{\theta}$ will have all of the constraints binding, i.e., satisfying a singular linear system with $|X|$ equations but $|X| + 1$ which takes care of the singularity and provides $J_{\boldsymbol{\theta}}$ and $h_{\boldsymbol{\theta}}(x)$, $\forall x \in X$ as the primal solution. In addition, the dual variables, $p_{\boldsymbol{\theta}}(x)$ for each state x, provide the vector $p_{\boldsymbol{\theta}}$ which is the unit eigenvalue related eigenvector of the transition probability matrix $[P_{xz}(u_\theta)]$, namely the vector of steady state probabilities $p_{\boldsymbol{\theta}}(x)$ for each state x when the system is controlled by policy θ ([15]).

Step *(ii)*, however, is not as straight forward: The convergence of the policy iteration algorithm requires the average cost to decrease in each iteration, i.e., $J_{\boldsymbol{\theta}(k+1)} \leq J_{\boldsymbol{\theta}(k)}$; and the change in the cost is expressed as follows ([15]):

$$J_{\boldsymbol{\theta}(k)} - J_{\boldsymbol{\theta}(k+1)} = \sum_{\forall x \in X} p_{\boldsymbol{\theta}(k+1)}(x) \delta_{\boldsymbol{\theta}(k),\boldsymbol{\theta}(k+1)}(x), \qquad (10)$$

where we define

$$\delta_{\boldsymbol{\theta}(k),\boldsymbol{\theta}(k+1)}(x) = \left[\left[g\big(x, u_{\boldsymbol{\theta}(k)}\big) + \sum_{\forall z \in X} P_{xz}\big(u_{\boldsymbol{\theta}(k)}\big) h_{\boldsymbol{\theta}(k)}(z) \right] - \left[g\big(x, u_{\boldsymbol{\theta}(k+1)}\big) + \sum_{\forall z \in Z} P_{xz}\big(u_{\boldsymbol{\theta}(k+1)}\big) h_{\boldsymbol{\theta}(k)}(z) \right]\right]. \qquad (11)$$

Therefore, in vector notation, we need $p'_{\boldsymbol{\theta}(k+1)} \delta_{(\boldsymbol{\theta}(k),\boldsymbol{\theta}(k+1))} \geq 0$. This non-negativity condition is trivially achievable in standard policy iteration, where the ability to select an improved policy $u_{k+1}(x)$ for each state i guarantees the non-negativity of each term in the weighted average above, namely, $\delta_{u_k,u_{k+1}}(x) \geq 0 \; \forall x \in X$. In our $\boldsymbol{\theta}$ parameter based functional approximation of the policy, however, there is no obvious way to guarantee non-negativity. Were we able to implement it, an intuitive policy parameter vector improvement would be to solve:

$$\boldsymbol{\theta}(k+1) = argmax \qquad p'_{\boldsymbol{\theta}(k+1)} \delta_{(\boldsymbol{\theta}(k),\boldsymbol{\theta}(k+1))}. \qquad (12)$$

The problem with the optimization problem in (12) is that $p_{\boldsymbol{\theta}(k+1)}$ is not known until $\boldsymbol{\theta}(k+1)$ is obtained and step *(i)* of the policy iteration is repeated. For this reason, we propose to approximate $p_{\boldsymbol{\theta}(k+1)}$ by $p_{\boldsymbol{\theta}(k)}$ while imposing a constraint in the optimization problem that forces $\boldsymbol{\theta}(k+1)$ to be close to $\boldsymbol{\theta}(k)$. Our implementable policy improvement step *(ii)* is thus:

$$\max_{\boldsymbol{\theta}(k+1) \in [\boldsymbol{\theta}(k) - \Delta\boldsymbol{\theta}, \boldsymbol{\theta}(k) + \Delta\boldsymbol{\theta}]} \qquad p'_{\boldsymbol{\theta}(k)} \delta_{(\boldsymbol{\theta}(k),\boldsymbol{\theta}(k+1))} \qquad (13)$$

with $\Delta\boldsymbol{\theta}$ having all non-negative entries. When the elements of $\Delta\boldsymbol{\theta}$ are small enough, optimization problem (13) satisfies the non-negativity convergence requirement since,

$$p_{\boldsymbol{\theta}(k)} \approx p_{\boldsymbol{\theta}(k+1)}, \qquad \text{follows from } \boldsymbol{\theta}(k) \approx \boldsymbol{\theta}(k+1). \quad (14)$$

In the iterative implementing of optimization problem (13), the four dimensional vector $\Delta\boldsymbol{\theta}$ is updated adaptively as follows: When the condition $J_{\boldsymbol{\theta}(k+1)} \leq J_{\boldsymbol{\theta}(k)}$ is violated, $\Delta\boldsymbol{\theta}$ is multiplied by ρ, $0 < \rho < 1$ and the policy improvement step repeated, otherwise $\Delta\boldsymbol{\theta}$ is multiplied by $1/\rho$. Numerical experience shows that, most of the time, entries of $\Delta\boldsymbol{\theta}$ do not have to be very small for the non-negativity condition to be satisfied and the policy iteration to converge.

VI. Numerical Results

In this section, we *(i)* discuss the adequacy of the Two-Dimensional Markov Chain (TDMC) representation of the RS signal, *(ii)* use an 11 point discretization of the control set to obtain the LP-based optimal solution and use it to observe the structure of the optimal policy and verify the price variance-utility loss properties derived in Section IV, and finally *(iii)* we simulate the performance of the discretized control set optimal policy, solve the sigmoid function based API policy, and show that it compares favorably to the LP-based optimal policy from both the accuracy and computational efficiency points of view.

TABLE I
Comparison of the Actual PJM Data and the TDMC Model on Time Frequencies of RS Signals

	$[-1,-0.5)$	$[-0.5,-0)$	$[0,0.5)$	$[0.5,1]$
PJM Data	% 9	% 39	% 41	% 11
TDMC Model	% 8	% 39	% 44	% 9

TABLE II
Comparison of the Actual Data and the TDMC Model on RS Signal Statistics

	$\mathbb{E}[y]$	σ_y^2	$Pr\{d=-1\}$	$Pr\{d=+1\}$
Data	0	0.158	0.5	0.5
Model	0	0.136	0.5	0.5

The following input is used in the base case problem definition of this section: $A = 50$ kW, $R = 30$ kW, $r = 1$ kW, $\Delta\bar{y} = 1/30$, $\lambda_M = 150$ /min, $\mu = 1$ /min, $\kappa = 100$ ¢/ kW^2, $U_M = 50$¢, $m = 10$, $\Delta t = 1/15$ min. We note that the objective function provided in (5) is scaled by $60/\Delta t$ in order to obtain the optimal cost for an hour rather than Δt minutes. Therefore, κ and U_M values given here correspond to the cost coefficient and the utility bound for an hour.

A. Model for RS Signal Dynamics

Since the TDMC model of RS signal dynamics is a key driver of the optimal solution, we placed particular attention to its calibration and tested its fidelity against historical data. The range of y was discretized by $\Delta\bar{y} = 1/30$, requiring 60 possible RS signal values for each direction. The transition probabilities were estimated from actual historical data of PJM's "Fast Response Regulation Signal", available in [14] and shown in Figure 1. The historical data was first mapped to the 120 discretized (y, d) states and a frequency distribution was estimated. The transition frequencies were then used to estimate the TDMC transition probabilities.

Table I compares historical data to simulated frequencies of $y(t)$ in four broad ranges. Finer range comparisons, not shown due to space limitations, proved quite accurate as well. Table II compares historical data and model simulated $\mathbb{E}[y]$, σ_y^2 estimates as well as the proportions of time $y(t)$ is in an up or down sub-trajectory. As such, our observations convinced us of the model's fidelity.

Here, it should be also mentioned that although ISOs aim to achieve energy neutrality by deploying tertiary reserves on an hourly basis, and the long run average of y is approximately zero, the past PJM data shows that energy neutrality does not always hold in a given hour. In fact, a two-sided t-test for hourly energy neutrality on the PJM RS signal that belongs to 10/31/11 reveals a p-value of 0.0107, which would imply higher tracking costs in practice than expected.

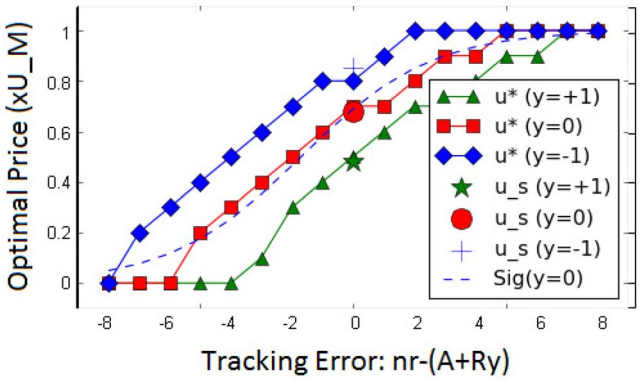


Fig. 3. Optimal Price Policy for $d = +1$.

TABLE III
LOW SENSITIVITY OF AVERAGE CONSUMPTION TO DYNAMIC OPTIMAL PRICES

Case	A	λ_M	$\bar{u}^*$	$\bar{\lambda}^*$	μ	$\bar{\lambda}^* r/\mu$
A1	40	150	$0.742U_M$	40.4	1	40.4
A2	40	150	$0.446U_M$	83	2	41.5
A3	50	90	$0.442U_M$	50.2	1	50.2
A4	50	100	$0.497U_M$	50.3	1	50.3
A5	50	150	$0.483U_M$	77.6	1.5	51.7

B. Structure of the Optimal Price Policy

Figure 3 shows the optimal prices ($u \in \{0, 0.1U_M, \ldots, U_M\}$, vertical axis) for different tracking error values ($nr-(A+Ry)$, horizontal axis), for $y = \{-1, 0, 1\}$ and $d = +1$. The optimal price policy is non-decreasing in the tracking error, and $u^* \approx u_s$ when the tracking error close to zero. Recall that u_s is the price that guides the system to $n = (A + Ry)/r$ in expectation in steady state when it is broadcast over a long enough period (see Section V-A).

In Figure 3, the shape of the curve that represents the optimal price policies for fixed y and d is similar to that of a sigmoid function. Moreover, increasing values of y shift the sigmoid-like optimal policy to the right, which results in a lower optimal price for a given tracking error. This observation motivated the approximation of optimal prices by an analytic sigmoid function whose parameters are selected by our API algorithm in Section VI-F.

Table III presents the observed time averages of optimal prices, $\bar{u}^*$, and connection rates ($\bar{\lambda}^*$) for different problem instances. As suggested in Corollary 1, the average prices satisfy $A = \lambda_M(1 - \bar{u}^*/U_M)r/\mu = \bar{\lambda}^* r/\mu$. Note that the optimal policy elicits average consumption that is slightly higher than A, a result indicating that the tracking error versus utility loss tradeoff weighs towards a slightly higher tracking error cost for the benefit of lower utility loss.

C. Price Variance-Utility Loss Relationship

Proposition 2 claims that the Utility Loss due to varying dynamic optimal prices equals $\lambda_M \sigma_u^2/(2U_M)$. Table IV results validate this relationship for different problem instances with different R values. The theoretical values of utility loss match with the experimental results obtained using optimal LP dual variables as explained in Section V-A. It is noteworthy that the price variance and associated utility loss increase with the RS reserve level. The SBO should consider this relationship in deciding its preferred RS reserve offer in the Hour Ahead market.

TABLE IV
COMPARISON OF THEORETICAL AND EXPERIMENTAL UTILITY LOSS (T.U.L. AND E.U.L., RESPECTIVELY) PER HOUR FOR DIFFERENT VALUES OF RS PROVISION

Case	σ_u^2	T.U.L. (¢)	E.U.L. (¢)
$R = 0.1$	107.57	161.35	161.34
$R = 5$	108	161.99	161.99
$R = 10$	110.35	165.53	165.52
$R = 15$	114.47	171.7	171.67
$R = 20$	119	178.5	178.49
$R = 25$	125.48	188.22	188.21
$R = 30$	133.74	200.61	200.88

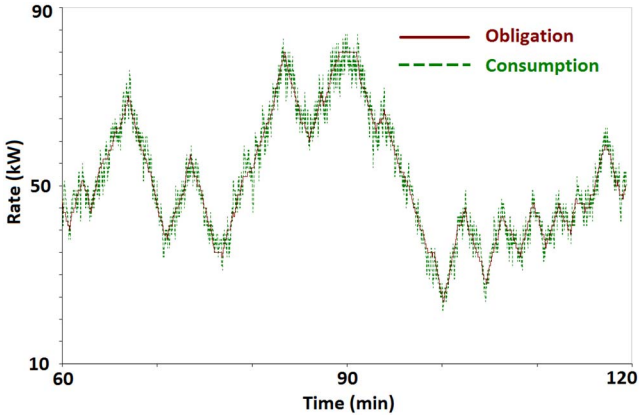


Fig. 4. Tracking Error under the Optimal Price Policy.

D. Simulated Consumption Trajectory Under the Optimal Price Policy

The objective of the optimal policy is to assist the SBO to track the RS signal the ISO broadcasts. Figure 4 displays a two hour building consumption trajectory simulated under the optimal policy. For given appliance consumption characteristics and the RS signal dynamics, Figure 4 shows that the SBO is able to successfully track the RS signal implied obligation plotted in addition to the actual consumption (the dotted green line shows the consumption rate $n(t)r$ of the building at time t, while the continuous red line shows the obligation $A + Ry(t)$). Since the building modeled by the input data is relatively small (i.e., small $\bar{\lambda}$), there is a large coefficient of variation in the associated $M/M/\infty$ queue, and the fluctuations of the consumption rate about the obligation trajectory are noticeable. The average tracking error, $(\mathbb{E}[|nr - (A + Ry)|])$, is equal to 2.1 kW, which is 7 % of the RS provision. As Section VI-A proves the success of the TDMC in modeling the RS signal behavior, in this simulation, a Markov chain model that is calibrated with 21,600 data points (24-hour) is used instead of the historical data.

TABLE V
GENERALIZATION OF THE PRICE POLICY TO MULTIPLE APPLIANCE CLASSES

Case	Class (i)	λ_i^M	r_i	μ_i	ϵ_T/R
C1	1	150	1	1	0.07
C2	1	50	1	1	0.13
	2	25	2	1	
	3	10	5	1	
C2	1	40	1	1	0.106
	2	100	1	2	
	3	15	2	1/2	

TABLE VI
INVARIANCE OF THE RELATIVE ABSOLUTE TRACKING ERROR TO THE APPLIANCE UTILITY DISTRIBUTION

Distribution	ϵ_T/R
Uniform(0,U_M)	0.07
Triangular(0,$U_M/2$, U_M)	0.07
Normal($U_M/2$, $U_M^2/25$)	0.07
Exponential($2/U_M$)	0.08

E. Experimenting Generalized Problem Settings

In the analytical derivations and experiments above, we considered a single appliance class with a utility distribution that is uniform over $[0, U^M]$. In reality, however, appliances are clustered in more than one classes. In this section, we numerically explore how the policies obtained under those assumptions can be extended to more general cases. First, we start with experimenting the cases where there are multiple appliance classes. To this end, we use the index j for the class specific values of the problem parameters, such as λ_j^M, r_j, μ_j, u_j. The optimal policies used in VI-D are obtained for an appliance class with $r = 1$ kW and $\mu = 1$/min. As the quantity r/μ specifies the expected energy consumption for an appliance per connection, we consider the optimal prices obtained for this base case as "unit" prices, u_b, and obtain class specific prices as $u_j = u_b r_j/\mu_j$. To measure the impact of this generalization, we compare the average values of ϵ_T/R over the course of a 24-hour simulation for various cases in Table V as the total consumption capacity, $\sum_i \lambda_i^M r_i/\mu_i$, is the same for all cases. Note that ϵ_T is the absolute tracking error as defined above. The results show that, although less successfully, the SBO is able to track the RS signal in the presence of multiple appliance classes and the ϵ_T/R ratio is below 13% in all cases.

Secondly, in Table VI, we show how the relative absolute tracking error, ϵ_T/R, differs when Assumptions 2 and 3 are relaxed, namely when non-uniform distributions are used to model the appliance utility. Numerical results show that the result is almost invariant to the distribution of the utility and the SBO tracks the RS signal as successfully as when the uniform distribution is used.

F. Sigmoid Function Approximation to Optimal Price Policy and Implementation of API Algorithm

Figure 3 motivates a continuous functional approximation of the optimal price policy by a sigmoid function whose parameters are optimized by the API algorithm of Section V-B. The value of this approximation relies on: *(i)* The effectiveness of the API algorithm in converging to the optimal parameter values, and *(ii)* the validity of the sigmoid function in representing the optimal price policy. We were able to verify the former by least squares fitting of a sigmoid function to the optimal discrete prices obtained from the LP solution approach. To investigate the latter, optimal objective function values of the two methods are compared in Table VII. As this comparison shows, not only does the API algorithm give close results to the LP solution, but the objective values obtained with the API algorithm generally dominate the LP solution for small discretization accuracy, i.e., $m = 5$.

TABLE VII
COMPARISON OF LP AND API SOLUTIONS

Case	Solution Time			J_{LP}		
	LP		API	$m = 5$	$m = 10$	J_{API}
	$m = 5$	$m = 10$				
B1	1h51m	3h06m	41m	-349	-349.8	-351.2
B2	1h48m	2h24m	25m	-305.2	-307	-306.5
B3	44m	3h16m	31m	-193.4	-196.1	-193.2
B4	1h38m	3h55m	28m	11.7	11.4	11.7

In API algorithm, the initial parameters are simply chosen as $\boldsymbol{\theta}(0) = [-1, 1, 0, 0]$ based on the observation that the optimal price is non-decreasing in the tracking error and non-increasing in the RS signal value, as per Figure 3. We also observe that the LP solution computational effort is particularly long because the transition probabilities over $\Delta t = 4$ sec result in dense constraints. Moreover, as we expect, increasing the discretization accuracy by using higher m values significantly increase the solution time. On the other hand, in the API algorithm, less than half of the total computation time is used in fine tuning around a small convergence tolerance, and the solution time can be reduced significantly by selecting large ϵ values without deviating from the optimal objective value significantly. The results reported in Table VII have been obtained with the following input: κ values are equal to 1, 5, 20 and 1 in cases B1,B2,B3 and B4, respectively, where $A = 50$ and $\lambda_M = 100$ for the cases B1,B2,B3. Case *B*4 uses $A = 100$ and $\lambda_M = 150$, and the contribution of the utility in the objective function is neglected in this case. In all cases, the API algorithm employed $\rho = 0.5$ and convergence tolerance $\epsilon = 0.1$.

VII. CONCLUSION

We have shown that optimal SBO managed policies can be obtained efficiently and implemented for effective provision of RS reserves by flexible loads. To this end, we investigated the structure and properties of the underlying Stochastic DP problem and proposed an efficient Approximate Policy Iteration algorithm to solve it, which we also extended to multiple appliance classes and non-uniform utility distributions. Exploring the impact of removing the assumptions on the optimal price policy analytically is a vital component to explore in future studies.

APPENDIX

A. Proof of Proposition 1

Proof: The expected number of connections during $[k\Delta t, (k+1)\Delta t]$ is $\lambda_M(1-u(k)/U_M)\Delta t$, and the expected energy consumption associated with these connections is $\Delta t\lambda_M(1-u(k)/U_M)r/\mu$, since r is the consumption rate of an active appliance and an appliance is expected to stay active for $1/\mu$ amount of time. Then, the expected average consumption committed over $[0,T]$, denoted by $\mathbb{E}[C_{[0,T]}]$, is

$$\mathbb{E}[C_{[0,T]}] = \frac{1}{T}\sum_{k=0}^{K-1}\Delta t\frac{\lambda_M r}{\mu}\left(1-\frac{u(k)}{U_M}\right) = \frac{\lambda_M r}{\mu}\left(1-\frac{\frac{1}{T}\sum_{k=0}^{K-1}u(k)\Delta t}{U_M}\right) \quad (15)$$

The result follows when $\frac{1}{T}\sum_{k=0}^{K-1}u(k)\Delta t = u_c$. Note that this proposition also holds when the price u updates are made in continuous time. This is evident when we consider $\Delta t \to 0$. ■

B. Proof of Corollary 1

Proof: For large κ, the tracking error penalty dominates the utility loss, and, for large enough λ_M and Assumption 1, the optimal price policy u^* results in average consumption rate of A kW. By Proposition 1 the dynamic pricing policy u^* and the static price policy $\bar{u}^*$ result in the same average consumption. However, under static price policy $\bar{u}^*$, $n(t)$ is equivalent to an $M/M/\infty$ queue which is Poisson distributed with mean $\bar{\lambda}^*/\mu$ implying average consumption $\bar{\lambda}^* r/\mu = A$. ■

C. Proof of Proposition 2

Proof: We express the total utility realized over $[0,T]$ under a dynamic price policy as the summation of the utilities realized over the disjoint sub-intervals, i.e.,

$$\mathbb{E}\left[\Phi_u^{[0,T]}\right] = \mathbb{E}\left[\sum_{k=0}^{K-1}\Phi_u^{[k\Delta t,(k+1)\Delta t]}\right] \quad (16)$$

$$= \sum_{k=0}^{K-1}\mathbb{E}\left[\Phi_u^{[k\Delta t,(k+1)\Delta t]}\right] \quad (17)$$

where the second equality follows from the fact that the Poisson connections in each sub-interval are independent from each other. The expected number of connections during $[k\Delta t, (k+1)\Delta t]$ is $\lambda_M(1-u(k)/U_M)\Delta t$. An appliance that connects at time $t \in [k\Delta t, (k+1)\Delta t]$ realizes a utility ϕ, where $\mathbb{E}[\phi|\phi \geq u(k)] = (u(k)+U_M)/2$. Then the expected utility realized over $[k\Delta t, (k+1)\Delta t]$ is given by

$$\mathbb{E}\left[\Phi_u^{[k\Delta t,(k+1)\Delta t]}\right] = \Delta t\lambda_M(1-u(k)/U_M)(u(k)+U_M)/2. \quad (18)$$

Therefore,

$$\mathbb{E}\left[\Phi_u^{[0,T]}\right] = \sum_{k=0}^{K-1}\Delta t\lambda_M\left(1-\frac{u(k)}{U_M}\right)\left(\frac{u(k)+U_M}{2}\right) \quad (19)$$

where $T = K\Delta t$. Letting $\epsilon(k) = u(k) - u_c$ during $[k\Delta t, (k+1)\Delta t]$, i.e., $u(k) = u_c + \epsilon(k)$ and therefore $\sum_{k=0}^{K-1}\epsilon(k) = 0$, Equation 19 can be written as

$$\begin{aligned}\mathbb{E}\left[\Phi_u^{[0,T]}\right] &= \lambda_M T\left(\frac{(U_M)^2-(u_c)^2}{2U_M} - \frac{\sum_{k=0}^{K-1}(\epsilon(k))^2\Delta t}{2U_M T}\right)\\ &= \frac{\lambda_M T}{2U_M}\left((U_M)^2-(u_c)^2\right) - \frac{\lambda_M T}{2U_M}\sum_{k=0}^{K-1}\frac{(\epsilon(k))^2\Delta t}{T}\\ &= \mathbb{E}\left[\Phi_{u_c}^{[0,T]}\right] - \frac{\lambda_M T}{2U_M}\sum_{k=0}^{K-1}\frac{(\epsilon(k))^2\Delta t}{T}. \end{aligned} \quad (20)$$

The first component of the expression above is the expected utility that would be realized under a static price policy u_c over $[0,T]$, while the second term, which we call *Utility Loss* represents a decrease in the static price utility whose magnitude increases in proportion to the time average of $(\epsilon(k))^2$. Since $\epsilon(k) = u(k) - u_c$, the term with the summation implies the variance of $u(k)$ over the trajectory $k = 0, 1, \ldots, K-1$, σ_u^2. Therefore, we can express the Utility Loss per unit time as

$$\frac{1}{T}\left(\mathbb{E}\left[\Phi_{u_c}^{[0,T]}\right] - \mathbb{E}\left[\Phi_u^{[0,T]}\right]\right) = \frac{\lambda_M\sigma_u^2}{2U_M}. \quad (21)$$

■

REFERENCES

[1] S. G. Whitley, "New challenges facing system operators," in *Proc. Cornell Univ. Workshop Past Present Future Power Grid*, New York, NY, USA, Aug. 2012. [Online]. Available: http://www.nyiso.com/public/webdocs/media_room/publications_presentations/NYISO_Presentations/NYISO_Presentations/New_Challenges_Facing_System_Operators-_S_Whitley_08_09_12.pdf

[2] Y. V. Makarov, C. Loutan, J. Ma, and P. de Mello, "Operational impacts of wind generation on California power systems," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 1039–1050, May 2009.

[3] G. Gownisankaran, S. S. Reynolds, and M. Samano, "Intermittency and the value of renewable energy," Nat. Bur. Econ. Res., Cambridge, MA, USA, Tech. Rep. 17086, May 2001.

[4] K. Porter *et al.*, "Review of industry practice and experience in the integration of wind and solar generation," GE Energy, Atlanta, GA, USA, Nov. 2012. [Online]. Available: http://bit.ly/21b6oE5

[5] ISONE. (2014). *Website*. [Online]. Available: http://www.iso-ne.com

[6] E. Bilgin and M. C. Caramanis, "Decision support for offering load-side regulation service reserves in competitive power markets," in *Proc. 52nd IEEE Conf. Decis. Control*, Florence, Italy, Dec. 2013, pp. 5628–5635.

[7] F. C. Schweppe, "Power: Power systems '2000': Hierarchical control strategies: Multilevel controls and home minis will enable utilities to buy and sell power at 'real time' rates determined by supply and demand," *IEEE Spectr.*, vol. 15, no. 7, pp. 42–47, Jul. 1978.

[8] A. W. Berger and F. C. Schweppe, "Real time pricing to assist in load frequency control," *IEEE Trans. Power Syst.*, vol. 4, no. 3, pp. 920–926, Aug. 1989.

[9] N. Lu and D. P. Chassin, "A state-queueing model of thermostatically controlled appliances," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1666–1673, Aug. 2004.

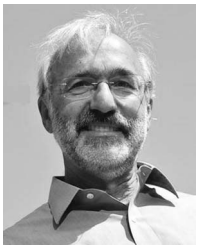
[10] M. D. Galus, S. Koch, and G. Andersson, "Provision of load frequency control by PHEVs, controllable loads, and a cogeneration unit," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4568–4582, Oct. 2011.

[11] I. C. Paschalidis, B. Li, and M. C. Caramanis, "Demand-side management for regulation service provisioning through internal pricing," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1531–1539, Aug. 2012.

[12] M. Caramanis, I. C. Paschalidis, C. Cassandras, E. Bilgin, and E. Ntakou, "Provision of regulation service reserves by flexible distributed loads," in *Proc. 51st IEEE Conf. Decis. Control*, Maui, HI, USA, Dec. 2012, pp. 3694–3700.

[13] I. C. Paschalidis and J. N. Tsitsiklis, "Congestion dependent pricing of network services," *IEEE ACM Trans. Netw.*, vol. 8, no. 2, pp. 171–184, Apr. 2000.
[14] PJM. (2013). *Market-Based Regulation*. [Online]. Available: http://pjm.com/markets-and-operations/ancillary-services/mkt-based-regulation.aspx
[15] D. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 2, 3rd ed. Belmont, MA, USA: Athena Scientific, 1995.

Enes Bilgin received the B.S. degree from Bilkent University in 2010, and the M.S. and Ph.D. degrees from Boston University in 2012 and 2014, respectively. He is currently an SMTS Operations Research Scientist with Advanced Micro Devices, Inc., and an Adjunct Faculty with the Ingram School of Engineering, Texas State University. His research interests include demand response in smart grids, Markov decision processes, and supply chain optimization in automated humanitarian missions. He was a recipient of numerous awards, including an Honorable Mention in the IBM/IEEE Smarter Planet Challenge.



Michael C. Caramanis received the B.S. degree from Stanford University in 1971, and the M.S. and Ph.D. degrees from Harvard University in 1972 and 1976, respectively. Since 1982, he has been a Mechanical Engineer with Boston University, where he is a Professor of Systems. He was the Chair of the Greek Regulatory Authority for Energy and the International Energy Charter's Investment Group from 2004 to 2008. He was active in power market implementations in England and Italy. He has co-authored the book *Spot Pricing of Electricity* (Kluwer, 1987) and 100+ refereed publications. His disciplinary background is in mathematical economics, optimization, and stochastic dynamic decision making. His recent application domain focus is marginal costing and dynamic pricing on smart power grids, grid topology control for congestion mitigation, and the extension of power markets to include distribution connected loads, generation, and resources.

Ioannis Ch. Paschalidis (M'96–SM'06–F'14) received the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1993 and 1996, respectively. In 1996, he joined Boston University, where he has been ever since. He is a Professor and the Distinguished Faculty Fellow with the Department of Electrical and Computer Engineering, the Division of Systems Engineering, and the Department of Biomedical Engineering, Boston University. He is the Director of the Center for Information and Systems Engineering. He has held visiting appointments with MIT and Columbia University, New York, NY, USA. His current research interests lie in the fields of systems and control, networking, applied probability, optimization, operations research, computational biology, and medical informatics.

Christos G. Cassandras received the B.S. degree from Yale University in 1977, the M.S.E.E. degree from Stanford University in 1978, and the S.M. and Ph.D. degrees from Harvard University in 1979 and 1982, respectively. From 1982 to 1984, he was with ITP Boston, Inc., where he worked on the design of automated manufacturing systems. From 1984 to 1996, he was a Faculty Member with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst. He is currently a Distinguished Professor of Engineering, the Head of the Division of Systems Engineering, and a Professor of Electrical and Computer Engineering with Boston University. He specializes in the areas of discrete event and hybrid systems, cooperative control, stochastic optimization, and computer simulation with applications to computer and sensor networks, manufacturing systems, and transportation systems. He has published over 350 papers in the above areas and five books. He was a recipient of several awards, including the 2011 IEEE Control Systems Technology Award, the Distinguished Member Award of the IEEE Control Systems Society (2006), the 1999 Harold Chestnut Prize (IFAC Best Control Engineering Textbook), the IBM/IEEE Smarter Planet Challenge Prize in 2011 and 2014, the 2012 Kern Fellowship, and the 1991 Lilly Fellowship. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL from 1998 to 2009, and has served on several editorial boards and as a Guest Editor for various journals. He was the 2012 President of the IEEE Control Systems Society. He is a Member of Phi Beta Kappa and Tau Beta Pi, and a Fellow of the IFAC.