Agent-Based Modeling of Retail Electrical Energy Markets With Demand Response

Kaveh Dehghanpour, *Student Member, IEEE,* M. Hashem Nehrir, *Life Fellow, IEEE,*
John W. Sheppard, *Fellow, IEEE,* Nathan C. Kelly, *Student Member, IEEE*

***Abstract*—In this paper, we study the behavior of a Day-Ahead (DA) retail electrical energy market with price-based Demand Response (DR) from Air Conditioning (AC) loads through a hierarchical multiagent framework, employing a machine learning approach. At the top level of the hierarchy, a retailer agent buys energy from the DA wholesale market and sells it to the consumers. The goal of the retailer agent is to maximize its profit by setting the optimal retail prices, considering the response of the price-sensitive loads. Upon receiving the retail prices, at the lower level of the hierarchy, the AC agents employ a Q-learning algorithm to optimize their consumption patterns through modifying the temperature set-points of the devices, considering both consumption costs and users' comfort preferences. Since the retailer agent does not have direct access to the AC loads' underlying dynamics and decision process (i.e., incomplete information) the data privacy of the consumers becomes a source of uncertainty in the retailer's decision model. The retailer relies on techniques from the field of machine learning to develop a reliable model of the aggregate behavior of the price-sensitive loads to reduce the uncertainty of the decision-making process. Hence, a multiagent framework based on machine learning enables us to address issues such as interoperability and decision-making under incomplete information in a system that maintains the data privacy of the consumers. We will show that using the proposed model, all the agents are able to optimize their behavior simultaneously. Simulation results show that the proposed approach leads to a reduction in overall power consumption cost as the system converges to its equilibrium. This also coincides with maximization in the retailer's profit. We will also show that the same decision architecture can be used to reduce peak load to defer/avoid distribution system upgrades under high penetration of Photo-Voltaic (PV) power in the distribution feeder.**

***Index Terms*—agent-based modeling, demand response, machine learning, retail electrical energy markets.**

I. Introduction

As the structure of power systems evolves along with the rapidly increasing penetration of variable renewable energy resources into the power grids, new techniques are introduced in the context of smart grids [1] to ensure the safe and optimal operation of the electrical energy systems [2]. In this context, Demand Response (DR) has been introduced to give the consumers the opportunity of participating in power system management and control processes.

This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award # DE-FG02-11ER46817.

K. Dehghanpour, H. Nehrir, and N. Kelly are with the Department of Electrical and Computer Engineering, Montana State University, Bozeman, MT 59717 USA (e-mail: hnehrir@ece.montana.edu; Kaveh.dehghanpour@msu.montana.edu; kellyna4@gmail.com).

J. Sheppard is with the Department of Computer Science, Montana State University, Bozeman, MT 59715 USA (email: john.sheppard@montana.edu)

DR programs are generally classified into two distinct categories [3]: 1) incentive-based DR programs, in which the system operator provides consumers with monetary incentives in return for various ancillary services, such as frequency and voltage regulation services, direct load control, and emergency DR, and 2) time-based DR programs are price-based procedures, including time-of-use pricing, peak-pricing, and real-time pricing. Time-based DR programs are of particular interest in this paper.

The basic idea in price-based DR is to introduce market mechanisms at the retail level, to which automated load control agents are able to respond. In this way, a level of price-sensitivity can be achieved on the demand side; that is, consumers change their consumption patterns in response to varying prices they receive. As shown in [4], in order to maintain the economic efficiency and viability of the markets in practice, the response of the consumers to energy prices needs to be considered and integrated within the pricing process of the market. To achieve this task, bilevel iterative decision models have been proposed at the electrical distribution level [5] [6]. This implies the need for developing smart metering and bidirectional communication networks between consumers and utility companies. As discussed in [7], the penetration of advanced metering devices throughout the U.S. increased to 36.3% of all the metering devices by July 2014, compared to 22.9% in 2011, and 8.7% in 2010 [3], which shows a promising trend in implementing DR programs.

In this paper, we present a price-based DR procedure for Day-Ahead (DA) planning and decision-making in retail electrical energy markets using an agent-based framework. While this problem has been addressed in the literature using different tools, such as multi-objective optimization [8], mixed integer linear programming [9], model predictive control [10], particle swarm optimization [11], and gradient-based methods [12], the novelty of this paper lies in the use of an agent-based approach, with agents employing techniques from the field of machine learning to model their environment and optimize their behavior. In this way, we can model and study the behavior of retail energy markets in a realistic context without burdening the examination with oversimplifying assumptions on the state of information of different entities in the system. Also, given that different computational tasks are distributed among agents, the proposed solution will be scalable for practical implementation. More specifically, agent-based modeling is employed in this paper to address the problems of interoperability and data privacy in retail power markets. In this paper, we assume that agents have no information on their

peers' private data, which addresses the concerns on privacy protection discussed in [3]. This data privacy leads to incomplete information of agents on their peers' behavior, which in turn, contributes to uncertainty in their decision models. In other words, an agent-based framework in combination with machine learning techniques corresponds to the natural state of decentralized and distributed decision-making structure of the interoperable retail energy markets. Interoperability is defined in [13] as: "the capability of two or more networks, systems, devices, applications, or components to exchange information between them and use the information exchanged." The role of interoperability in grid modernization and integration of new resources in power systems is discussed in [14]. Noting that two of the most essential properties of agents in a multi-agent setting are *autonomy* and *social capability* (i.e., the ability to exchange data with peers) [15], the connection between multi-agent systems and interoperable systems becomes clear and well-founded. Each component of an interoperable network can be viewed as an autonomous agent that interacts with other components.

Also, employing a multi-agent system approach introduces a certain degree of independency in modeling different decision and control mechanisms in the system. For instance, in a multi-agent setting, the decision problem of the retailer and loads can be decoupled completely, since each of them is being handled by distinct agents. This has enabled us to test and compare different decision and control procedures (as has been done in this paper) without having to re-design the whole model from scratch. More on the merits of agent-based modeling can be found in [16].

As shown in the literature [17], Thermostatically Controlled Loads (TCLs) have a high potential for being candidate appliances to participate in a DR program. Due to their considerable thermal capacity, TCLs, can temporarily deviate from their desired consumption pattern without causing significant discomfort to the consumers. Specifically, in this paper, we consider Air Conditioning (AC) loads as primary agents that participate in the price-based DR program. In this way we can observe the effects of the dynamics of the loads on the market.

In [5], the retail market is formulated as a bi-level decision problem, introducing a two-stage pricing mechanism, through a distributed convex optimization problem. Limiting assumptions have been made to keep the optimization problems convex. Also, it is assumed that the pricing mechanism has access to the complete information on the decision problems of the loads. In [9], the reaction of demand side to energy prices is modeled as simple elasticity coefficients that appear as constraints in the optimization problem of a profit-oriented utility company. While this generic approach provides valuable insight into the decision-making of a utility company, it does not capture the dynamic behavior of loads in the retail pricing process. Another DR scheme is proposed and solved in [18] considering the uncertainty of wind power and grid energy price. In this work, also, generic models have been considered for loads. Moreover, the DR problem is solved through a central optimization problem with access to all consumers' data. In [19], purchase bidding strategies of an energy coalition with a non-profit aggregator and DR is studied, again, through

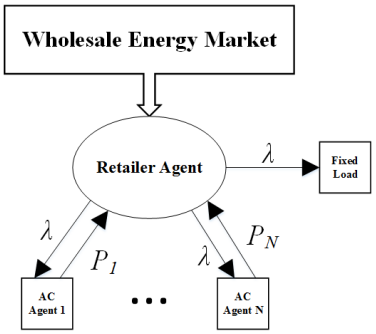


Fig. 1. The overall architecture of the agent-based model.

a bi-level decision problem. The authors have relied on Monte-Carlo simulations and stochastic optimization to account for the price-sensitivity of the loads in decision-making. Since no learning mechanism is adopted to direct the search process, the number of iterations required to solve the problem is very high (in the order of 1000). Another iterative pricing mechanism is introduced in [12] to flatten the load profile through peak reduction. In this paper, the utility company relies on estimated gradient of price sensitive demand along the retail prices. Hence, to guarantee the global optimality and proper convergence of the gradient method the authors have kept the optimization problem convex. In an interesting work, the problem of response of the consumers to retail prices is discussed within a bi-level decision problem [6]. In this paper the authors use an iterative scheme and a simulated-annealing-based price control strategy to perform retail energy pricing. The nonlinear dynamics of the loads are ignored and simplified to keep the decision problem of the loads convex.

In our paper, we propose a distributed decision-making framework for implementing a DR program to address several issues that we believe have been ignored and under-studied in the previous works: 1) we avoid simplifying the load models to obtain convex decision problems. We will show that even simple nonlinear first-order load models lead to non-convex optimization problems on the retailer-side, 2) we keep the decision problem of consumers as simple as possible for ease of implementation, 3) we address the effect of uncertainty in the retailer's decision model resulting from incomplete knowledge about the behavior of the price-sensitive loads and their private settings. It is also critical to investigate how different forecasting tools can be incorporated in the decision model of the retailer to limit the uncertainty of the problem, 4) we study how variations in users' preferences in terms of cost-sensitivity affect the equilibrium of the market.

The proposed hierarchical agent-based framework in this work consists of two levels, as shown in Fig. 1. At the top level, a retailer agent buys energy on the wholesale market and sells it to the consumers, consisting of both fixed and price-sensitive loads. Hence, the retailer can be viewed as a load aggregator and power supplier. The primary objective of the retailer is to maximize its profit from sales of energy. For that purpose, the retailer develops a model of the aggregate response of price-sensitive loads to retail prices. This model is

basically a forecasting tool that is learned through interaction with consumers. We consider two possibilities at this stage: a first case in which the retailer employs a linear model, and a second case, for which the retailer uses a nonlinear model in the form of an Artificial Neural Network (ANN) to approximate the aggregate response of the price-sensitive loads. While the linear model is learned using multiple linear regression and QR decomposition [20], the ANN is parameterized via Bayesian Regularized Back Propagation (BRBP) [21]. Using the distinct learned models (i.e., load forecasting tools), the retailer formulates the DA profit maximization problem, which is solved using Particle Swarm Optimization (PSO) [22]. One objective of this paper is to compare the performances of the two modeling approaches. In other words, we will investigate the efficiency of a linear model versus a non-linear model and their effects on the revenue stream of the retailer.

At the lower level of the hierarchy, the AC agents optimize their consumption patterns independently using their local controllers, by setting proper temperature set-points after receiving the retail prices from the retailer, considering both the cost of consumption and the consumers' comfort levels along with the predicted ambient temperature. The consumers have the freedom to determine the trade-off between increasing cost reduction and reducing deviation from comfort zone, through private settings in their decision model. This problem is formulated as a Markov Decision Process (MDP) [23] and solved via Q-learning [23]. Upon calculating their DA expected consumption profiles, the AC agents send this information as feedback signals to the retailer agent. At this level, we define two case studies: *mild DR* and *active DR*. For the case of mild DR, the population of AC agents shows less sensitivity to retail prices. In *active DR*, however, the portion of consumers that actively seek to cut their consumption costs (at the expense of higher deviations in temperature set-points) dominate the population of the AC agents. These two case studies help us understand the behavior of the retail market as the level of the price-sensitivity of the demand-side participants in the market changes.

The proposed agent-based model is a bi-level sequential decision-making process: the retailer updates its model of the consumers, and revises the retail prices based on the newly received feedback data on aggregate AC consumption levels. On the other end, the AC agents revise their consumption pattern based on the newly received retail prices. This sequential decision-making process relies on the existence of a bidirectional communication network and local decision-making algorithms. As will be demonstrated, using this method, the system converges to its equilibrium. Also, it will be shown that the approach of the system to its equilibrium coincides with a reduction in total consumption cost and magnitude, which shows a promising ground for practical implementation of the algorithm. On the retailer side, the approach of the model to equilibrium coincides with profit maximization.

While our proposed pricing scheme leads to reduction of overall consumption level, it can also lead to creation of minor secondary peaks at later hours of the day, as is shown in Section V. The minor secondary peak could lead to congestion and overloading of the distribution system in the presence of Photo-Voltaic (PV) power in the system. We will show that in addition to profit maximization, the same retail pricing mechanism can be applied by the aggregator to reduce the peak load and mitigate the problem of congestion, as a secondary objective, in high PV penetration scenarios, in order to avoid/defer distribution system upgrades. In summary the contributions of the paper are as follows:

- Introducing an agent-based approach based on the concept of "learning" to model the retail energy markets with DR. The performance of different machine learning techniques at the retail level are compared, under different case studies.
- Using an MDP and Q-learning to model the behavior of price-sensitive AC loads, considering the uncertainty of their initial conditions. Monte-Carlo simulation was used to obtain the aggregate response of the population of ACs.
- The problem of uncertainty of the decision-making of the retailer (due to incomplete information on the state of price-sensitive loads) is addressed using machine learning techniques to design load forecasting tools.
- The effects of variations in consumers' private settings and preferences on the equilibrium of the market are addressed.

The rest of this paper is organized as follows: in Sections II, III, and IV the basic functionality of the agent-based framework at the two levels of the hierarchy is described. In Section V, the numerical results are shown and discussed. The main conclusions are presented in Section VI.

II. AC Agents' Decision Problem

In this section the functionality of the AC agents in the market will be discussed, and their overall decision-making problem will be explained. The decision problem of each AC load is solved by the individual controller of that load using consumer's private settings. The market operation is on an hour-by-hour basis.

A heterogeneous population of price-sensitive AC loads is participating in the retail market. The AC agents are in charge of controlling the AC loads by determining optimal DA temperature set-points, based on the forecasted DA ambient temperature and estimated initial states. When determining the temperature set-points, AC agents need to consider three issues: the dynamics of the AC loads, the total cost of energy, and the consumers' comfort level. Note that the DA hourly retail prices act as constant inputs to the loads' decision problems.

The dynamics of TCLs, including ACs, can be described by a non-linear first order system of differential equations as shown in [24]. Using the load dynamics, the DA temperature set-points, and forecasted DA ambient temperature vector, the AC agent is able to estimate the level of consumed power for different hours of the next day. We assume that the temperature forecasting is performed by a Numerical Weather Prediction (NWP) unit, and the predicted set-point values are treated as given inputs in the ACs decision-making units. Note that the dynamics of the load is a source of uncertainty for the problem, since the initial room temperature and thermostat status are not

TABLE I
AVERAGE VALUES OF AC LOAD PARAMETERS

Parameter	Value
R	2°C/kW
C	10 kWh/°C
P_N	14 kW
δ	1°C
T_{des}	19°C
η	2.5

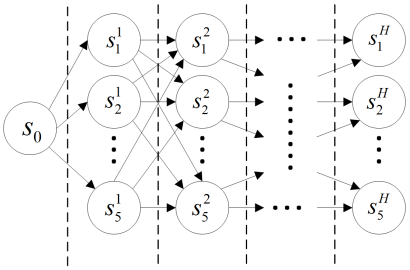


Fig. 2. Proposed MDP for AC level decision-making.

known, *a priori*, by the agents. The dynamics of a single AC is shown in (1).

$$\frac{dT}{dt} = \frac{1}{RC}(T_{am} - T(t) - P_N \cdot R \cdot m(t)),$$
$$m(t) = \begin{cases} m(t) = 1, & \text{if } T > T_{set} + \frac{\delta}{2} \\ m(t) = 0, & \text{if } T < T_{set} - \frac{\delta}{2} \\ m(t) = m(t-1), & \text{otherwise.} \end{cases} \quad (1)$$

where, parameters R, C, T, T_{am}, and P_N denote room/house thermal resistance, thermal capacitance, inside temperature, ambient temperature, and nominal AC cooling power, respectively. The variable $m(t)$ is a binary variable that represents the ON/OFF thermostat status, with δ being the operational dead-band of the device. The nominal electrical power of the AC load (P_e) is obtained using :

$$P_e = \frac{P_N}{\eta} \quad (2)$$

where, η is the load efficiency. The *average* values of load parameters, selected according to [24], are given in Table I.

To perform the decision-making, the problem is formulated as an MDP (Fig. 2). At each state the agent can select from a set of available actions. The selected action then leads the agent to another state, based on a state transition function. Also, each state transition results in a penalty value for the agent, according to the MDP's penalty function. The goal of the agent is to minimize its aggregate penalty by finding the optimal action at each state (note that each AC agent is equipped with its own private MDP).

The immediate penalty within the proposed MDP is defined by two parameters: the estimated DA consumed power, which is obtained by the load dynamic model, and the violation of consumer's comfort level, which is defined by the absolute value of temperature set-point deviation from the consumer's desired temperature. Hence, the immediate penalty function consists of two competing terms: one objective is to minimize total energy costs, and the other is to stay within the consumer's comfort zone as often as possible. The consumer has the freedom of balancing these two objectives by assigning weights to them. The states of the MDP specify the allowed discretized values of deviation of temperature set-point from the desired temperature at each hour of the next day.

Five states are defined for each hour of the day, with each state corresponding to certain degrees of deviation in the temperature set-point (T_{set}) from the desired temperature (T_{des}) at each specific hour. The average value of T_{des} over the population of AC loads is given in Table I. The deviation values, in Celsius, are selected from the set $\{+2, +1, 0, -1, -2\}$, corresponding to states s_1 through s_5, respectively. For instance, s_2^j implies +1°C deviation in temperature set-point of the AC from the desired temperature at the j^{th} hour of the day, with j changing from 1 to $H = 24$ (with H denoting the planning period). The immediate penalty (π) for an action at the $(j-1)^{th}$ hour is calculated as follows:

$$\pi^j = \frac{|T_{set}^j - T_{des}|}{N_1} w_1 + \frac{p^j \lambda^j}{N_2} w_2. \quad (3)$$

The first term in (3) (i.e., $\frac{|T_{set}^j - T_{des}|}{N_1} w_1$) penalizes deviations from the desired temperature level at the j^{th} hour (with N_1 as normalizer). The second term (i.e., $\frac{p^j \lambda^j}{N_2} w_2$) penalizes the total cost of consumed electrical energy for the j^{th} hour of the day (with p^j, λ^j and N_2 denoting total energy consumption at the j^{th} hour, retail price at the j^{th} hour, and a normalizer term, respectively). Here, N_1 and N_2 are equal to the maximum temperature set-point deviation and maximum consumption cost, respectively. Hence, the first term serves as a measure of the consumer's comfort level, while the second term acts as a measure of the tendency of the consumer to cut energy costs. Different consumers have different preferences on balancing their energy costs and comfort levels, which is modeled as the two weights in the penalty function, w_1, and w_2, with: w_1, $w_2 \in [0, 1]$, $w_1 + w_2 = 1$.

Q-learning [23], which is a type of model-free reinforcement learning algorithm, is used to obtain the optimal sequence of temperature set-points for the device. Using Q-learning, an agent can find the optimal course of action without having full knowledge of transition and penalty functions of the MDP. The basic idea of Q-learning is to assign a Q-value to each state-action pair at time t, i.e., $Q(s_t, a_t)$, and update it at each encounter, in a way to reinforce good behavior. The Q-values correspond to the long-term "worth" of state-action pairs. Hence, each AC agent develops a private look-up table that contains the Q-values of the state-action pairs of the proposed MDP. The update mechanism for the Q-values in the look-up table is shown below:

$$Q(s_t, a_t) := Q(s_t, a_t) + \alpha_t \cdot (-\pi_t + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t)) \quad (4)$$

where, α_t is a variable learning rate, π_t is the immediate penalty (obtained according to (3)), and γ is a discount factor. In order to take uncertainties of load dynamics into account, (i.e., uncertain initial room temperature and initial thermostat state) the learning process will be repeated for a high number

of episodes with different initial conditions. Consequently, we can ensure that the ACs will have an expected desirable behavior under different real-time scenarios.

Note that the coefficients w_1, and w_2 in (3) are "user-defined", for each AC agent. This means that based on private preferences, the consumers are able to modify the rate of "price-sensitivity" of their appliance, even "turning off" the cost-sensitive module altogether, by setting $w_2 = 0$. Hence, based on the distribution of w_1, and w_2 in the population of AC agents, two types of DR programs are defined. We will investigate and compare the maximum profit level of the retailer under these two programs.

1) Mild DR: In this case, the value of w_1 is selected according to a uniform distribution over the interval (i.e., $w_1 \sim U[0, 1]$). This mean that the number of AC agents that value comfort level over savings in monetary costs are roughly the same as the number of AC agents that actively seek to reduce consumption costs.
2) Active DR: In this case, w_1 values for the AC agents are selected based on a uniform distribution over the interval (i.e., $w_1 \sim U[0, 0.5]$), which implies that the AC agents that value consumption cost savings over comfort level constraints dominate the population. An active DR corresponds to increased price-sensitivity of consumers in the retail markets.

Employing Q-learning to address the decision problem of consumers has several advantages that are discussed below:

- Q-learning is model free. Hence, the decision strategy is independent of the agent's knowledge of the AC load model. This implies that the proposed method can be generalized to more complex AC load models. The model-free nature of Q-learning provides the decision-making agents with higher levels of flexibility and controllability over the classical optimization approaches, such as linear programming [25]. While in linear programming we need to linearize the underlying dynamics of the loads to solve the optimization problem, in a model-free approach such as Q-learning, the solution strategy is independent of the properties of the underlying models. Hence, we are able to capture the effects of the non-linearity of the models on the decision problem.
- Another advantage of Q-learning is its simplicity. The whole computational process of the algorithm is based on a look-up table and an update rule (equation (4)). The ease of implementation of an algorithm is crucial, specifically at electrical distribution level and for home energy management systems.
- Also, through Q-learning, the uncertainty of the system can be considered in the decision-making process. This is achieved through episodic learning, as explained previously. The update process, (3), takes place under different episodes. Each of these episodes represent different scenarios that reflect our incomplete and uncertain knowledge of different variables in the decision model. In this paper episodic learning is performed to account for the uncertainty of the initial state of the AC devices in real-time (i.e., initial ON/OFF status and initial temperature according to probability distribution functions in [24]).
- Q-learning is a reinforcement learning method. Thus, unlike supervised learning schemes, in Q-learning we do not need to provide the decision-making agents with correct or optimal solution samples beforehand. Through interactions with their environment the agents are able to obtain the optimal course of action to maximize their pay-off level. In this case, provided with a load model, the AC agents are able to track the optimal temperature set-points for given prices at different hours of the day.

Also, other alternative methods were tried instead of Q-learning, such as the value-iteration [23] method and non-linear programming [25] (solved using PSO). However, Q-learning showed better performance in terms of speed of convergence and quality of final results.

III. Retailer Agent's Decision Problem

The retailer agent (i.e., the aggregator) develops a model based on the feedback signals it receives from the AC agents to approximate the aggregate behavior of price-sensitive loads as a function of the retail price vector composed of hourly prices. Hence, the goal of the retailer is to perform a function approximation procedure. This model is learned and incorporated into the profit maximization problem of the retailer agent. The outcome of the optimization problem is the optimal retail price vector, which is consequently sent to the price-sensitive loads via the communication network. The same decision model and pricing mechanism can be used for peak reduction, with minor changes. We discuss peak shaving (Sections III.C and V.C) as a secondary objective for the retailer agent, in the presence of PV power in the distribution feeder.

Two distinct modeling approaches on the retailer side are studied and compared in this paper: using a linear model, and a non-linear ANN-based model. Each of these modeling approaches leads to a distinct optimization problem formulation for the retailer. In this section, we also address the problem of solving the profit maximization/peak reduction for each adopted model.

A. Linear Model

This model represents the aggregate consumed power of the price-sensitive loads at each hour of the day as a linear combination of hourly retail prices, as shown below:

$$\begin{bmatrix} P^1 \\ \vdots \\ P^H \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1H} \\ \vdots & \ddots & \vdots \\ a_{H1} & \dots & a_{HH} \end{bmatrix} \begin{bmatrix} \lambda^1 \\ \vdots \\ \lambda^H \end{bmatrix} + \begin{bmatrix} P_0^1 \\ \vdots \\ P_0^H \end{bmatrix} \tag{5}$$

where P^j and λ^j denote the total consumed energy of the AC loads and the retail energy price at the j^{th} hour of the day, with j changing from 1 to $H = 24$. This model is learned through multiple linear regression, employing QR decomposition. Equivalently, (5) can be written as,

$$\boldsymbol{P} = \boldsymbol{A\lambda} + \boldsymbol{P_0}. \tag{6}$$

Using (6), the retailer agent is able to develop and maximize its total profit to obtain the optimal set of retail prices. The profit maximization problem is formulated as follows:

$$\max_{\lambda^1,\ldots,\lambda^H} \sum_{i=1}^{H} (\lambda^i - \lambda_g^i)(P^i + P_f^i), \tag{7}$$

where, λ_g^i, and P_f^i denote the wholesale DA energy price, and fixed power consumption, respectively (all at the i^{th} hour of the next day). Using (6), this optimization problem can be written as:

$$\min_{\lambda^1,\ldots,\lambda^H} \sum_{i=1}^{H} -(\lambda^i - \lambda_g^i)(\boldsymbol{a_i} \cdot \boldsymbol{\lambda} + P_0^i + P_f^i), \tag{8}$$

with $\boldsymbol{a_i}$ being the i^{th} row of matrix $\boldsymbol{A}$. Employing algebraic manipulations, (8) is transformed into a non-convex constrained quadratic programming problem as follows [25]:

$$\min_{\lambda} -\boldsymbol{\lambda^T A \lambda} + (\boldsymbol{\lambda_g^T A - P_0^T - P_f^T})\boldsymbol{\lambda} + (\boldsymbol{\lambda_g^T P_0 + \lambda_g^T p_f}),$$
$$s.t. \begin{cases} \boldsymbol{\lambda_{min}} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\lambda_{max}} \\ \frac{1}{H}\sum_{i=1}^{H} \lambda^i = \frac{1}{H}\sum_{i=1}^{H} \lambda_g^i. \end{cases} \tag{9}$$

where "$\preceq$" denotes element-wise "$\leq$" operator for vectors. The constraints in (9) are designed to ensure two properties: the retail prices remain bounded, and the average retail price would be constant (in this case equal to average wholesale prices). These properties can be viewed as regulatory requirements or even mutual agreements among the retailer and its customers, to keep the prices "fair". The retailer has a short-term monopoly over the consumers due to long-term contracts with them. Therefore, the solution to the optimization would always be $\boldsymbol{\lambda} = \boldsymbol{\lambda_{max}}$, without the introduced constraint on the average retail prices, given in (9). Also, under fixed and frozen retail energy pricing, the fixed retail prices reflect the long term average wholesale electricity prices [26]. Hence, we believe the constraints in the optimization problem (9) are necessary to connect the wholesale and retail market models, even under time-varying retail tariffs.

The output of (9) gives us the optimal retail prices for each hour of the next day. These prices are generated at each iteration and sent to the consumers. Upon receiving the power consumption feedback signals from the AC agents, model (6) is updated and (9) is solved again to update the prices.

B. Nonlinear ANN-based Model

Unlike a linear model, an ANN is able to capture the nonlinearity of the loads' aggregate behavior. To implement an ANN-based model of the collective behavior of AC agents, the retailer agent employs a three-layer feedforward structure of artificial neurons [27]. The input and output layers consist of $\mathbf{24}$ neurons each. The hidden layer consists of $\mathbf{25}$ neurons, chosen using cross validation. BRBP is adopted for learning the weights of the connections within the network. BRBP is computationally burdensome; however, it produced superior results for this application (i.e., higher predictive capabilities), compared to other learning approaches. The functionality and advantages of BRBP are discussed in [21], and [28].

Basically, after training, the ANN will be able to map the retail price vector to the aggregate power consumption vector of the AC agents. This nonlinear mapping can be shown as follows:

$$\boldsymbol{P} = \boldsymbol{Net(\lambda)}. \tag{10}$$

Then the retailer's profit maximization problem (with the same set of constraints as (9)) can be formulated as,

$$\min_{\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_H} -(\boldsymbol{Net(\lambda)} + \boldsymbol{P_f})^T(\boldsymbol{\lambda} - \boldsymbol{\lambda_g}),$$
$$\boldsymbol{s.t.} \begin{cases} \boldsymbol{\lambda_{min}} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\lambda_{max}} \\ \frac{1}{H}\sum_{i=1}^{H} \boldsymbol{\lambda}^i = \frac{1}{H}\sum_{i=1}^{H} \boldsymbol{\lambda}_g^i. \end{cases} \tag{11}$$

Since an ANN acts as a "black box" within the objective function, solving (11) directly is not easy (in contrast to (9), which was based on a linear model). However, by defining a "fitness" function for (11), population-based algorithms can be applied to solve it. In this paper we have chosen PSO as a solver to (11). Also, to provide a fair comparison of the linear and ANN-based models, PSO has been used to solve (9), as well.

C. Peak Reduction

The energy pricing mechanism employed by the load aggregator agent for profit maximization can also be used to reduce peak value of the load to avoid/defer distribution system upgrades. The basic difference with problems (9) and (11) is that the objective function of the peak reduction problem would be the maximum load value, instead of profit level. Hence, the energy pricing mechanism is performed as follows:

$$\min_{\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_H} (\max_{i=\{1,\ldots,H\}} (\boldsymbol{Net(\lambda + P_f - P_R)})),$$
$$\boldsymbol{s.t.} \begin{cases} \boldsymbol{\lambda_{min}} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\lambda_{max}} \\ \frac{1}{H}\sum_{i=1}^{H} \boldsymbol{\lambda}^i = \frac{1}{H}\sum_{i=1}^{H} \boldsymbol{\lambda}_g^i. \end{cases} \tag{12}$$

where, $\boldsymbol{P_R}$ denotes the forecasted renewable power values for the given decision window. As we will show in the result section, the problem of peak reduction and secondary peaks in presence of solar power is of critical importance at distribution level. Hence, we have solved (12) for a distribution system with high penetration of PV power.

D. Solution Strategy

In order to incorporate the constraints of the optimization problems into the PSO, different heuristics have been introduced in the literature [29]. Here, to deal with the equality constraint, we have added a penalty term to the fitness function to penalize deviations of the particles from the feasible region. To handle the inequality constraints, at each iteration, the elements of retail price vector ($\boldsymbol{\lambda}$) that violate the maximum and minimum price limits are removed and replaced with the maximum or minimum price values, depending on the constraint boundary that was crossed. Hence, the augmented fitness function is as follows,

$$\boldsymbol{Fitness(\lambda)} = \boldsymbol{Profit(\lambda)} - \gamma\Big|\frac{\mathbf{1}}{\boldsymbol{H}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{H}} \boldsymbol{\lambda}^{\boldsymbol{i}} - \frac{\mathbf{1}}{\boldsymbol{H}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{H}} \boldsymbol{\lambda}_{\boldsymbol{g}}^{\boldsymbol{i}}\Big|, \tag{13}$$

where, γ is the penalty coefficient which is treated as another tunable parameter in the model. The profit function, $\boldsymbol{Profit(\lambda)}$, is obtained for the linear and ANN-based models according to the objective functions of (9), and (11), respectively. For the problem of peak reduction, we simply replace $\boldsymbol{Profit(\lambda)}$ with $\boldsymbol{Peak(\lambda)}$ in (13). $\boldsymbol{Peak(\lambda)}$ denotes the peak load level for retail price $\boldsymbol{\lambda}$, which is obtained using the objective function of optimization problem (12).

Considering the fitness function given by (13), the base dynamics of the PSO algorithm is according to the following update rules (denoting the positions of the i^{th} particle by $\boldsymbol{X_i}$ and its speed by $\boldsymbol{V_i}$):

$$\boldsymbol{V_i^{k+1}} = \omega_k \boldsymbol{V_i^k} + c_1 r_1 (\boldsymbol{pBest_i^k} - \boldsymbol{X_i^k}) + c_2 r_2 (\boldsymbol{gBest^k} - \boldsymbol{X_i^k})$$

$$\boldsymbol{X_i^{k+1}} = \boldsymbol{X_i^k} + \boldsymbol{V_i^{k+1}}$$

$$\omega_k = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{k_{max}} k, \qquad (14)$$

with ω_k acting as a weight parameter ω_{max} and ω_{min} represent the maximum and minimum values of the weight), c_1 and c_2 used as tunable coefficients, r_1 and r_2 being uniformly generated random numbers from the [0,1] interval, $\boldsymbol{pBest_i^k}$ and $\boldsymbol{gBest^k}$ representing the personal and global best solutions, and k_{max} denoting the maximum number of iterations of the algorithm.

Parameter tuning for (14) is performed based on numerical tests on the model. The values of the parameters that showed the best performances are as follows: $c_1 = 0.1$, $c_2 = 7$, $\omega_{max} = 0.9$, $\omega_{max} = 0.1$, and $k_{max} = 15000$. Also, the size of the swarm of particles is selected to be 50. The large number of the particles is due to the relatively high dimensionality of the decision variable (i.e., $\boldsymbol{X} \in \mathcal{R}^{24}$).

IV. Putting The Pieces Together

Fig. 3 shows a flow-diagram of the step-by-step iterative process of the model on the retailer side and the consumer (AC) side, referring to the equations used at each step. The algorithm needs a number of iterations to converge. At each iteration, the retailer agent updates the linear model or the ANN, and calculates the optimal retail prices based on the learned model. These prices are sent to the AC agents that obtain their optimal consumption patterns using Q-learning. Then, the AC agents send back their expected consumption levels to the retailer, to be used for the next iteration. The minimum number of iterations needed for the convergence of the system depends on the sample complexity of the model that the retailer employs [30]. Basically, sample complexity determines the required number of samples (i.e., iterations) to learn and develop reliable models and avoid overfitting. At each iteration, using the learned model, the retailer has the opportunity to predict the aggregate response of the AC agents to the obtained optimal price vector. The prediction Mean Absolute Error (MAE) is used as a measure of deciding whether overfitting occurs or not. As lower values of MAE are achieved through iterations, the learned model becomes more reliable for decision-making. Note that the prediction MAE is a measure of the uncertainty of the decision model of the retailer agent, caused by incomplete information on private control processes and individual settings of AC agents (due to data privacy).

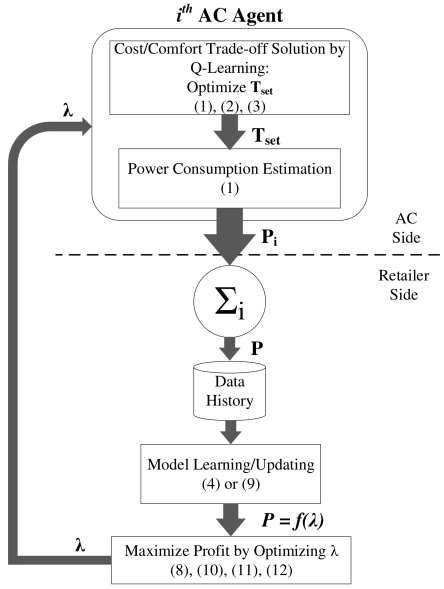


Fig. 3. Flow-diagram of the agent-based model at each iteration.

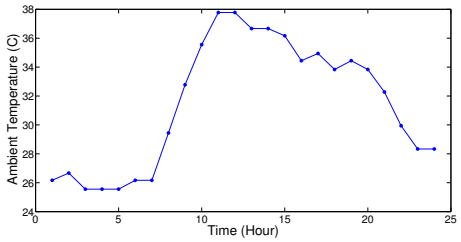


Fig. 4. Forecasted ambient temperature.

V. Numerical Experiments and Results

The proposed method is tested on a sample distribution feeder with one retailer agent and 200 AC agents. The parameters of the AC agents are selected according to log-normal distribution functions used in [24]. The weight values w_1, and w_2 are determined using uniform random distribution functions to represent the two different types of DR programs (mild and active DR) discussed in Section II. Also, the predicted DA ambient temperature for a summer-day, adopted from [31], is shown in Fig. 4. The daily fixed load data for the feeder is based on [32], and [33]. The peak value of the fixed load profile is 1.4 MW. On the other hand, the peak value of the aggregate AC consumption level without DR is 0.8 MW, which is around 35% of the total load peak value. The DA energy prices are chosen from real DA price data of the PJM market [34]. Evaluation of the performance of the AC agents and the retailer agents is discussed in this section.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2016.2631453, IEEE Transactions on Smart Grid

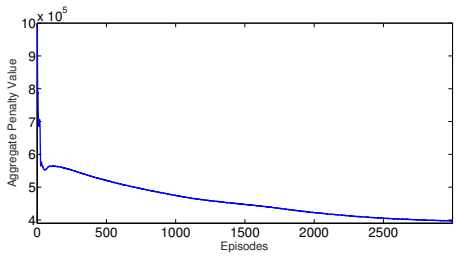
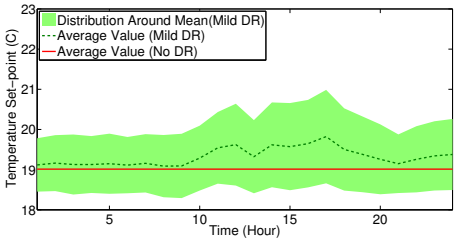
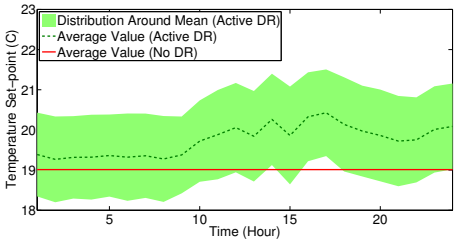


Fig. 5. Aggregate accumulated penalty of the AC agent using Q-learning.



(a) Mild DR



(b) Active DR

Fig. 6. Temperature set-point distribution over time.

A. AC Agents' Performance

To verify the performance of the AC agents, the total accumulated penalty values of all the 200 agents over the episodes is shown in Fig. 5 for a certain retail price vector. As can be seen, the penalty curve is decreasing in episodes, which implies that the AC agents are able to reduce the overall cost using Q-learning. Also, in Fig. 6 the final DA average temperature set-point distribution of all the devices is shown for the mild and active DR programs. By comparing Fig. 6a and Fig. 6b, we observe that in an active DR program, the average temperature set-points tend to show higher deviations from the case without DR. Also, comparing the two cases of mild and active DR we observe a higher standard deviation in the temperature set-point profile for the latter. This implies that as price-sensitivity increases on the consumer side, the retailer faces a higher level of uncertainty (i.e., it would be more difficult to predict the response of the AC agents to prices).

As observed in Fig. 6, the AC loads go through a "pre-cooling" period during the first few hours (i.e., average temperature set-point is kept around the average desired value) in order to be able to remain deactivated during the hours with higher prices without violating the temperature constraints. In the final few hours of the day, a slight increase in the average temperature set-points is observed, which brings the

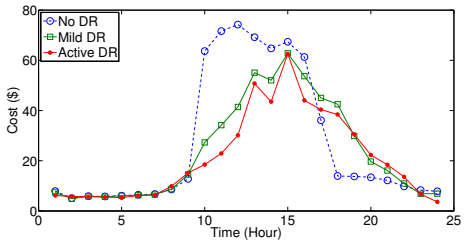


Fig. 7. Overall cost profile of the AC agents.

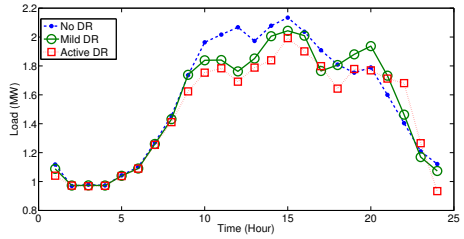


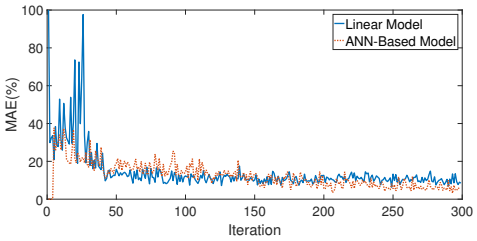
Fig. 8. DA load profile of the system.

consumption cost back almost to its initial levels, as shown in Fig. 7. The drop and a slight shift in the consumption cost is shown in this figure. The total payment of the AC agents for consuming energy for the day in different scenarios are: \$653 (without DR), \$570.3 (mild DR), and \$512.4 (active DR). Hence, using the case without DR scenario as a base, the reductions in cost of consumption are equal to 12.6%, and 21.5% for mild and active DR cases, respectively. These values have been obtained based on the optimal retail prices received from the retailer, as discussed in the next subsection. Hence, the total payment of the AC agents is equal to the maximum revenue value of the retailer.

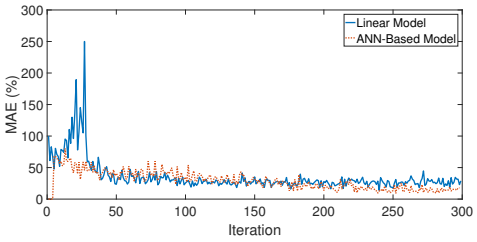
The overall DA load profile (consisting of both price-sensitive and fixed electrical demand) is shown in Fig. 8. As is demonstrated in this figure, the DR program leads to a decrease in the peak value and shifting of the load to the later hours of the day. The peak load drops from 2.134 MW to around 2.084 MW (2.34% decrease). The total reductions in the AC power consumption level compared to the case without DR are equal to 6.03% (mild DR), and 14.67% (active DR). Hence, the AC agents are able to reach considerable consumption cost reductions with relatively low cuts in their consumed power levels. The percentage reduction in cost is approximately between 1.5 to 2 times the percentage reduction in overall consumption of AC loads, up to the point where the load response reaches its maximum level and is saturated.

B. Retailer Agent's Performance

On the retailer side where the profit maximization problem is solved, the retailer agent's prediction MAE is shown in Fig. 9 for the mild and active DR programs. For the case of mild DR (Fig. 9a), the MAE of prediction converges to 6.08% (ANN-based model), and 9.95% (linear model). Hence, the long term MAE of the ANN-based model falls below that of the linear model. However, in the short term (iterations 50 to

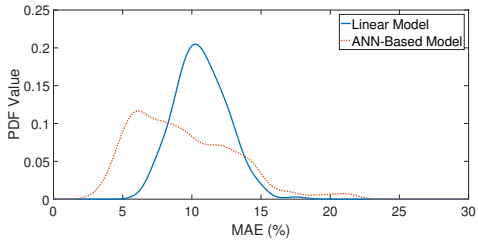


(a) Mild DR

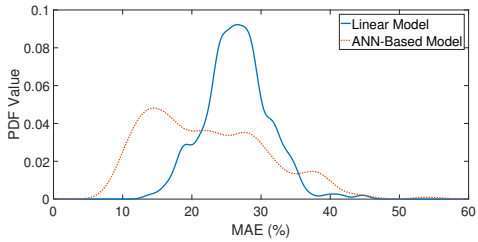


(b) Active DR

Fig. 9. Retailer prediction error.



(a) Mild DR



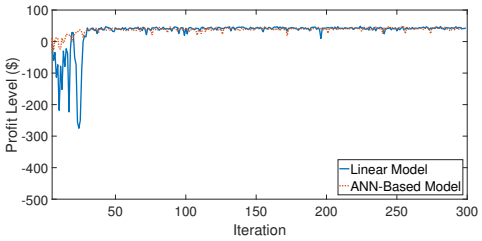
(b) Active DR

Fig. 10. The estimated PDF of the retailer's prediction error.

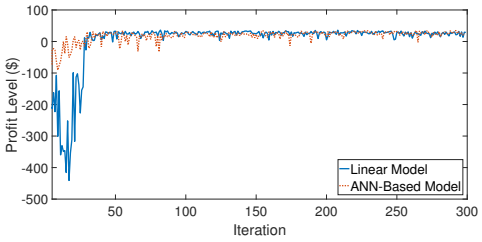
100) the linear model is able to show similar or even better performance than the ANN. The faster convergence of the linear model implies that we can get to the optimal operation point of the multi-agent system in fewer iterations compared to ANN. For the case of active DR, similar observations can be made. Here, the DA prediction errors are generally higher compared to the mild DR situations. However, as can be seen in Fig. 9b, the long term prediction MAE of the ANN is 14.66%, which is considerably lower than that of the linear model (27.89%). This suggests that as the behavior of the demand side in response to time-varying retail prices gets more uncertain, a non-linear and more powerful tool such as ANN is able to outperform the linear function approximation approach. Hence, ANN is able to better capture the response of the loads to the prices and reduce the uncertainty in the decision model that is caused by data privacy (i.e., the incomplete information of the retailer agent on the state of AC loads.) The estimated Probability Distribution Functions (PDF) of the prediction MAE of the two models are depicted in Fig. 10 for the mild and active cases. Although the prediction MAE has a symmetric, almost Gaussian shape distribution under the linear model, for the ANN-based model it is skewed. While the mean of the MAE is lower for the ANN (implying superior performance), it has a higher standard deviation (mild DR: 3.67%, active DR: 8.67%) compared to the case of the linear model (mild DR: 1.85%, active DR: 4.71%). Moreover, the standard deviations of the PDFs increase considerably for the case of active DR.

Now the question is whether the enhanced prediction accuracy of the ANN leads to monetary gains for the retailer agent. The profit level of the retailer over the iterations is shown in Fig. 11 for the cases of mild and active DR. As can be seen in the figures, after the initial phase of learning, where the retailer is collecting enough samples to address the problem of overfitting, the profit level of the retailer agent increases and reaches its maximum amount. In the case of mild DR, the total profit of the retailer per day is \$42.4 for the linear model and \$41.3 for the ANN-based model. The mean average difference in the optimal retail price under the two models is around 1%. Hence, for the case of mild DR, the performances of the two models in terms of profit are quite close for both models, and no meaningful difference is observed (the profit under the linear model is 2.6% higher than the ANN-based model). However, as the price-responsivity of the AC agents increases (i.e., the system gets more uncertain), the superior predictive capability of ANN leads to higher profit levels for the retailer, compared to the linear model. For the case of active DR, the total profit of retailer per day is \$26.4 under the ANN-based model and \$23.9 under the linear model. Hence, using the ANN-based model leads to 10.5% improvement in the profit level of the retailer, compared to the linear model. The mean average difference between the optimal retail prices increases to the value of 3.1%. Also, for both DR cases, the ANN produces a more stable profit stream as is observed in the figures. Another notable result is that as the DR program gets more active (i.e., the AC agents reduce their consumption costs more aggressively), the profit level of the retailer from the sales of energy also decreases (from 7.3% of total revenue for the case of mild DR to 4.8% for active DR). This drop is not only observed in the total profit, but also in the unit profit values (i.e., profit level per sold energy unit). For comparison, the total profit level of the retailer agent under no DR from loads is equal to \$96.4, which corresponds to 14.77% of the total revenue. The reduction in the profit level of the retailer, as the DR program gets more active, is due to the overall reductions in the cost of power consumptions as the consumers strategically modify their consumption profile in response to the retail prices they receive. This implies that as a result of the DR program, the consumers will be less captive to the actions of the profit-oriented retailers in the markets and are able to affect the equilibrium of the retail market to their benefit.

The final results of the optimization problems (i.e., optimal



(a) Mild DR



(b) Active DR

Fig. 11. Retailer's profit throughout the iterations.

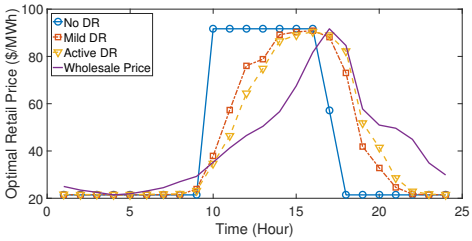


Fig. 12. Optimal retail prices along with the wholesale price signal.

retail prices) are shown in Fig. 12, along with the input DA wholesale price signal (adopted from PJM data history [34]). By comparing the three cases of price-sensitivity (i.e., no-DR, mild DR, and active DR), it is observed that increased sensitivity of the consumers to consumption costs leads to smoother retail price signals. In Fig. 13 the correlation values between the wholesale and retail prices are shown. As can be seen, the correlation level between the two markets tend to increase as the demand response program gets more active.

C. Peak Shaving

Due to the shifting of AC loads towards the later hours of the day with lower energy prices, a minor secondary peak is created in the load profile around 19:00 PM to 21:00 PM,

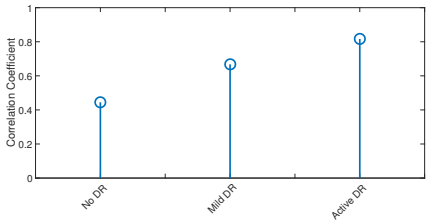


Fig. 13. Correlation between the wholesale and retail prices as a function of price-sensitivity of loads.

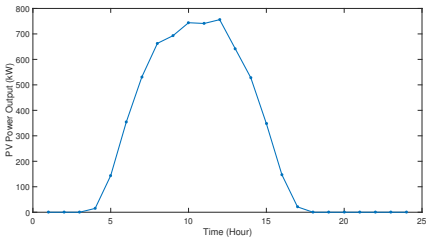


Fig. 14. PV power profile in the distribution feeder.

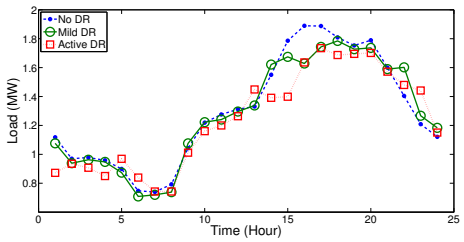


Fig. 15. Feeder load profile, considering PV power generation under different DR scenarios.

as shown in Fig. 8. While this minor peak is smaller than the main original peak of the load, in distribution systems with high penetration of PV generation the minor peak can contribute to network congestion and overloading (since there is a sharp drop in PV power around this period). Hence, we have also addressed the problem of peak load reduction using the dynamic pricing scheme in a distribution feeder with a high penetration of PV power (35% of the peak load). The PV power profile used for the simulations is adopted from [35], and is shown in Fig. 14.

In this section simulations are performed, assuming that the goal of the retailer is to minimize peak load through optimal energy pricing by solving (12). The results show that employing the proposed pricing scheme in the distribution system with PV penetration, the retailer is able to reduce the estimated peak load value from 1.89 MW to 1.78 MW (mild DR) and 1.73 MW (active DR), as shown in Fig. 15. This implies that with 35% of the total load being price-sensitive AC loads in the distribution system, the peak load can be reduced by 5.5% (mild DR), and 8.5% (active DR). Hence, as the DR program gets more active higher levels of reduction in the peak load are achieved. Also, the peak-to-average load ratio (load factor) of the distribution feeder has improved from a value of 1.45 to around 1.39. It can also be observed from Fig. 15 that employing the proposed retail energy pricing mechanism to reduce peak load by the retailer does not lead to secondary peaks (unlike the original problem of the profit maximization).

VI. Conclusion

In this paper, we introduced an agent-based framework for studying the behavior of a DA retail market with DR from AC loads. The proposed approach employs machine learning techniques to model the behavior of the agents at different levels of the hierarchical framework. Q-learning is employed

to solve the decision-making problem of the consumers. On the retailer side, different techniques (linear modeling and ANN-based modeling) are compared with each other, based on the linearity and non-linearity of the developed model by the retailer. Due to the modular characteristic of the proposed model, the framework can be generalized easily to include more complex and advanced models, without a need for significant changes in its basic functionality. The numerical results show that through this framework the consumers are able to cut their consumption costs, while the retailer maximizes its profit from the sales of energy, subject to the behavior of the loads in terms of cost-sensitivity. Also, the results suggest that as the penetration level of price-sensitive appliances increases in the system (which leads to higher uncertainty), it would be beneficial (in terms of revenue) for the profit-oriented retailer to employ more advanced (non-linear) tools, such as ANNs, instead of a linear method, to capture the behavior of the consumers. As has been demonstrated in the paper, the same pricing mechanism can also be applied to reduce load peak value. The simulation results for high penetration of PV power in the retail market suggest that as the DR program gets more active, higher levels of peak reduction are observed.

REFERENCES

[1] M. E. Kantarci and H. T. Mouftah, "Energy efficient information and communication infrastructures in the smart grid: a survey on interactions and open issues," *IEEE Communication Survey & Tutorials*, vol. 17, no. 1, pp. 179–197, 2015.

[2] S. A. Pourmousavi and M. H. Nehrir, "Real-time central demand response for primary frequency regulation in microgrids," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1988–1996, Dec. 2012.

[3] "FERC staff report: Assessment of demand response and advanced metering - Dec. 2012," https://www.ferc.gov/legal/staff-reports/12-20-12-demand-response.pdf.

[4] D. S. Kirschen, "Demand-side view of electricity market," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 179–197, May 2003.

[5] B. Moradzadeh and K. Tomsovic, "Two-stage residential energy management considering network operational constraints," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2339–2347, Dec. 2013.

[6] L. P. Qian, Y. J. Zhang, J. Huang, , and Y. Wu, "Demand response management via real-time electricity price control in smart grids," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1268–1280, Jul. 2013.

[7] "FERC staff report: Assessment of demand response and advanced metering - Dec. 2015," https://www.ferc.gov/legal/staff-reports/2015/demand-response.pdf.

[8] Y. Y. Hong, J. K. Lin, C. P. Wu, and C. C. Chuang, "Multi-objective air-conditioning control considering fuzzy parameters using immune clonal selection programming," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1603–1610, Dec. 2012.

[9] A. Safdarian, M. F. Firuzabad, and M. Lehtonen, "Integration of price-based demand response in discos short-term decision model," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2235–2245, Sep. 2014.

[10] C. Chen, J. Wang, Y. Heo, , and S. Kishore, "MPC-based appliance cheduling for residential building energy management controller," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1401–1410, Sep. 2013.

[11] S. Li, D. Zhang, A. B. Roget, , and Z. ONeill, "Integrating home energy simulation and dynamic electricity price for demand response study," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 779–788, Mar. 2014.

[12] P. Samadi, H. Mohsenian-Rad, V. W. S. Wong, and R. Schober, "Real-time pricing for demand response based on stochastic approximation," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 789–798, Mar. 2014.

[13] "European Information & Communication Technology Association (EICTA): Interoperability white paper - Jun. 2004," http://agoria.be/www1.wsc/webextra/prg/.

[14] S. Widergren, A. Levinson, J. Mater, and R. Drummond, "Smart grid interoperability maturity model," in *IEEE Power and Energy Society General Meeting*, Minneapolis, MN, 2010, pp. 1–6.

[15] M. Wooldridge, *An introduction to multiagent systems*. New Jersey: John Wiley & Sons, 2009.

[16] A. Weidlich and D. Veit, "A critical survey of agent-based wholesale electricity market models," *Energy Economics*, vol. 30, no. 4, pp. 1728–1759, 2008.

[17] A. Molina-Garcia, F. Bouffard, and D. S. Kirschen, "Decentralized demand side contribution to primary frequency control," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 411–419, Feb. 2011.

[18] M. Rahimiyan, L. Baringo, and A. J. Conejo, "Energy management of a cluster of interconnected price-responsive demands," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 645–655, Mar. 2014.

[19] D. Menniti, F. Costanzo, N. Scordino, and N. Sorrentino, "Purchase-bidding strategies of an energy coalition with demand response capabilities," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1241–1255, Aug. 2009.

[20] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. New Jersey: Prentice-hall, 1974.

[21] D. J. C. MacKay, "A practical bayesian framework for backpropagation networks," *Neural Computation*, vol. 4, no. 3, pp. 448–472, May 1992.

[22] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," *Evolutionary Computation Proceedings, IEEE*, 1998.

[23] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. Cambridge: MIT press, 1998.

[24] S. Bashash and H. K. Fathy, "Modeling and control of aggregate air conditioning loads for robust renewable power management," *IEEE Trans. Control. Syst. Technol.*, vol. 21, no. 4, pp. 1318–1327, Jul. 2013.

[25] R. Baldick, *Applied optimization: formulation and algorithms for engineering systems*. Cambridge: Cambridge University Press, 2006.

[26] H. Allcott, "Real-time pricing and electricity markets," *Harvard University*, pp. 1–77, Jan. 2009.

[27] S. Samarasinghe, *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*. Boca Raton: CRC press, 1992.

[28] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, May 1992.

[29] A. R. Jordehi, "A review on constraint handling strategies in particle swarm optimization," *Neural Computing and Applications*, vol. 26, no. 6, pp. 1265–1275, Aug. 2015.

[30] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 525–536, Mar. 1998.

[31] "Weather Underground," https://www.wunderground.com/history.

[32] NREL, "Randomized hourly load data for use with taxonomy distribution feeders," https://catalog.data.gov/harvest/object/, Nov. 2015.

[33] A. Hoke, R. Butler, J. Hambrick, , and B. Kroposki, "Steady-state analysis of maximum photovoltaic penetration levels on typical distribution feeders," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 350–357, Apr. 2013.

[34] "PJM market data," http://www.pjm.com/.

[35] "Electric Power Research Institute (EPRI): Distributed PV monitoring and feeder analysis - Jun. 2012," http://dpv.epri.com/measurement_data.html.