Dynamic Stochastic Demand Response With Energy Storage

Yuanzhang Xiao, *Member, IEEE,* and Mihaela van der Schaar, *Fellow, IEEE*

***Abstract*—We consider a power system with an independent system operator (ISO), and distributed aggregators who have energy storage and purchase energy from the ISO to serve their customers. All the entities in the system are *foresighted*: each aggregator seeks to minimize its own *long-term* payments for energy purchase and operational costs of energy storage by deciding how much energy to buy from the ISO, and the ISO seeks to minimize the *long-term* total cost of the system (e.g. energy generation costs and the aggregators' costs) by dispatching the energy production among the generators. The decision making of the foresighted entities is complicated for two reasons. First, the information is decentralized among the entities, namely each entity does not know the others' states. Second, an aggregator's current decision affects its future costs due to the coupling introduced by the energy storage. We propose a design framework in which the ISO provides each aggregator with a conjectured future price, and each aggregator distributively minimizes its own long-term cost based on its conjectured price as well as its locally-available information. We prove that the proposed framework can achieve the social optimum despite being decentralized and involving complex coupling. Simulation results show that the proposed foresighted demand side management achieves significant reduction in the total cost, compared to the optimal myopic demand side management (up to 60% reduction), and the foresighted demand side management based on the Lyapunov optimization framework (up to 30% reduction).**

I. Introduction

Power systems are currently undergoing drastic changes on both supply and demand sides. On the supply side, renewable energy (e.g., wind energy, solar energy) is increasingly used to make the power systems "greener"; however, this introduces high uncertainty in the energy generation. Hence, on the demand side, energy storage is increasingly used as an important solution to cope with this uncertainty [8].

We consider a power system consisting of several heterogeneous energy generators on the supply side, an independent system operator (ISO) that operates the system, and multiple heterogeneous aggregators and their customers on the demand side. The ISO receives energy purchase requests from the aggregators as well as reports of (parameterized) energy generation cost functions from the generators, and based on these, dispatches the energy generators and determines the unit energy prices. The aggregators, located in different geographical areas, determine how much to buy from the grid (and how much to store in their storage) to provide energy for their customers (e.g., households, office buildings).

Y. Xiao is with Department of Electrical Engineering and Computer Science, Northwestern University. Email: xyz.xiao@gmail.com.

M. van der Schaar is with Department of Electrical Engineering, University of California Los Angeles. Email: mihaela@ee.ucla.edu.

The key feature that sets apart our paper from existing works [1]–[10] is that there are *multiple foresighted* decision makers in the system. In contrast to the works that study the decision problems of a *single* aggregator [1]–[6], we consider the decision problems of *the ISO and multiple aggregators*, whose decisions impact the others' costs. In contrast to the works that assume the aggregators to be *myopic* and to minimize their *short-term* (e.g., one-day or hourly) costs [7]–[10], we allow the aggregators to minimize their *long-term* costs. As a result, our proposed solution achieves lower long-term average costs by taking into account the externality among the decision makers and by allowing foresighted demand-side management (DSM) strateiges.

The challenges of our problem result from the presence of *multiple foresighted* decision makers and the decentralized information among the decision makers. The total cost depends on the generation cost functions (e.g. the speed of wind for wind energy generation, the amount of sunshine for solar energy generation, and so on), the status of the transmission lines (e.g. the flow capacity of the transmission lines), the amount of electricity in the energy storage, and the demand from the customers, all of which change over time due to supply and demand uncertainty. However, no entity knows all the above information: the ISO knows only the generation cost functions and the status of the transmission lines, and each aggregator knows only the status of its own energy storage and the demand of its own customers. Hence, the DSM strategy needs to be decentralized, such that each entity can make decisions solely based on its *locally-available* information[1].

To overcome the challenges due to information decentralization, we propose a decentralized DSM strategy based on conjectured prices. Specifically, each aggregator makes decisions based on its conjectured price, and its local information about the status of its energy storage and the demand from its customers. In other words, each aggregator summarizes all the information that is not available to it into its conjectured price. Note, however, that the price is determined based on the generation cost functions and the status of the transmission lines, which is only known to the ISO. Hence, the aggregators' conjectured prices are determined by the ISO. We propose a simple online algorithm for the ISO to update the conjectured prices based on its local information, and prove that by using

[1]Even if the aggregators are willing to share all their private information with the ISO such that the ISO can make centralized decisions, the resulting centralized decision problem becomes intractable quickly as the size of the power network grows (e.g. for the IEEE 118-bus system). For large power networks, it is not only desirable, but also necessary to have an decentralized solution in which each entity is able to solve one subproblem after decomposing the intractable centralized problem.

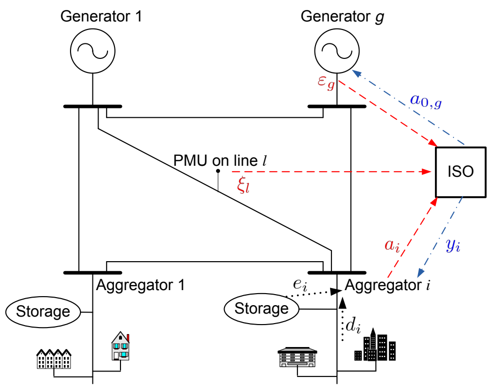

Fig. 1. The system model of the smart grid. The information flow to the ISO is denoted by red dashed lines, the information flow to the aggregators is denoted by black dotted lines, and the information flow sent from the ISO is denoted by blue dash-dot lines.

the algorithm, the ISO obtains the optimal conjectured prices under which the aggregators' (foresighted) best responses minimize the total cost of the system. In addition, we propose an online learning algorithm for the entities to learn and converge to the optimal DSM strategy when the system dynamics are unknown. The proposed online learning algorithm utilizes the concept of post-decision states [11][12] and exploits the independence of the state dynamics in the system, which results in a faster learning speed and a better run-time performance compared to conventional learning algorithms such as Q-learning [13].

The rest of the paper is organized as follows. We introduce the system model in Section II, and then formulate the design problem in Section III. We describe the proposed optimal decentralized DSM strategy in Section IV. Through simulations, we demonstrate the performance gains of the proposed strategy in Section V. Finally, we conclude the paper in Section VI.

II. System Model

We consider a real-time electricity market with one ISO indexed by 0, G heterogeneous generators indexed by g, I heterogeneous aggregators indexed by i, and L transmission lines (see Fig. 1 for an illustration). The ISO schedules the energy generation of generators and determines the unit prices of energy for the aggregators. The generators provide the ISO with the information of their energy generation cost functions, based on which the ISO can minimize the total cost of the system. Each aggregator, equipped with energy storage, provides energy for its customers (e.g. residential households, commercial buildings), and determines how much energy to buy from the ISO. In summary, the decision makers (or the entities) in the system are the ISO and the I aggregators. We denote the set of aggregators by $\mathcal{I} = \{1, \ldots, I\}$. In the following, we refer to the ISO or an aggregator generally as entity $i \in \{0\} \cup \mathcal{I}$, with entity 0 being the ISO and entity $i \in \mathcal{I}$ being aggregator i.

A. The System Setup

As discussed before, different entities possess different local information. Specifically, the ISO receives reports of the energy generation cost functions, denoted by $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_G)$, from the generators, and measures the status of the transmission lines such as the phases or capacities, denoted by $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_L)$, by using the phasor measurement units (PMUs). We summarize the energy generation cost functions and the status of the transmission lines into the ISO's state $s_0 = (\boldsymbol{\varepsilon}, \boldsymbol{\xi}) \in S_0$, which is unknown to aggregators. Each aggregator receives energy consumption requests from its customers, and manages its energy storage. We summarize the aggregate demand d_i from aggregator i's customers and the amount e_i of energy in aggregator i's storage into its state $s_i = (d_i, e_i) \in S_i$, which is only known to aggregator i.

Physically, the demand and the storage value are continuous variables. In practice, however, they are usually quantized and represented as discrete variables. For the demand, the aggregators usually quantize it before submitting its demand bids to the ISO. This is because it is impossible to submit an arbitrary real number as the demand bid (due to communication constraint). For example, the demand bids in the California ISO (CAISO) dataset are multiples of 10kW (see Reference [17]). For the storage value, the aggregators need to quantize it if they use any digital system to manage their storages. Therefore, we assume that all the sets $S_0, \ldots, S_I$ of states are finite.

The ISO's action is how much energy each generator should produce, denoted by $a_0 \in A_0(s_0) \subset \mathbb{R}_+^G$, where $A_0(s_0)$ is the action set under state s_0. In practice, most aggregators submit inelastic/self-schedule demand bids (i.e., how much energy to procure), instead of elastic/economic bids (i.e, how much energy to procure and the price they are willing to pay), to the ISO [16]. Therefore, it is reasonable to assume that all the aggregators submit inelastic bids. Denote aggregator i's action, namely the amount of energy to purchse, by $a_i \in A_i(s_i) \subset \mathbb{R}_+$, where $A_i(s_i)$ is the action set under state s_i. We denote the joint action profile of the aggregators as $\boldsymbol{a} = (a_1, \ldots, a_I)$, and the joint action profile of all the aggregators other than i as $\boldsymbol{a}_{-i}$. We allow the action set to be dependent on the current state, in order to impose constraints on each entity's behavior. For example, we require that the aggregator must fulfill its customers' demand. Hence, given aggregator i's state $s_i = (d_i, e_i)$, we have

$$A_i(s_i) = \{a_i : d_i \leq a_i + e_i \leq E_i\}, \tag{1}$$

where E_i is the maximum capacity of aggregator i's storage. We could also impose constraints on the charging/discharging rates of the storage. The aggregator charges the storage when $a_i > d_i$ and discharges the storage when $a_i < d_i$. Hence, the maximum charging/discharging rate constraint can be written as $-r_i^{discharge} \leq a_i - d_i \leq r_i^{charge}$, where $r_i^{discharge}$ and r_i^{charge} are the maximum discharging and charging rates.

We divide time into periods $t = 0, 1, 2, \ldots$, where the duration of a period is determined by how fast the demand or supply changes or how frequently the energy trading decisions are made. In each period t, the entities act as follows:

- The ISO observes its state s_0.
- Each aggregator i observes its state s_i.
- Each aggregator i chooses its action a_i, namely how much energy to purchase from the ISO, and tells its

amount a_i of energy purchase to the ISO.

- Based on its state s_0 and the aggregators' action profile $\boldsymbol{a}$, the ISO determines the price $y_i(s_0, \boldsymbol{a}) \in Y_i$ of electricity at each aggregator i, and announces it to each aggregator i. The ISO also determines its action a_0 of how much energy each generator should produce.
- Each aggregator i pays $y_i(s_0, \boldsymbol{a}) \cdot a_i$ to the ISO.

The instantaneous cost of each entity depends on its current state and its current action. Each aggregator i's total cost consists of two parts: the operational cost and the payment. Each aggregator i's operational cost $c_i : S_i \times A_i \to \mathbb{R}$ is a convex increasing function of its action a_i. An example operational cost function of an aggregator can be

$$c_i(s_i, a_i) = m_i(e_i, a_i - d_i),$$

where $m_i(e_i, a_i)$ is the maintenance cost of the energy storage that is convex [8]. It may depend on both the amount of energy in the storage and the charging/discharging rate $a_i - d_i$. Then we write each aggregator i's total cost, which is the cost aggregator i aims to minimize, as the sum of the operational cost and the payment, namely

$$\bar{c}_i = c_i + y_i(s_0, a_i, \boldsymbol{a}_{-i}) \cdot a_i.$$

Note that each aggregator's payments depends on the others' actions through the price. Although each aggregator i observes its realized price y_i, it does not know how its action a_i influences the price y_i, because the price depends on the others' actions $\boldsymbol{a}_{-i}$ and the ISO's state s_0, neither of which is known to aggregagtor i.

The energy generation cost of generator g is denoted $c_g(\varepsilon_g, a_{0,g})$, which is assumed to be convex increasing in the energy production level $a_{0,g}$. An example cost function is

$$(q_{0,g} + q_{1,g}a_{0,g} + q_{2,g}a_{0,g}^2) + q_{r,g}(a_{0,g} - a_{0,g}^-)^2, \qquad (2)$$

where $a_{0,g}^-$ is the production level in the previous time slot. In this case, the energy generation cost function of generator g is a vector $\varepsilon_g = (q_{0,g}, q_{1,g}, q_{2,g}, q_{r,g}, a_{0,g}^-)$. In the cost function, $q_{0,g} + q_{1,g} \cdot a_{0,g} + q_{2,g} \cdot a_{0,g}^2$ is the quadratic cost of producing a_0 amount of energy, and $q_{r,g} \cdot (a_{0,g} - s_{0,g})^2$ is the ramping cost of changing the energy production level. We denote the total generation cost by $c_0 = \sum_{g=1}^{G} c_g$. The ISO's cost, denoted $\bar{c}_0$, is then the sum of generation costs and the aggregators' costs, i.e. $\bar{c}_0 = \sum_{i=0}^{I} c_i$.

Note, importantly, that the example cost functions above are for illustrative purpose; we can define a variety of cost functions as long as they satisfy the convexity assumption.

Denote aggregator i's state transition probability by ρ_i, where $\rho_i(s_i'|s_i, a_i)$ is the probability that the next state is s_i', when the current state is s_i and the current action is a_i. We assume that each entity's state transition is Markovian. Note that we need this assumption for our theoretical results. As we will show in the simulations, even when the state transition is not Markovian, our proposed solution can track the nonstationary dynamics (i.e. time-varying state transitions). We also assume that conditioned on the ISO's action a_0 and the aggregators' action profile $\boldsymbol{a}$, each entity's state transition is independent of each other.

B. The DSM Strategy

At the beginning of each period t, each aggregator i chooses an action based on its state. Specifically, entity i's strategy is a mapping from its set of states to its set of actions, namely $\pi_i : S_i \to A_i$. The joint strategy profile of all the entities is written as $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_I)$. Since each entity's strategy depends only on its local information, the strategy $\boldsymbol{\pi}$ is *decentralized*.

The joint strategy profile $\boldsymbol{\pi}$ and the initial states $(s_0^0, s_1^0, \ldots, s_I^0)$ induce a probability distribution over the sequences of states and prices, and hence a probability distribution over the sequences of total costs $\bar{c}_i^0, \bar{c}_i^1, \ldots$. Taking expectation with respect to the sequences of costs, we have entity i's expected long-term cost given the initial state as

$$\bar{C}_i(\boldsymbol{\pi}|(s_0^0, s_1^0, \ldots, s_I^0)) = \mathbb{E}\left\{(1-\delta)\sum_{t=0}^{\infty}\left(\delta^t \cdot \bar{c}_i^t\right)\right\}, \qquad (3)$$

where $\delta \in [0, 1)$ is the discount factor. The discount factor is commonly used in dynamic programming (e.g., see References [6], [11]-[13]). It can be interpreted as the discounting of future cost. Specifically, the cost incurred in the future is discounted (i.e., lowered) when we add it in the total cost.

III. The Design Problem

The designer, namely the ISO, aims to maximize the social welfare, namely minimize the long-term total cost in the system. In addition, we need to satisfy the constraints due to the capacity of the transmission lines, the supply-demand requirements, and so on. We denote the constraints by $\boldsymbol{f}(s_0, a_0, \boldsymbol{a}) \leq \boldsymbol{0}$, where $\boldsymbol{f}(s_0, a_0, \boldsymbol{a}) \in \mathbb{R}^N$ with N being the number of constraints. We assume that the electricity flow can be approximated by the direct current (DC) flow model, in which case the constraints $\boldsymbol{f}(s_0, a_0, \boldsymbol{a}) \leq \boldsymbol{0}$ are linear in each a_i. These constraints model the possible congestion in electricity markets (i.e., when the constraints are binding). Hence, the design problem can be formulated as

$$\begin{aligned} \min_{\boldsymbol{\pi}} \quad & \sum_{s_0^0, s_1^0, \ldots, s_I^0} \left\{\sum_{i=0}^{I} C_i(\boldsymbol{\pi}|(s_0^0, s_1^0, \ldots, s_I^0))\right\} \qquad (4) \\ s.t. \quad & \boldsymbol{f}(s_0, \pi_0(s_0), \ldots, \pi_I(s_I)) \leq \boldsymbol{0}, \ \forall (s_0, \ldots, s_I). \end{aligned}$$

Note that in the above optimization problem, we use aggregator i's cost C_i instead of its total cost $\bar{C}_i$, because its payment is transferred to the ISO and is thus canceled in the total cost. Note also that we sum up the social welfare under all the initial states. This is a common technique used in dynamic programming: instead of minimizing the social welfare at each initial state, we minimize the sum of social welfare at all initial states, which is easier to solve. The optimal stationary strategy profile that maximizes this expected social welfare will also maximize the social welfare given any initial state. We write the solution to the design problem as $\boldsymbol{\pi}^\star$ and the optimal value of the design problem as $C^\star$.

IV. Optimal Foresighted DSM

A. The aggregator's Decision Problem and Conjectured Price

Contrary to the designer, each aggregator aims to minimize its own long-term total cost $\bar{C}_i(\boldsymbol{\pi}|(s_0^0, s_1^0, \ldots, s_I^0))$. In other words, each aggregator i solves the following problem:

$$\pi_i = \arg\min_{\pi_i'} \bar{C}_i(\pi_i', \boldsymbol{\pi}_{-i} | (s_0^0, s_1^0, \ldots, s_I^0)).$$

Assuming that the aggregator knew all the information, the optimal solution to the above problem would satisfy the following:

$$V_i(s_0, s_i, \boldsymbol{s}_{-i}) = \min_{a_i \in A_i} (1-\delta)\bar{c}_i(s_0, s_i, a_i, \boldsymbol{a}_{-i})$$

$$+\delta \cdot \sum_{s_0', s_i', \boldsymbol{s}_{-i}'} \left\{ \rho_0(s_0'|s_0) \prod_{j=1}^{I} \rho_j(s_j'|s_j, a_j) V(s_0', s_i', \boldsymbol{s}_{-i}') \right\}.$$

The above equations would be the Bellman equations, if the aggregator knew all the information such as the other aggregators' strategies $\boldsymbol{\pi}_{-i}$ and states $\boldsymbol{s}_{-i}$, and the ISO's state s_0. However, such information is never known to the aggregator.

One way to decouple the interaction among the aggregators is to endow each aggregator with a conjectured price. In this paper, we adopt a simple conjecture that the price is constant, and will prove later that this simple conjecture is sufficient to achieve the social optimum. Denote the conjectured price as $\tilde{y}_i$, we can rewrite aggregator i's decision problem as

$$\begin{aligned} \tilde{V}^{\tilde{y}_i}(s_i) = \min_{a_i \in A_i} (1-\delta) \quad & [c_i(s_i, a_i) + \tilde{y}_i \cdot a_i] \\ + \quad & \delta \cdot \sum_{s_i'} \left[\rho_i(s_i'|s_i, a_i) \tilde{V}^{\tilde{y}_i}(s_i') \right]. \end{aligned} \tag{5}$$

We can see from the above equations that given the conjectured price $\tilde{y}_i$, each aggregator can make decisions based only on its local information.

In Fig. 2, we illustrate the entities' decision making and information exchange in the design framework based on conjectured prices. In the proposed design framework, the ISO sends the conjectured prices to the aggregators before the aggregators make decisions. This additional procedure of exchanging conjectured prices allows the ISO to lead the aggregators to the optimal DSM strategies. Note that the conjectured price is generally not equal to the real price charged at the end of the period, and is not equal to the expectation of the real price in the future. In this sense, the conjectured prices can be considered as control signals sent from the ISO to the aggregators, which can help the aggregators to compute the optimal strategies.

The remaining question is how to determine the optimal conjectured prices, such that when each aggregator reacts based on its conjectured price, the resulting strategy profile maximizes the social welfare.

B. The Optimal Decentralized DSM Strategy

We propose a distributed algorithm used by the ISO to iteratively update the conjectured prices and by the aggregators to update their optimal strategies. We summarize the algorithm in Table I, and prove that the algorithm can achieve the optimal social welfare in the following theorem.

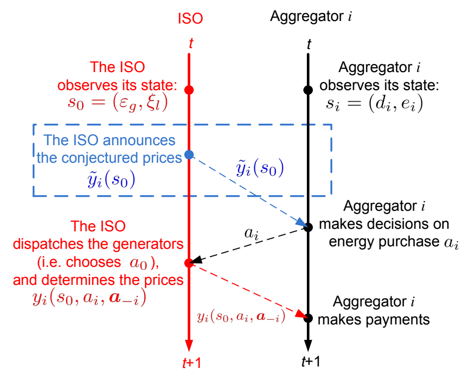

Fig. 2. Illustration of the entities' decision making and information exchange in the design framework based on conjectured prices.

TABLE I

DISTRIBUTED ALGORITHM TO COMPUTE THE OPTIMAL DECENTRALIZED DSM STRATEGY.

Input: Each entity's performance loss tolerance ϵ_i

Initialization: $k = 0$, $\lambda(0) = 0$, $\bar{a}_i(0) = 0, \forall i \in \mathcal{I}$, $\tilde{y}_i(0) = 0, \forall i \in \mathcal{I} \cup \{0\}$.

repeat

Each aggregator i solves

$$\tilde{V}_i^{\tilde{y}_i^{(k)}(s_0)}(s_i) = \min_{a_i \in A_i} \left\{ (1-\delta) \left[c_i(s_i, a_i) + \tilde{y}_i^{(k)}(s_0) \cdot a_i \right] + \delta \cdot \sum_{s_i'} \left[\rho_i(s_i'|s_i, a_i) \tilde{V}_i^{\tilde{y}_i^{(k)}(s_0)}(s_i') \right] \right\}$$

The ISO solves

$$\tilde{V}_0^{\tilde{\boldsymbol{y}}_0(k)}(s_0) = \min_{a_i \in A_i} \left\{ (1-\delta) \left[\sum_g c_g(s_0, a_0) + \tilde{\boldsymbol{y}}_0(k)^T \cdot a_0 \right] + \delta \cdot \sum_{s_0'} \left[\rho_0(s_0'|s_0, a_0) \tilde{V}_0^{\tilde{\boldsymbol{y}}_0(k)}(s_0') \right] \right\}$$

Each aggregator i reports its purchase request $\pi_i^{\tilde{y}_i^{(k)}(s_0)}(s_i)$

The ISO updates $\bar{a}_i(k+1) = \bar{a}_i(k) + \pi_i^{\tilde{y}_i^{(k)}(s_0)}(s_i)$ for all $i \in \mathcal{I}$

The ISO updates the conjectured prices:

$\tilde{y}_i^{(k+1)}(s_0) = \left(\boldsymbol{\lambda}(k+1)^T \cdot \frac{\partial f(s_0, \boldsymbol{a})}{\partial a_i} \right)^T$, where $\Delta(k) = \frac{1}{k+1}$ and

$\boldsymbol{\lambda}(k+1) = \left\{ \boldsymbol{\lambda}(k) + \Delta(k) \boldsymbol{f}\left(s_0, \pi_0^{\tilde{\boldsymbol{y}}_0(k)}(s_0), \frac{\bar{a}_1(k+1)}{k+1}, \ldots, \frac{\bar{a}_I(k+1)}{k+1} \right) \right\}^+$

until $\|\tilde{V}_i^{\tilde{y}_i^{(k+1)}(s_0)} - \tilde{V}_i^{\tilde{y}_i^{(k)}(s_0)}\| \leq \epsilon_i$

Theorem 1: The algorithm in Table I converges to the optimal strategy profile, namely

$$\lim_{k \to \infty} \left| \sum_{s_0^0, s_1^0, \ldots, s_I^0} \sum_{i=0}^{I+1} C_i(\boldsymbol{\pi}^{\tilde{y}^{(k)}} | (s_0^0, s_1^0, \ldots, s_I^0)) - C^\star \right| = 0.$$

Proof: See the appendix. ∎

Remark 1: The proposed algorithm is intended to run *only once*, before the system starts operation. The output of the proposed algorithm is a *contingent plan* of electricity purchase for each aggregator, and a *contingent plan* of electricity pricing for the ISO. Specifically, each aggregator will have a plan of how much electricity to buy from the grid, based on the status of its energy storage and its demand; the ISO will have a plan of how to price the electricity, based on the status of the generators and the grid.

Remark 2: In real time (i.e., when the system starts operation), the aggregators and the ISO can simply act according to this predetermined plan (i.e., look up what action to take at each state based on the plan). Therefore, there is no issue of computational complexity in real time. Note that the proposed DSM mechanism is adaptive, namely the decision makers will adapt their actions based on different states. The aggregators will purchase different amounts of electricity under different energy storage status and demands; the ISO will set different electricity prices under different amounts of available renewable energy.

We summarize the operations needed to implement the proposed DSM mechanism and their main features in Table II.

Finally, we can further reduce the computational complexity of implementing the algorithm in Table I by decomposing the ISO's decision problem. Note that the generators' energy generation cost functions are independent of each other. Then we have the following theorem.

Theorem 2: Given the conjectured price $\tilde{\boldsymbol{y}}_0(k)$, the ISO's value function $\tilde{V}_0^{\tilde{\boldsymbol{y}}_0(k)}$ can be calculated by $\tilde{V}_0^{\tilde{\boldsymbol{y}}_0(k)}(s_0) = \sum_{g=1}^{G} \tilde{V}_{0,g}^{\tilde{y}_{0,g}(k)}(\varepsilon_g)$, where $\tilde{V}_{0,g}^{\tilde{y}_{0,g}(k)}$ solves

$$\tilde{V}_{0,g}^{\tilde{y}_{0,g}(k)}(\varepsilon_g) = \min_{a_{0,g}}(1-\delta)\left[c_g(\varepsilon_g, a_{0,g}) + \tilde{y}_{0,g}(k)^T \cdot a_{0,g}\right] + \delta \cdot \sum_{\varepsilon_g'} \left[\rho_0(\varepsilon_g'|\varepsilon_g, a_{0,g}) \tilde{V}_{0,g}^{\tilde{y}_{0,g}(k)}(\varepsilon_g')\right].$$

Proof: The proof follows directly from Lemma 1 in the appendix. ■

From the above theorem, we know that the dimensionality of the ISO's decision problem is $\sum_{g=1}^{G} |\mathcal{E}_g|$, where $|\mathcal{E}_g|$ is the cardinality of the set of generator g's generation cost functions. The dimensionality increases linearly with the number of generators, instead of exponentially with the number of generators and transmission lines without decomposition.

C. Learning Unknown Dynamics

In practice, each entity may not know the dynamics of its own states (i.e., its own state transition probabilities) or even the set of its own states. When the state dynamics are not known a priori, each entity cannot solve their decision problems using the distributed algorithm in Table I. In this case, we can adapt the online learning algorithm based on post-decision state (PDS) in [11], which was originally proposed for wireless video transmissions, to our case.

The main idea of the PDS-based online learning is to learn the post-decision value function, instead of the normal value function. Each aggregator i's post-decision value function is defined as $U_i(\tilde{d}_i, \tilde{e}_i)$, where $(\tilde{d}_i, \tilde{e}_i)$ is the post-decision state. The difference from the normal state is that the PDS $(\tilde{d}_i, \tilde{e}_i)$ describes the status of the system after the purchase action is made but before the demand in the next period arrives. Hence, the relationship between the PDS and the normal state is

$$\tilde{d}_i = d_i, \ \tilde{e}_i = e_i + (a_i - d_i).$$

Then the post-decision value function can be expressed in terms of the normal value function as follows:

$$U_i(\tilde{d}_i, \tilde{e}_i) = \sum_{d_i'} \rho_i(d_i', \tilde{e}_i - (a_i - \tilde{d}_i)|\tilde{d}_i, \tilde{e}_i) \cdot V_i(d_i', \tilde{e}_i - (a_i - \tilde{d}_i)).$$

In PDS-based online learning, the normal value function and the post-decision value function are updated in the following way:

$$\begin{aligned} V_i^{(k+1)}(d_i^{(k)}, e_i^{(k)}) = \quad & \min_{a_i} \quad (1-\delta) \cdot c_i(d_i^{(k)}, e_i^{(k)}, a_i) \\ & + \quad \delta \cdot U_i^{(k)}(d_i^{(k)}, e_i^{(k)} + (a_i - d_i^{(k)})), \end{aligned}$$

$$\begin{aligned} U_i^{(k+1)}(d_i^{(k)}, e_i^{(k)}) \quad &= \quad (1-\alpha^{(k)}) U_i^{(k)}(d_i^{(k)}, e_i^{(k)}) \\ &+ \quad \alpha^{(k)} \cdot V_i^{(k)}(d_i^{(k)}, e_i^{(k)} - (a_i - d_i^{(k)})). \end{aligned}$$

We can see that the above updates do not require any knowledge about the state dynamics. It is proved in [11] that the PDS-based online learning will converge to the optimal value function.

V. Simulation Results

In this section, we validate our theoretical results and compare against existing DSM strategies through extensive simulations. We use the widely-used IEEE test power systems with the data (e.g. the topology, the admittances and capacity limits of transmission lines) provided by University of Washington Power System Test Case Archive [14]. We describe the other system parameters as follows (these system parameters are used by default; any changes in certain scenarios will be specified):

- A period is one hour. The discount factor is $\delta = 0.99999$.
- The demand of the aggregators follows the distribution obtained from the California ISO (CAISO) dataset [18]. Specifically, we chose 10 different demand distributions from the dataset, and assigned them to the aggregators randomly.
- The aggregators have energy storage of the capacity chosen randomly from $[5, 45]$ MW. The charging and discharge rates are constrained.
- The aggregators have linear energy storage cost [8]:

  $$c_i(s_i, a_i) = g \cdot (a_i - d_i)^+,$$

  where g is the index of the generator.
- We index energy generators starting from the renewable energy generators. All the renewable energy generators have linear energy generation cost functions: [17]

  $$c_g(a_{0,g}) = g \cdot a_{0,g},$$

  where the unit energy generation cost has the same value as the index of the generator (these values are adapted from [17], which cited that the unit energy generation cost ranges from \$0.19/MWh to \$10/MWh). Although the energy generation cost function is deterministic, the maximum amount of energy production is stochastic (due to wind speed, the amount of sunshine, and so on). The maximum amounts of energy production of all the renewable energy generators follow the distribution from the National Renewable Energy Laboratory (NREL) dataset [19].

TABLE II
OPERATIONS NEEDED TO IMPLEMENT OUR PROPOSED DSM MECHANISM.

Timeline	Before run time	In real time (e.g., each hour)
Operations	Run the algorithm in Table I	Act according to the plan computed by the algorithm in Table I
Who performs the operation	The aggregators and the ISO joinly	Each entity individually
Computation time	Up to a few minutes (see Table III for details)	Negligible

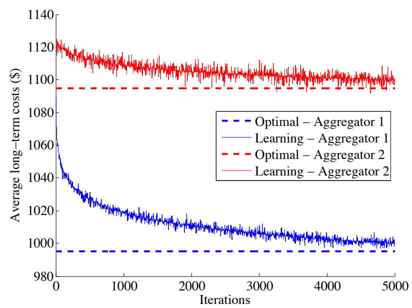

Fig. 3. Convergence of the PDS-based learning algorithm. The two aggregators' average long-term costs converge to the optimal one under the PDS-based learning algorithm.

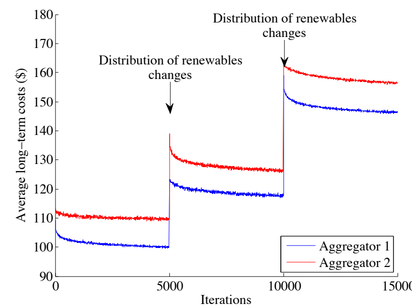

Fig. 4. The PDS-based learning algorithm tracks the optimal solution when the underlying distribution of the energy generation is time-varying.

- The rest of energy generators are conventional energy generators that use coal, all of which have the same energy generation cost function: [8]

$$c_g(a_{0,g}) = \underbrace{\frac{1}{2}(a_{0,g})^2}_{\text{generation cost}} + \underbrace{\frac{1}{10}(a_{0,g} - a_{0,g}^-)^2}_{\text{ramping cost}}.$$

  In other words, the conventional energy generators have fixed (i.e., not stochastic) generation cost functions.
- The status of the transmission lines is their capacity limits. The nominal values of the capacity limits are the same as specified in the data provided by [14]. In each period, we randomly select a line with equal probability, and decrease its capacity limit by 10%.
- There are initial installation costs of energy storage [20].

We compare the proposed DSM strategies with the following schemes.

- Centralized optimal strategies ("Centralized"): We assume that there is a central controller who knows everything about the system and solves the long-term cost minimization problem as a single-user MDP. This scheme serves as the benchmark optimum.
- Myopic strategies ("Myopic") [7]–[10]: In each period t, the aggregators myopically minimizes their current costs, and based on their actions, the ISO minimizes the current total generation cost.
- Single-user Lyapunov optimization ("Lyapunov") [1]–[6]: We let each aggregator adopt the stochastic optimization technique proposed in [1]–[6]. Based on the aggregators' purchases, the ISO minimizes the current total generation cost.

TABLE III
COMPUTATION TIME NEEDED FOR RUN THE ALGORITHM IN TABLE I.

Test cases	**Storage capacity**	**Computation time (sec)**
IEEE 14-bus	5 MW	15.23
	25 MW	33.66
	45 MW	90.13
IEEE 118-bus	5 MW	98.94
	25 MW	189.95
	45 MW	390.23

A. Learning and Tracking The Optimal Policies Without Knowledge of State Transition Probabilities

Before comparing against the other solutions, we show that the proposed PDS-based learning algorithm converges to the optimal solution (namely the optimal value function is learned). The optimal solution is obtained by the proposed algorithm in Table I assuming the statistical knowledge of the system dynamics. We consider the IEEE 14-bus system, in which each aggregator has a energy storage of 45 MW. For illustrative purpose, we show the convergence of the learning algorithm in terms of the average long-term costs only for two aggregators in Fig. 3 (note that the pricing plans of the ISO also converges).

We also demonstrate that the proposed PDS-based learning algorithm can track the optimal solution when the state transition probabilities are time-varying (e.g., the wind energy distribution changes over seasons). In Fig. 4, we change the distributions of wind energy after every 5000 time slots. We can see that the learning algorithm can track the optimal solution even when the underlying distribution is time-varying.

In Table III, we list the computation time needed for running the algorithm in Table I.[2] With a larger storage capacity, the number of states is larger, resulting in a longer computational

[2]The simulation is run on a laptop with a 1.8 GHz Intel Core i5 CPU and a 4 GB 1600 MHz memory.

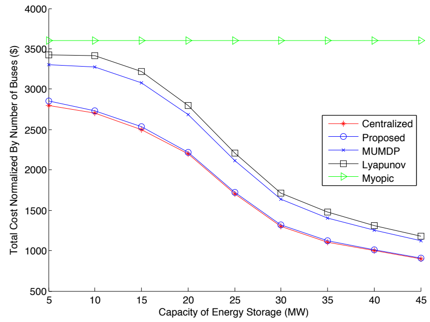

Fig. 5. The normalized total cost per hour versus the capacity of the energy storage in the IEEE 14-bus system.

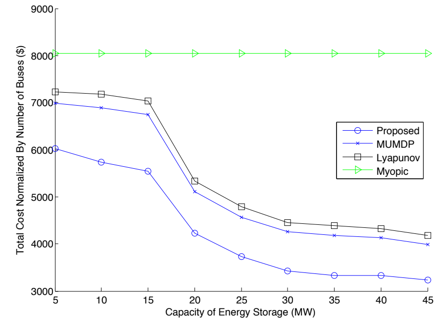

Fig. 6. The normalized total cost per hour versus the capacity of the energy storage in the IEEE 118-bus system.

time. However, even in the complicated IEEE 118-bus test case, the computation time is less than a few minutes. Since the algorithm is implemented before the run time, such a computation time is efficient.

B. Performance Evaluation

Now we evaluate the performance of the proposed DSM strategy in various scenarios.

1) Impact of the energy storage: First, we study the impact of the energy storage on the performance of different schemes. We assume that all the generators are conventional energy generators using fossil fuel, in order to rule out the impact of the uncertainty in renewable energy generation (which will be examined next). The performance criterion is the total cost per hour normalized by the number of buses in the system. We compare the normalized total cost achieved by different schemes when the capacity of the energy storage increases from 5 MW to 45 MW.

Fig. 5–6 show the normalized total cost achieved by different schemes under IEEE 14-bus system and IEEE 118-bus system, respectively. Note that we do not show the performance of the centralized optimal strategy under IEEE 118-bus system, because the number of states in the centralized MDP is so large that it is intractable to compute the optimal solution. This also shows the computational tractability and the scalability of the proposed distributed algorithm. Under IEEE 14-bus system, we can see that the proposed DSM strategy achieves almost the same performance as the centralized optimal strategy. The slight optimality gap comes from the performance loss experienced during the convergence process of the conjectured prices. Compared to the DSM strategy based on single-user Lyapunov optimization, our proposed strategy can reduce the total cost by around 30% in most cases. Compared to the myopic DSM strategy, our reduction in the total cost is even larger and increases with the capacity of the energy storage (up to 60%).

In Table IV, we show the individual average costs of the ISO and the aggregators in the IEEE 14-bus system. We can see that all the entities reduce their costs by using the proposed scheme.

TABLE IV
AVERAGE COSTS OF THE ISO AND EACH AGGREGATOR UNDER THE PROPOSED SCHEME AND THE MU-MDP SCHEME

	Storage capacity: 5MW		Storage capacity: 45MW	
	MU-MDP	Proposed	MU-MDP	Proposed
ISO	$2738.8	$2255.5	$1338.7	$1004.0
Bus 4	$3419.8	$2816.3	$1194.7	$896.0
Bus 5	$3735.5	$3076.3	$1190.7	$893.0
Bus 7	$4068.2	$3350.3	$987.2	$740.4
Bus 9	$4337.6	$3572.1	$1471.6	$1103.7
Bus 10	$3452.2	$2843.0	$1164.5	$873.4
Bus 11	$2494.4	$2054.2	$1009.5	$757.1
Bus 12	$2949.3	$2428.8	$1560.4	$1170.3
Bus 13	$2755.5	$2269.2	$1140.0	$855.0
Bus 14	$4827.0	$3975.2	$1042.9	$782.2

2) Impact of the uncertainty in renewable energy generation: Now we examine the impact of the uncertainty in renewable energy generation. For a given test system, we let half of the generators to be renewable energy generators. Recall that the maximum amounts of energy production of the renewable energy generators are stochastic and follow the same uniform distribution. We set the mean value of the maximum amount of energy production to be 100 MW, and vary the range of the uniform distribution. A wider range indicates a higher uncertainty in renewable energy production. Hence, we define the uncertainty in renewable energy generation as the maximum deviation from the mean value in the uniform distribution.

Fig. 7 shows the normalized total cost under different degrees of uncertainty in renewable energy generation. Again, the proposed strategy achieves the performance of the centralized optimal strategy. We can see that the costs achieved by all the schemes increase with the uncertainty in renewable energy generation. We can also see from the simulation that when the aggregators have larger capacity to store energy, the increase of the total cost with the uncertainty is smaller. This is because the energy storage can smooth the demand, in order to mitigate the impact of uncertainty in the renewable energy generation. This shows the value of energy storage to reduce the cost.

VI. CONCLUSION

In this paper, we proposed a methodology for performing optimal foresighted DSM strategies that minimize the long-term total cost of the power system. We overcame the hurdles of information decentralization and complicated coupling among the various entities present in the system, by decoupling

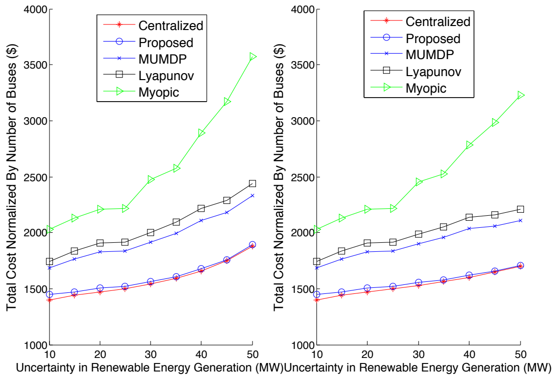

Fig. 7. The normalized total cost per hour versus the uncertainty in renewable energy generation in the IEEE 14-bus system. The aggregators have energy storage of capacity 25 MW and 50 MW, respectively.

their decision problems using conjectured prices. We proposed an online algorithm for the ISO to update the conjectured prices, such that the conjectured prices can converge to the optimal ones, based on which the entities make optimal decisions that minimize the long-term total cost. We prove that the proposed method can achieve the social optimum, and demonstrate through simulations that the proposed foresighted DSM significantly reduces the total cost compared to the optimal myopic DSM (up to 60% reduction), and the foresighted DSM based on the Lyapunov optimization framework (up to 30% reduction).

Appendix A
Proof of Theorem 1

The proof consists of three key steps. First, we prove that by penalizing the constraints $\boldsymbol{f}(s_0, a_0, \boldsymbol{a})$ into the objective function, the decision problems of different entities can be decentralized. Hence, we can derive optimal decentralized strategies for different entities under given Lagrangian multipliers. Then we prove that the update of Lagrangian multipliers converges to the optimal ones under which there is no duality gap between the primal problem and the dual problem, due to the convexity assumptions made on the cost functions. Finally, we validate the calculation of the conjectured prices.

First, suppose that there is a central controller that knows everything about the system. Then the optimal strategy to the design problem (4) should result in a value function V^* that satisfies the following Bellman equation: for all $s_0, s_1, \ldots, s_I$, we have

$$V^*(s_0, \boldsymbol{s}) = \min_{a_0,\boldsymbol{a}} \left\{ \begin{array}{l} (1-\delta)\cdot\sum_{i=0}^{I} c_i(s_i, a_i) + \\ \delta\cdot\sum_{s_0',\boldsymbol{s}'} \rho(s_0', \boldsymbol{s}'|s_0, \boldsymbol{s}, a_0, \boldsymbol{a}) V^*(s_0', \boldsymbol{s}') \end{array} \right\} \quad s.t.\ \boldsymbol{f}(s_0, a_0, \boldsymbol{a}) \le 0. \tag{6}$$

Defining a Lagrangian multiplier $\boldsymbol{\lambda}(s_0) \in \mathbb{R}_+^N$ associated with the constraints $\boldsymbol{f}(s_0, a_0, \boldsymbol{a}) \le 0$, and penalizing the constraints on the objective function, we get the following Bellman equation:

$$V^{\boldsymbol{\lambda}}(s_0, \boldsymbol{s}) = \min_{a_0,\boldsymbol{a}} \left\{ (1-\delta)\cdot\left[\sum_{i=0}^{I} c_i(s_i, a_i) + \boldsymbol{\lambda}^T(s_0)\cdot \boldsymbol{f}(s_0, a_0, \boldsymbol{a})\right] + \delta\cdot\sum_{s_0',\boldsymbol{s}'} \rho(s_0', \boldsymbol{s}'|s_0, \boldsymbol{s}, a_0, \boldsymbol{a}) V^{\boldsymbol{\lambda}}(s_0', \boldsymbol{s}') \right\}. \tag{7}$$

In the following lemma, we can prove that (7) can be decomposed.

Lemma 1: The optimal value function $V^{\boldsymbol{\lambda}}$ that solves (7) can be decomposed as $V^{\boldsymbol{\lambda}}(s_0, \boldsymbol{s}) = \sum_{i=0}^{I} V_i^{\boldsymbol{\lambda}}(s_i)$ for all $(s_0, \boldsymbol{s})$, where $V_i^{\boldsymbol{\lambda}}$ can be computed by entity i locally by solving

$$V_i^{\boldsymbol{\lambda}(s_0)}(s_i) = \min_{a_i} \left\{ (1-\delta)\cdot\left[c_i(s_i, a_i) + \boldsymbol{\lambda}^T(s_0)\cdot f_i(s_0, a_i)\right] + \delta\cdot\sum_{s_i'} \rho_i(s_i'|s_i, a_i) V_i^{\boldsymbol{\lambda}}(s_i') \right\}.$$

Proof: This can be proved by the independence of different entities' states and by the decomposition of the constraints $\boldsymbol{f}(s_0, a_0, \boldsymbol{a})$. Specifically, in a DC power flow model, the constraints $\boldsymbol{f}(s_0, a_0, \boldsymbol{a})$ are linear with respect to the actions $a_0, a_1, \ldots, a_I$. As a result, we can decompose the constraints as $\boldsymbol{f}(s_0, a_0, \boldsymbol{a}) = \sum_{i=0}^{I} f_i(s_0, a_i)$. ■

We have proved that by penalizing the constraints $\boldsymbol{f}(s_0, a_0, \boldsymbol{a})$ using Lagrangian multiplier $\boldsymbol{\lambda}(s_0)$, different entities can compute the optimal value function $V_i^{\boldsymbol{\lambda}(s_0)}$ distributively. Due to the convexity assumptions on the cost functions, we can show that the primal problem (4) is convex. In addition, there always exists a strictly feasible solution. Hence, there is no duality gap. In other words, at the optimal Lagrangian multipliers $\boldsymbol{\lambda}^*(s_0)$, the corresponding value function $V^{\boldsymbol{\lambda}^*(s_0)}(s_0, \boldsymbol{s}) = \sum_{i=0}^{I} V_i^{\boldsymbol{\lambda}^*(s_0)}(s_i)$ is equal to the optimal value function $V^*(s_0, \boldsymbol{s})$ of the primal problem (6). It is left to show that the update of Lagrangian multipliers converge to the optimal ones. It is a well-known result in dynamic programming that $V^{\boldsymbol{\lambda}(s_0)}$ is convex and piecewise linear in $\boldsymbol{\lambda}(s_0)$, and that the subgradient of $V^{\boldsymbol{\lambda}(s_0)}$ with respect to $\boldsymbol{\lambda}(s_0)$ is $\boldsymbol{f}(s_0, a_0, \boldsymbol{a})$ (it is a subgradient since the function $V^{\boldsymbol{\lambda}}(s_0)$ may not be differentiable with respect to $\boldsymbol{\lambda}(s_0)$) [11]. Note that we use the sample mean of a_0 and $\boldsymbol{a}$, whose expectation is the true mean value of a_0 and $\boldsymbol{a}$. Since $\boldsymbol{f}(s_0, a_0, \boldsymbol{a})$ is linear in a_0 and $\boldsymbol{a}$, the subgradient calculated based on the sample mean has the same mean value as the subgradient calculated based on the true mean values. In other words, the update is a stochastic subgradient descent method. It is well-known that when the stepsize $\Delta(k) = \frac{1}{k+1}$, the stochastic subgradient descent on the dual variable (i.e. the conjectured prices) $\boldsymbol{\lambda}$ will converge to the optimal $\boldsymbol{\lambda}^*$ [15].

Finally, we can write the conjectured prices by taking the derivatives of the penalty terms. For aggregator i, its penalty is $\boldsymbol{\lambda}^T(s_0)\cdot f_i(s_0, a_i)$. Hence, its conjectured price is

$$\frac{\partial \boldsymbol{\lambda}^T(s_0)\cdot f_i(s_0, a_i)}{\partial a_i} = \boldsymbol{\lambda}^T(s_0)\cdot\frac{\partial f_i(s_0, a_i)}{\partial a_i}. \tag{8}$$

References

[1] L. Jia and L. Tong, "Optimal pricing for residential demand response: A stochastic optimization approach," *Proc. IEEE Allerton Conference*, 2012.

[2] L. Jia, L. Tong, and Q. Zhao, "Retail pricing for stochastic demand with unknown parameters: An online machine learning approach," *Proc. IEEE Allerton Conference*, 2013.

[3] L. Huang, J. Walrand and K. Ramchandran, "Optimal demand response with energy storage management," *Technical Report*. Available: "http://arxiv.org/abs/1205.4297".

[4] L. Huang, J. Walrand, and K. Ramchandran, "Optimal power procurement and demand response with quality-of-usage guarantees," *Proc. IEEE Power and Energy Society General Meeting*, Jul. 2012.

[5] B.-G. Kim, Y. Zhang, M. van der Schaar, and J.-W. Lee, "Dynamic pricing and energy consumption scheduling with reinforcement learning," accepted and to appear in *IEEE Trans. Smart Grid*.

[6] Y. Zhang and M. van der Schaar, "Structure-aware stochastic load management in smart grids," accepted and to appear in *Infocom 2014*.

[7] A. Malekian, A. Ozdaglar, E. Wei, "Competitive equilibrium in electricity markets with heterogeneous users and ramping constraints," *Proc. IEEE Allerton Conference*, 2013.

[8] K. M. Chandy, S. H. Low, U. Topcu, and H. Xu, "A simple optimal power flow model with energy storage," *Proc. IEEE Conference on Decision and Control (CDC)*, Dec. 2010.

[9] Italo Atzeni, Luis G. Ordóñez, Gesualdo Scutari, Daniel P. Palomar, and Javier R. Fonollosa, "Noncooperative and cooperative optimization of distributed energy generation and storage in the demand-side of the smart grid," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2454–2472, May 2013.

[10] Italo Atzeni, Luis G. Ordóñez, Gesualdo Scutari, Daniel P. Palomar, and Javier R. Fonollosa, "Demand-side management via distributed energy generation and storage optimization," *IEEE Trans. on Smart Grids*, vol. 4, no. 2, pp. 866–876, June 2013.

[11] F. Fu and M. van der Schaar, "Learning to compete for resources in wireless stochastic games," *IEEE Trans. Veh. Tech.*, vol. 58, no. 4, pp. 1904–1919, May 2009.

[12] V. Borkar and S. Meyn, "The ODE method for convergence of stochastic approximation and reinforcement learning," *SIAM J. Control Optimization*, vol. 28, pp. 447–469, 1999.

[13] R. Sutton and A. Barto, "Reinforcement learning: An introduction," MIT Press, 1998.

[14] "Power Systems Test Case Archive," Available: http://www.ee.washington.edu/research/pstca/

[15] H. Kushner and G. Yin, Stochastic Approximation and Recursive Algorithms and Applications. Springer, 2003.

[16] M. Kohansal and H. Mohsenian-Rad, "A closer look at demand bids in California ISO energy market," *IEEE Trans. Power Syst.*, vol. 31, no. 4, Jul. 2016.

[17] E. B. Fisher, R. P. O'Neill, and M. C. Ferris, "Optimal transmission switching," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1346–1355, Aug. 2008.

[18] California ISO Open Access Same-time Information System (OASIS) [Online]. Available: http://oasis.caiso.com

[19] National Renewable Energy Laboratory, "The role of energy storage with renewable electricity generation," *Tech. Rep.*, 2010.

[20] G. L. Kyriakopoulos and G. Arabatzis, "Electrical energy storage systems in electricity generation: Energy policies, innovative technologies, and regulatory regimes," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 1044-1067, Apr. 2016.