Reinforcement Learning-Based Demand Response Management in Smart Grid Systems With Prosumers

Fisayo Sangoleye, *Student Member, IEEE*, Jenilee Jao, *Student Member, IEEE*, Kimberly Faris, Eirini Eleni Tsiropoulou, *Senior Member, IEEE*, and Symeon Papavassiliou, *Senior Member, IEEE*

***Abstract*—In this article, we introduce a reinforcement learning-based price-driven demand response management (DRM) mechanism in smart grid systems consisting of prosumers. Our proposed approach accounts for the prosumers' behavioral characteristics and models the emerging interactions among all the involved actors in the smart grid system, i.e., prosumers, energy management system (EMS), and utility companies. In particular, an off-policy reinforcement learning is introduced enabling the EMS to determine the optimal price that should be announced to the prosumers on an hourly-basis toward minimizing the overall system's cost. In this process, the utility companies' hourly-based wholesale price and the prosumers' energy generation and consumption characteristics are considered as input. At the same time, the prosumers' optimal amount of purchased energy is determined in a real-time manner. The presented numerical results demonstrate the success of the proposed DRM model to deal with the incomplete information availability scenarios, regarding the prosumers' energy selling and purchasing patterns, compared to the state of the art. Also, the detailed comparative evaluation against other price-based DRM approaches, e.g., cap-based and day-ahead pricing, shows the benefits of the proposed DRM model in terms of adapting in a real-time manner to the prosumers' energy demand, while jointly minimizing the overall system's long-term cost.**

***Index Terms*—Decision-making, demand response management (DRM), prosumers, reinforcement learning, smart grid systems, system modeling.**

I. Introduction

SMART grid systems are novel power-grid systems which are capable to sense and measure the consumer's power consumption by exploiting a smart metering infrastructure enabled by advanced communication and information technologies [1]. Smart grid systems are expected to orchestrate the energy generation, consumption, and conservation by implementing intelligent demand response management (DRM) mechanisms. The ultimate goal of DRM mechanisms is to efficiently balance the energy supply and demand, thus leveling the energy consumption during peak hours [2].

Current studies on DRM mechanisms have considered smart grid systems consisting either of consumers [3] simply purchasing energy from the smart grid system, or prosumers, who can purchase energy from the grid and/or generate energy via using renewable energy sources [4]. The existing DRM mechanisms can be categorized as: 1) price-based, where the consumers' energy consumption is controlled via the announced energy prices by the utility companies [5], and 2) incentive-based, where the prosumers' energy consumption behavior is guided via appropriate designed incentives [6]. Also, the main theoretical tools that have been used to design DRM mechanisms are: game theory, capturing the interdependencies among the consumers' energy consumption decisions [7], and learning-based approaches enabling the grid operator to forecast the energy consumption patterns [8].

In this article, we introduce a reinforcement learning-based price-based DRM mechanism in a smart grid system consisting of prosumers, while accounting for the prosumers' characteristics, and the interactions among all the involved entities in the smart grid system. Two main markets are studied to capture all the involved entities interactions in the smart grid system, i.e., the wholesale market where the energy management system (EMS) purchases energy from the utility companies, and the retail market, where the prosumers purchase energy from the EMS. Our ultimate goal is to minimize the overall long-term smart grid system's cost via determining the optimal announced energy price by the EMS to the prosumers, while the latter ones dynamically adapt and determine their optimal energy consumption following the proposed price-based DRM mechanism.

A. Related Work

Game theory has been widely used in the recent literature to design DRM mechanisms [9]. In [10], the authors formulate the interactions among the distributed power sources, the energy storage devices, and the consumers as a noncooperative game among them, where each entity aims at maximizing its profit by determining its optimal energy generation or consumption,

Manuscript received 27 October 2021; revised 6 May 2022 and 29 January 2023; accepted 17 February 2023. This work was supported by the National Science Foundation EPSCoR Program under Award #OIA-1757207. *(Corresponding author: Eirini Eleni Tsiropoulou.)*

Fisayo Sangoleye, Jenilee Jao, Kimberly Faris, and Eirini Eleni Tsiropoulou are with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131 USA (e-mail: fsangoleye@unm.edu; jenjao@unm.edu; kimberly.faris@cosmiac.org; eirini@unm.edu).

Symeon Papavassiliou is with the School of Electrical and Computer Engineering, National Technical University of Athens, 15780 Athens, Greece (e-mail: papavass@mail.ntua.gr).

Digital Object Identifier 10.1109/JSYST.2023.3248320

respectively. The authors prove that the game is an exact potential game and show the existence and uniqueness of a Nash equilibrium that enables the overall smart grid system to operate at a stable point. An aggregate game among the consumers is formulated in [11], enabling the consumers to collaborate among each other in order to determine their optimal aggregate energy consumption following a price-based DRM model. A price-based DRM mechanism following the hourly billing model is introduced in [12]. The consumers' optimal energy consumption is derived as the Nash equilibrium of the noncooperative game among the consumers who aim to maximize their satisfaction by purchasing energy. A social pricing mechanism is proposed in [13] that is derived from the overall energy demand in the smart grid system. Toward determining the optimal energy consumption of each consumer, a noncooperative game is formulated among them and a corresponding equilibrium is calculated. Also, both cases of rational and risk-aware consumers are studied based on the principles of expected utility theory and prospect theory, respectively. Focusing on the multienergy interactions, a trilayer multienergy day-ahead market structure and operation mechanism is introduced in [14], to support the trading of electricity, heat, and natural gas. In [15], a novel evolutionary game model is proposed based on the bounded rationality of bidders to support the consumers to determine the optimal demand response bidding strategy under scenarios of incomplete information.

The special category of *Stackelberg games* has been extensively used in the literature to study DRM given the inherent property of those games to handle the hierarchical relationship between the electricity market and the prosumers [16]. A Stackelberg game is formulated among the utility companies (leaders) and the consumers (followers) aiming at maximizing the utility companies' profit and the consumers' welfare, under several different underlying goals. Such objectives include the following.

1) Determining the unique optimal number of utility companies to maximize their profit [17].
2) Enabling the prosumers to select the most beneficial utility company based on a reinforcement learning mechanism in order to optimize their long-term welfare [18].
3) Maximizing the utility companies' trading probability with the consumers, while the latter ones optimize their flexible loads during the day [19].
4) Using the price-based DRM mechanism as an incentivization scheme to motivate the consumers to participate in the DRM program and determine the optimal reduction of energy consumption [20].
5) Coordinating the renewable energy share among the prosumers who aim at maximizing their welfare by selling or buying a corresponding amount of renewable energy [21].
6) Optimizing the number of transactions with the prosumers [22].
7) Minimizing the cost both for the microgrid operator (MGO) and the consumers [23].

Learning mechanisms have recently attracted the interest of the academic community in order to deal with DRM problems in smart grid systems. A price-based DRM mechanism is introduced in [24] via designing a deep reinforcement learning model based on a dueling deep Q network structure. The proposed model optimizes the energy exchange regarding interruptible loads considering the time of use tariff and different energy consumption patterns for the consumers. Also, in [25], a learning model is proposed to enable the MGO to learn the prosumers energy consumption patterns regarding the home heating, ventilation, and air conditioning energy needs. Then, the MGO can design the optimal demand response policies to maximize its profit and satisfy the consumers' energy prerequisites. A deep reinforcement learning approach is introduced in [26] in order to jointly optimize the charging scheduling, order dispatching, and vehicle rebalancing for large-scale shared electric vehicles fleet operator.

B. Contributions and Outline

Following the above discussion and analysis, we should stress that the problem of incomplete information availability and scenarios, regarding the prosumers' energy selling or purchasing patterns, has not been properly addressed and still remains an open and unresolved issue [27]. Even the recently adopted and applied learning mechanisms mainly focus on learning the prosumers' energy exchange patterns, and do not consider the MGO's long-term optimal announced price while accounting for the prosumers' characteristics. Moreover, the game-theoretic DRM mechanisms though offering interesting results, still suffer from the drawback of excessive communications overhead between the MGO and the prosumers in order to conclude to an optimal announced price for the MGO and an optimal energy exchange pattern for the prosumers.

In this research work, we strive exactly to tackle these issues and drawbacks, by proposing a reinforcement learning-based price-driven DRM mechanism in smart grid systems consisting of prosumers. In a nutshell, an electricity management system (EMS) coordinates the energy exchange among the utility companies and the prosumers aiming at minimizing the overall system cost, including both the prosumers' and the EMS's cost. The minimization of the overall system's cost is achieved by determining the optimal retail price announced by the EMS per hour of the day, thus, introducing a price-based DRM mechanism accounting for the prosumers' characteristics, and the interactions among all the involved entities in the smart grid system, i.e., utility companies, EMS, and prosumers. A thorough evaluation of the proposed framework is performed by using real data for the years of 2020–2021 from the U.S. Energy Information Administration. The main contributions of this research work that differentiate it from the rest of the existing literature are summarized as follows.

1) A novel model and architecture of a smart grid system is introduced consisting of the utility companies, the energy management systems (EMS), and the prosumers, while identifying and properly reflecting all the interactions among the actors of the involved markets. The utility companies set their wholesale energy selling price to the EMS, and the latter one determines the optimal retail energy selling price to the prosumers in order to minimize the overall system's cost and bring the overall smart grid system into a stable operation point.

2) The prosumers' and the EMS's characteristics are captured in order to define, determine, and update in a real-time manner, their experienced cost from the energy exchange. Specifically, the prosumers' cost for buying energy from the EMS, as well as its experienced dissatisfaction from postponing its energy purchase to a future time due to the increased price, are captured in realistic functions. Also, the EMS's cost for purchasing energy from the utility companies in the wholesale market, and its profit from selling energy to the prosumers in the retail market, are represented in properly formulated cost functions.
3) A reinforcement learning-based price-based DRM mechanism is designed to determine the EMS retail energy price per hour of the day toward minimizing the overall cost experienced in the smart grid system. The proposed mechanism considers as input the utility companies' hourly-based wholesale price, and the prosumers' energy generation and consumption characteristics. The latter information can be derived from national catalogues. In our case, we have used real data from the U.S. Energy Information Administration online available datasets.
4) A detailed set of numerical results stemming from real datasets show the performance and operation of the proposed reinforcement learning price-based DRM mechanism. In addition, the reaction of the prosumers to the retail energy price is studied considering different prosumers' behavioral patterns regarding the postponement of their energy consumption for a future time. Finally, a detailed set of comparative results to alternative existing DRM mechanisms, reveals the benefits and tradeoffs of our proposed framework.

The rest of this article is organized as follows. Section II describes the introduced smart grid system's architecture and the considered system model, while Section III captures the EMS and the prosumers' characteristics. Subsequently, in Section IV the proposed reinforcement learning price-based DRM mechanism is introduced and its operation is analyzed. Detailed numerical and comparative results, obtained via modeling and simulation, are provided and discussed in Section V. Finally, Section VI concludes this article.

II. Prosumers-Based Smart Grid System Overview

A smart grid system is considered, consisting of the utility companies, the EMS, and the prosumers. Those three different types of entities create an energy market, while a simplified illustration of the wholesale energy market and the retail energy market components are presented in Fig. 1. Each prosumer is equipped with an advanced metering infrastructure (AMI) and an energy management controller (EMC). The EMC enables the scheduling of the energy usage for the corresponding prosumer given its energy needs and its personal flexibility to postpone part of its needs. The AMI is used to support the bidirectional communication between the prosumer and the EMS, where the latter one provides energy to the prosumers through a retail energy market. Specifically, the EMS buys energy from the utility companies through a wholesale energy market, where the utility companies sell energy at an hourly dynamic wholesale price. Then, the EMS sells energy to the prosumers by using an hourly dynamic retail price. We study the interactions among the utility companies, the EMS, and the prosumers at each time slot. Toward capturing the time-dependency in the involved entities' interactions, we introduce the set of periods $\mathcal{H} = \{0, 1, \ldots, H-1\}$, where each period represents a real time (i.e., hour) of the day. Practically, without loss of generality, we can consider that the periods represent the actual hours of the day, i.e., $H = 24$. Then, we map each time slot t to a corresponding period based on $h^t = \mathrm{mod}(t, H), t \geq 0$.

Focusing on the prosumers, their corresponding set is denoted as $N = \{1, \ldots, n, \ldots, |N|\}$. Each prosumer's energy consumption can be divided into two categories: 1) shiftable, i.e., elastic needs, such as water heating, electric vehicle charging, etc., and 2) nonshiftable, i.e., nonelastic needs, such as refrigerator, alarm system, etc. [28]. The corresponding set of appliances of each prosumer n that consume energy is denoted as $A_n = \{1, \ldots, a_n, \ldots, |A_n|\}$. At time slot t, an appliance can be on, i.e, $\delta^t_{a_n} = 1$, or off, i.e., $\delta^t_{a_n} = 0$. Thus, the cumulative energy demand at time slot t is given as follows:

$$d_n^t = \sum_{\forall a_n \in A_n} \delta^t_{a_n} E^t_{a_n} \text{ [kWh]} \tag{1}$$

where $E^t_{a_n}$ [kWh] is the energy consumption of the appliance a_n in time slot t. Given the retail price announced by the EMS in a given period h, where time slot t belongs to a corresponding period h^t the prosumer determines the amount of energy that it will purchase in order to minimize its cost, the latter consisting of the monetary cost, as well as its experienced dissatisfaction by postponing part of its energy needs for a future time. Simultaneously, the prosumer is capable of producing energy based on renewable sources of energy, such as solar photovoltaic panels. The prosumer's energy generation in time slot t is denoted as g_n^t[kWh]. The prosumer's energy generation and demand characteristics can be extracted from available national datasets. In our article, we have used real data for the years of 2020–2021 from the U.S. Energy Information Administration [29].

At each time slot t, each prosumer decides whether or not to enter the retail energy market in order to purchase energy from the EMS based on its energy generation and demand characteristics. It is noted that each prosumer is equipped with an energy storage device, e.g., Lithium-ion batteries, where b_n^t[kWh] denotes the available stored energy at time slot t. We also assume that, in the realistic scenarios considered here, the storage devices suffice to store the prosumers' energy surplus, without exceeding their physical limits. Therefore, if $g_n^t + b_n^{t-1} \geq d_n^t$, then the prosumer can cover its energy needs based on its own generated energy and its already available energy at its storage devices. In this scenario, the prosumer does not enter the retail energy market. Also, the prosumer's energy surplus is stored for future usage, i.e., $b_n^{t+1} = b_n^t + (g_n^t - d_n^t)$. In the opposite scenario, if $g_n^t + b_n^{t-1} < d_n^t$, then the prosumer needs to buy energy from the EMS in order to cover part or all of its energy needs. Let us denote as e_n^t[kWh] the amount of purchased energy from the EMS.

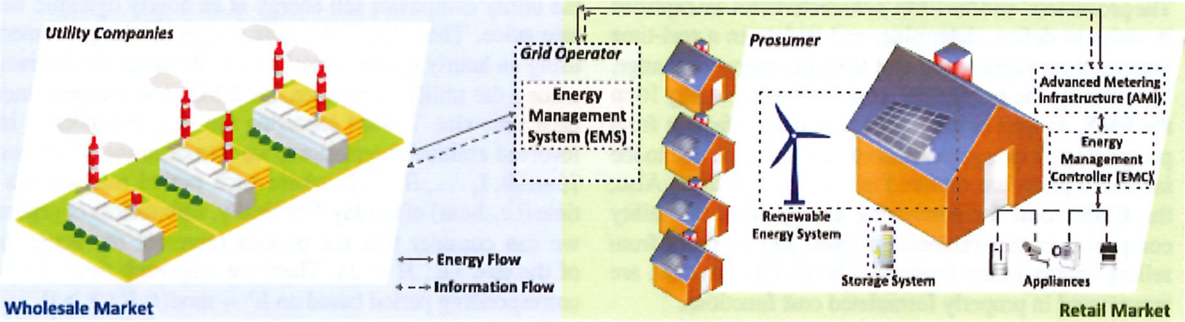


Fig. 1. Illustration of the interactions among the utility companies, the energy management system (EMS), and the prosumers.

It is highlighted that the prosumer will actually purchase an amount of energy e_n^t that minimizes its experienced monetary cost and dissatisfaction from postponing part of its energy needs for a future time slot, based on the retail energy price announced by the EMS. Let us denote as

$$l_n^t = d_n^t - b_n^{t-1} - g_n^t \text{ [kWh]} \tag{2}$$

the total amount of energy that the prosumer n may potentially purchase from the EMS. Then, for the actual amount of energy that the prosumer purchases, it holds true that $e_n^t \leq l_n^t$.

In the next section, we capture the prosumers' and the EMS's characteristics, while appropriately defined cost functions from the energy exchange are introduced in order to capture the interactions of the utility companies, the EMS, and the prosumers.

III. Prosumers' and Electricity Management System's Characteristics and Interactions

The prosumer's dissatisfaction from postponing part of its energy needs for a future time slot is denoted as $\mathcal{D}_n(l_n^t - e_n^t)$, where $\mathcal{D}_n : \mathbb{R}^+ \to \mathbb{R}^+$ and is called dissatisfaction function. The dissatisfaction function is a strictly increasing function with respect to the amount of energy requirement, that the prosumer decides to postpone for a future time slot. For presentation purposes in this article, we capture the dissatisfaction function, as follows:

$$\mathcal{D}_n(l_n^t - e_n^t) = s_n^h(l_n^t - e_n^t)^{x_n} \tag{3}$$

where $x_n \in \mathbb{R}^+, s_n^h \in \mathbb{R}^+$. The sensitivity parameters s_n^h and x_n capture the prosumer's level of dissatisfaction by postponing part of its energy consumption, reflecting and offering for a more personalized perspective. Specifically, a prosumer is more sensitive in terms of its experienced dissatisfaction by postponing part of its energy consumption for greater values of the sensitivity parameters. Also, the time-dependent sensitivity parameter s_n^h captures the variability of the prosumer's dissatisfaction based on the peak (or low) energy demand hours of the day. Specifically, a prosumer is more sensitive regarding its dissatisfaction if it postpones its energy demand during a peak hour, where it may have greater need to purchase and use the necessary energy.

Furthermore, as highlighted before, the prosumer is charged for the amount of energy e_n^t that it decides to purchase for covering its needs, which is also another factor affecting its perceived experience. Therefore, the overall cost that the prosumer experiences by postponing part of its energy usage and/or purchasing an amount of energy, is given as follows:

$$PC_n^t(l_n^t, e_n^t) = \mathcal{D}_n(l_n^t, e_n^t) + S^t(e_n^t) \tag{4}$$

where $S^t(e_n^t) = k^h \cdot e_n^t$ denotes the EMS selling function of energy to the prosumers and $k^h \in \mathbb{R}^+$[\$/kWh] denotes the retail price of energy, as announced by the EMS at a specific period (i.e., hour) of the day. The selling function $S^t(e_n^t)$ represents the profit that the EMS makes by selling energy to the prosumers in the retail market. It is highlighted that, without loss of generality, we consider an hourly-based pricing model, where the EMS can dynamically announce a different retail price at every hour of the day in order to optimize its profit and minimize its experienced cost.

The goal of each prosumer is to minimize its experienced cost while accounting for the monetary cost to purchase energy and the dissatisfaction cost to postpone part of its energy needs. Thus, given the announced retail energy price by the EMS, each prosumer's goal is to determine its optimal amount of purchased energy e_n^{t*} in order to minimize its experienced cost, while considering its nonshiftable and shiftable energy needs. The optimal solution of the corresponding optimization problem is given as follows:

$$e_n^{t*}(l_n^t) = \underset{0 \leq e_n^t \leq l_n^t}{\operatorname{argmin}} \; PC_n^t(l_n^t, e_n^t). \tag{5}$$

Following, and focusing on the characteristics of the EMS, we study the EMS interactions with the utility companies and the prosumers. As explained before, the EMS buys the energy from the utility companies in a wholesale market and sells energy to the prosumers in a retail market. The utility companies announce an hourly-based wholesale price of the energy based on collected data and statistics on the energy demand and a cap-based price accounting for the energy demand to improve their profit and avoid peaks of energy demand. Therefore, the wholesale billing function announced by the utility companies to the EMS is given as follows:

$$b^t\left(\sum_{\forall n \in N} e_n^{t*}(l_n^t)\right) = a^t \sum_{\forall n \in N} e_n^{t*}(l_n^t) + \beta_{h^t}^t \left(\sum_{\forall n \in N} e_n^{t*}(l_n^t)\right)^2 \tag{6}$$

where $a^t \in \mathbb{R}^+[\$/\text{kWh}]$ denotes the wholesale hourly-based price and $\beta^t_{h^t}$ denotes the cap-based energy price, which is a random variable, whose expected value changes in an hourly-basis following the total energy demand. In practical scenarios, the wholesale hourly-based price and the cap-based energy price are announced by the utility companies at the point of the energy exchange among them and the EMS, in the wholesale market. Our proposed model could be easily extended by introducing a different wholesale billing function per utility company selling energy to the EMS, for any corresponding amount of requested energy.

The EMS collects revenue by selling energy to the prosumers at a retail price k^h, where the price as explained before is set in a dynamic manner following an hourly-basis pricing model. Thus, the overall cost experienced by the EMS is derived as follows:

$$EC^t(\mathbf{l}^t, \mathbf{b}, \mathbf{k}) = b^t \left(\sum_{\forall n \in N} e^{t*}_n(l^t_n) \right) - \sum_{\forall n \in N} S^t(e^{t*}_n(l^t_n)) \quad (7)$$

where $\mathbf{l}^t$ is the prosumers' vector of the amount of needed energy at time slot t, $\mathbf{b}$ denotes the ulility companies' wholesale price vector, and $\mathbf{k}$ is the EMS's retail price vector. It is noted that the overall cost experienced by the EMS, as expressed in (7), depends on the EMS's retail price vector $\mathbf{k}$, given that the selling function $S^t(e^{t*}_n(l^t_n))$ depends on the retail energy price k^h.

Considering the prosumers' cost (4) and the EMS cost (7), the overall cost experienced in the smart grid system at each time slot t is determined as follows:

$$SC^t(\mathbf{e}^t, \mathbf{k}) = w_1 EC^t(\mathbf{l}^t, \mathbf{b}, \mathbf{k}) + w_2 \sum_{\forall n \in N} PC^t_n(l^t_n, e^t_n) \quad (8)$$

where $w_1, w_2 \in \mathbb{R}^+, w_1 + w_2 = 1$ denote the corresponding weight to value more the EMS cost (system-centric approach) or the prosumers' cost (prosumer-centric approach). It is noted that the weights w_1, w_2 are introduced to provide the enhanced flexibility to the smart grid system to weigh more the EMS cost or the prosumers' cost based on the examined use case scenarios. In the general case, where those costs are treated with the same importance, these weights are assumed equal. Also, $\mathbf{e}^t$ denotes the vector of purchased energy by the $|N|$ prosumers at time slot t.

IV. Reinforcement Learning Enabling Demand Response Management

In this section, capitalizing on the above presented modeling and formulation, we introduce and design a reinforcement learning price-based DRM approach. In particular, the proposed DRM mechanism exploits the theory of reinforcement learning in order to enable the EMS to determine the optimal price k^h per hour of the day toward minimizing the overall cost in the smart grid system, as defined in (8). At the same time, the prosumers also determine their optimal amount of purchased energy $e^{t*}_n \forall n \in N \forall t$, by responding to the announced retail price by the EMS in order to minimize the overall system's cost. Toward achieving this goal, an off-policy reinforcement learning algorithm is used to enable the EMS learn the optimal prices in order to minimize the long-term system's cost. Specifically, the EMS executes the proposed off-policy reinforcement learning algorithm and announces the retail price to the consumers at each iteration of the algorithm.

Specifically, we formulate the system's cost minimization problem as a Markov decision process (MDP) problem. The MDP problem is defined by the decision maker's actions, a set of states for the smart grid system, and the overall system's cost function. In our examined problem, the decision maker is the EMS, and its actions are the announced retail prices k^h in an hourly-basis. In the smart grid system under consideration, the state s^t of the system at time slot t is defined as a combination of the current period h^t, the amount of energy that the prosumers can potentially purchase $\mathbf{l}^t$, and the wholesale pricing function b^t, as follows:

$$s^t = (h^t, \mathbf{l}^t, b^t). \quad (9)$$

Focusing on a realistic smart grid system implementation we consider $H = 24$ hours. Also, we quantize the maximum feasible intervals of the total amount of energy that the prosumers can potentially purchase and the wholesale prices into L and B levels, respectively. This information can be derived by historical statistical data without requesting to reveal any sensitive information from the prosumers' side [29], as only their total amount of purchased energy influences the retail market price. Given that the transition of the total amount of energy that the prosumers can purchase and the wholesale price from the one period to the next one depend on the state s^t and the action k^h, then the sequence of the states $s^t, t = 0, 1, 2, \ldots$ follows an MDP with action k^h. Specifically, based on the action (i.e., announced retail price k^h by the EMS to the prosumers) taken at each iteration of the proposed reinforcement learning algorithm, the prosumers determine their optimal amount of purchased energy [based on (5)] and the EMS is driven to a new state s^t (9).

Based on the previous analysis, we develop an off-policy reinforcement learning algorithm, based on the principles of Q-learning, by focusing on the effect of performing an action k^h in the state s^t. We define the Q value based on the following relation:

$$Q(s, k) = Q(s, k) + \alpha(r + \gamma \max Q(s', k) - Q(s, k)) \quad (10)$$

where α denotes the learning rate of the algorithm capturing how thoroughly our proposed model explores the available states of the examined system, and γ is the discount factor capturing the importance of future rewards. The factor r is the reward from moving from one state s^t to the next one s^{t+1} and it is formulated as being inversely proportional to the experienced system cost. To determine the optimal prices announced by the EMS per period h^t, and the corresponding optimal amount of energy purchased by each prosumer per time slot t, we follow the subsequent steps.

1) *First*, we initialize the Q function (10) at some arbitrary values.
2) *Second*, we select an action k^h using the ϵ-greedy policy and move to the new state. Based on the ϵ-greedy policy, we select the action that gives the maximum Q value with probability $(1 - \epsilon)$, and we explore a randomly selected action with probability ϵ. Following this process,

we enable the EMS to probabilistically select actions that provide low overall cost in the system, while also allowing the EMS to explore pricing actions that could potentially lead to a decrease in the overall system's cost in the long-term. Furthermore, in order to keep a balance between exploration of new actions and the convergence time to an optimal strategy, we adopt a decaying ϵ-greedy policy. Following this learning policy, our algorithm starts exploring the available actions with high ϵ values, e.g., $\epsilon = 0.9$, to aggressively learn the possible actions and corresponding outcomes. Then, as the iterations proceed, the ϵ value follows a decaying rule to balance the exploration and convergence time.

3) *Third*, we update the Q value of a previous state according to the Q-learning update rule, as presented in (10).
4) *Finally*, we repeat the second and the third step in our algorithm until we reach a terminal state and the algorithm has converged.

It is highlighted that at the second step of our algorithm, the EMS selects an action k^h and announces the retail price to the prosumers. Then, the latter ones determine their optimal amount of purchased energy e_n^{t*} by solving the optimization problem presented in (5).

It is noted that our proposed reinforcement learning price-based DRM mechanism enables the interactions among the utility companies, the EMS, and the prosumers in an autonomous and distributed manner, without the need of a centralized entity taking optimal decisions about the examined system. The developed DRM model can dynamically adapt to the conditions of the smart grid system, and appropriately conclude to the optimal announced hourly-based prices by the EMS and the corresponding optimal amount of purchased energy by each prosumer.

V. Numerical Results

In this section, a detailed numerical evaluation is presented to study the performance and the inherent characteristics of the proposed reinforcement learning price-based DRM model for smart grid systems with prosumers. Initially, in Section V-A, the pure operation and the performance of the proposed framework is presented. Section V-B studies the operation of the smart grid system and the effectiveness of the proposed DRM mechanisms under different scenarios of the prosumers' behavioral characteristics. Section V-C focuses on a scalability analysis scenario, and Section V-D demonstrates a detailed comparative analysis of the proposed approach against alternative DRM strategies.

Throughout our evaluation, unless otherwise explicitly stated, we consider a smart-grid system with the following parameters: $|N| = 30$, $|H| = 24$, $\gamma = 0.95$, $\alpha = 0.1$, $w_1 = 0.4$, $w_2 = 0.6$, initial value $\epsilon = 0.9$, and final value $\epsilon = 0.001$, $a^t = 0.02$, $\beta^t_{h^t} \in [0.1, 0.6]$, $s_n^h = 0.12$, $x_n = 2$, $k^h \in [0, 1]$, $d_n^t \in [0.14, 0.23]$, $g_n^t \in [0.13, 0.20]$. The values of parameters considered in our simulations have been extracted from the real data for the years of 2020–2021 from the U.S. Energy Information Administration considering the Southwest region of USA [29]. The proposed framework's evaluation was conducted using a Dell XPS desktop with 11th Gen Intel core i9-11900 K 5.3 GHz processor, and 64 GB available RAM. In the rest of the analysis, the system cost, EMS cost, and prosumer's cost (reflected in the vertical axis of the corresponding figures) are presented in relative monetary units, while the purchased energy in relative energy units. It is noted that the prosumer's dissatisfaction is a unitless metric.

A. Pure Operation and Performance

In this section, we study the pure operation and performance of our proposed framework. We consider three different learning scenarios adopting three different decaying ϵ-greedy policies, i.e., decreasing the ϵ value every 100, 200, and 300 iterations of the proposed reinforcement learning algorithm (Section IV). Figs. 2(a)–(c) and 3(a) and (b) present the overall system's cost, the total prosumers' dissatisfaction, the total amount of purchased energy, the EMS cost, and the total prosumers' cost, respectively, as a function of the real execution time of the proposed reinforcement learning price-based DRM mechanism for the three considered scenarios. It is noted that the demonstrated execution time consists of the reinforcement learning algorithm's execution time in order to determine the EMS's pricing and the prosumers' execution time in order to determine the prosumers' optimal amount of purchased energy (5). It is highlighted that the optimization problem in (5) is solved in a distributed manner by each prosumer.

The results reveal that under the scenario where the EMS thoroughly explores its optimal actions (decay cycle = 300), i.e., optimal announced price, the minimum cost is achieved [Fig. 2(a)]. Also, under this scenario, the prosumers' cost is minimized [Fig. 3(b)] via enabling the prosumers to buy a large amount of energy [Fig. 2(c)], thus, minimize their experienced dissatisfaction [Fig. 2(b)], as the prosumers cover their energy needs. Consequently, a thorough exploration of the EMS available actions favors the prosumers, while the EMS achieves an almost zero cost [Fig. 3(a)]. The exact opposite observations are derived for the other two scenarios, i.e., decay cycle = 100 or 200, where the EMS performs less exploration of its optimal actions. Furthermore, the results demonstrate that the proposed DRM mechanism converges fast to an optimal announced price and purchased amount of energy, making it suitable for realistic implementation in a real-life scenario.

B. Prosumers Behavioral Characteristics

In this section, we assess the effectiveness of our proposed DRM mechanism to adapt to the prosumers' behavioral characteristics and their sensitivity in postponing part of their energy needs to a future time slot. Initially, we consider four, scenarios regarding the prosumers' behavioral characteristics, namely: 1) Homogeneous, 2) Random, 3) Clustered, and 4) Heterogeneous. The four scenarios are differentiated based on the sensitivity parameter s_n^h (3). Under the homogeneous scenario, all the prosumers have exactly the same sensitivity $s_n^h \forall n \in N$, with respect to their dissatisfaction. Under the heterogeneous scenario, half of the prosumers' population has a very low sensitivity, while the remaining half of the population demonstrates a very high sensitivity to the dissatisfaction. Under the clustered scenario, three equal-sized clusters of prosumers are considered,

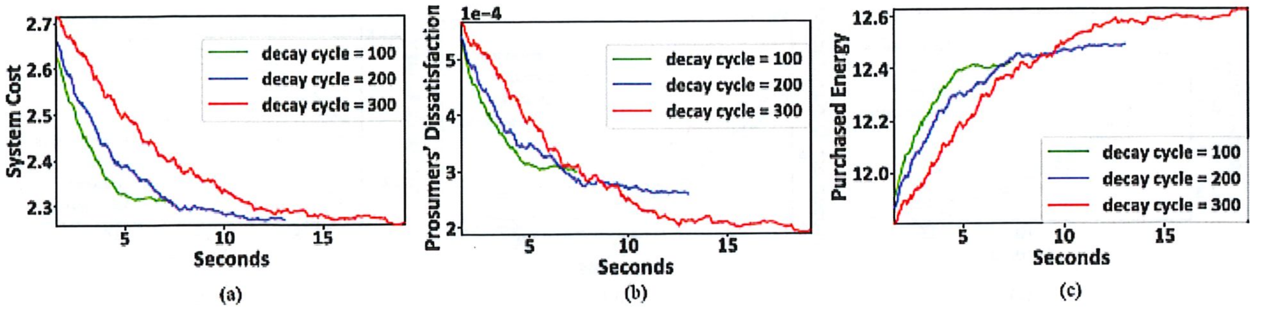


Fig. 2. (a) System cost SC^t, (b) prosumers' dissatisfaction $\sum_{\forall n\in N} \mathcal{D}_n$, and (c) prosumers' total amount of purchased energy $\sum_{\forall n\in N} e_n^{t*}$, under three different learning scenarios.

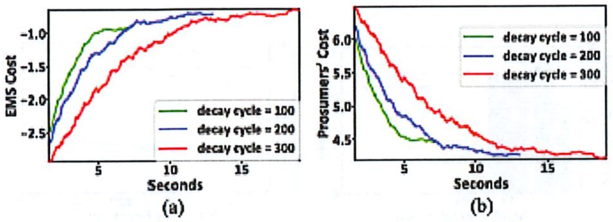


Fig. 3. (a) EMS cost EC^t, and (b) prosumers' total cost $\sum_{\forall n\in N} PC_n^t$, under three different learning scenarios.

characterized by low, medium, and high sensitivity. Under the random scenario, each prosumer is characterized by a random sensitivity with reference to its experienced dissatisfaction. It is highlighted that under the three comparative scenarios, the average prosumer sensitivity to dissatisfaction, considering all the prosumers, is the same for fairness in the comparison. Figs. 4(a)–(c) and 5(a) and (b) present the overall system's cost, the total prosumers' dissatisfaction, the total amount of purchased energy, the EMS cost, and the total prosumers' cost, respectively, as a function of the real execution time of our proposed mechanism for the three considered scenarios.

The results show that a great heterogeneity regarding the prosumers' behavioral characteristics, i.e., Heterogeneous scenario, favors the operation of the smart grid system by achieving lower cost [Fig. 4(a)], owing to the low cost achieved by the EMS [Fig. 5(a)] and the prosumers [Fig. 5(b)]. Specifically, under the scenario of high heterogeneity among the prosumers, the EMS cannot learn the unique behavioral characteristics of each prosumer, thus, it announces a holistic/generic price to all of them. The latter phenomenon brings the prosumers in an unfavorable situation, where they are not incentivized to buy a lot of energy [Fig. 4(c)], as the price for many of them is quite high. Thus, the highly heterogeneous prosumers tend to postpone a larger amount of energy to be purchased in a future time slot, and experience high dissatisfaction [Fig. 4(b)].

Further extending our analysis, we consider three comparative scenarios of 1) low, 2) medium, and 3) high sensitivity of all the prosumers regarding their dissatisfaction perceived by postponing part of their energy needs for a future time slot. Similarly, Figs. 6(a)–(b), and 7(a) and (b) present the overall system's cost, the prosumers' dissatisfaction, the total amount of purchased energy, the EMS cost, and the total prosumers' cost, respectively, as a function of the real execution time of our DRM mechanism for the three dissatisfaction sensitivity scenarios.

The results reveal that the more sensitive the prosumers are to the dissatisfaction, the more energy they purchase [Fig. 6(c)] in order to cover their energy needs, and thus, reduce their dissatisfaction value [Fig. 6(b)]. Thus, given that they buy more energy, their cost gets higher [Fig. 7(b)], mainly due to the high cost that they incur for buying a large amount of energy. Accordingly, taking into account the high energy demand, the EMS makes more profit by selling the energy [Fig. 7(a)]. In a nutshell, by jointly considering the prosumers' and the EMS cost, when the prosumers' dissatisfaction sensitivity increases, the overall system's cost becomes high as well [Fig. 6(a)].

C. Scalability Evaluation

In this section, a scalability analysis of our proposed DRM mechanism is provided for a large-scale system in terms of the number of prosumers and for a different number of states of the proposed reinforcement learning-based price-driven DRM mechanism. Fig. 8(a) presents the system's cost and the execution time of our framework, for an increasing number of states considered in the proposed reinforcement learning-based price-driven DRM mechanism, i.e., the EMS announces a retail price every 4, 2, and 1 h, respectively. The results show that as the number of states increases, the system's cost decreases, as the EMS can more accurately adapt to the prosumers' energy demand characteristics, while the execution time increases. Furthermore, the four comparative scenarios introduced at the beginning of Section V-B are considered regarding the heterogeneity of the prosumers with respect to their dissatisfaction sensitivity (i.e., Homogeneous, Random, Clustered, Heterogeneous). In particular, Fig. 8(b) presents the execution time of our proposed reinforcement learning price-based DRM mechanism as a function of the number of prosumers residing in a large-scale smart grid system (up to 1000 prosumers) for the aforementioned four scenarios. First, the results show that for a heterogeneous prosumers' population more time is required by the EMS to learn its optimal announced hourly pricing in order to minimize the overall system's cost. Second, we observe that for all scenarios the execution time grows in a much slower trend than linearly, thus overall remains relatively low and appropriate for a real-life application.

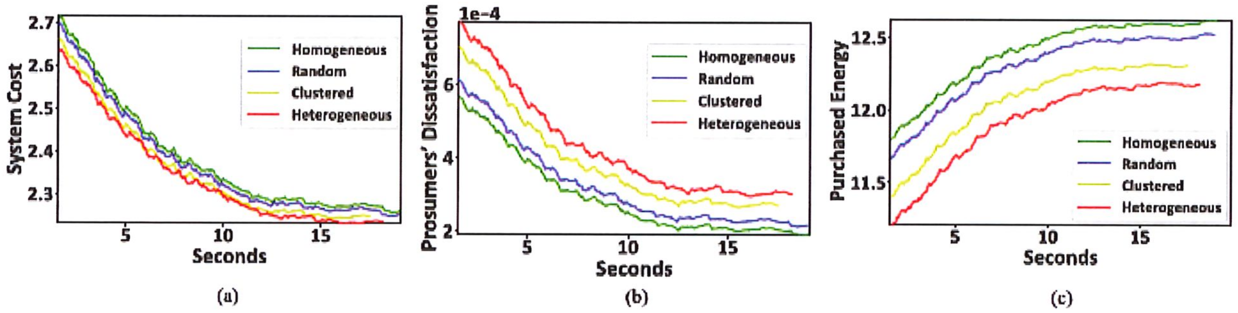


Fig. 4. (a) System cost SC^t, (b) prosumers' dissatisfaction $\sum_{\forall n \in N} \mathcal{D}_n$, and (c) prosumers' total amount of purchased energy $\sum_{\forall n \in N} e_n^{t*}$, under (i) heterogeneous, (ii) random, (iii) clustered, and (iv) homogeneous prosumers' dissatisfaction sensitivity.

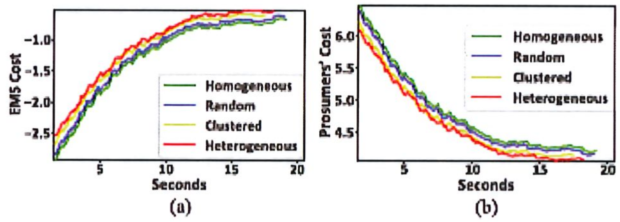


Fig. 5. (a) EMS cost EC^t, (b) prosumers' cost $\sum_{\forall n \in N} PC_n^t$, under (i) heterogeneous, (ii) random, (iii) clustered, and (iv) homogeneous prosumers' dissatisfaction sensitivity.

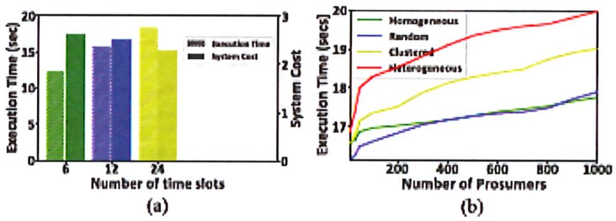


Fig. 8. Scalability evaluation with respect to (a) number of states, and (b) increasing number of prosumers, under different scenarios of dissatisfaction sensitivity populations (homogeneous, random, clustered, heterogeneous).

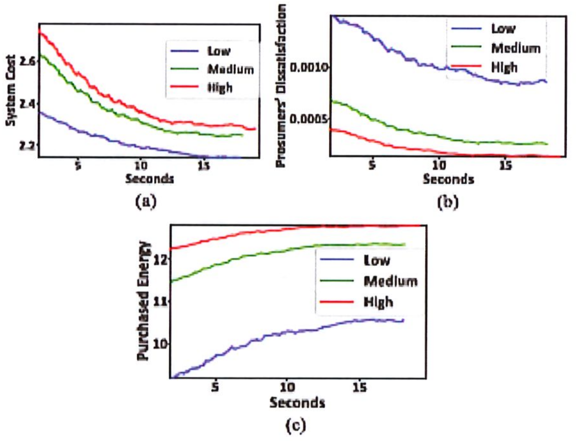


Fig. 6. (a) System cost SC^t, (b) prosumers' dissatisfaction $\sum_{\forall n \in N} \mathcal{D}_n$, and (c) prosumers' total amount of purchased energy $\sum_{\forall n \in N} e_n^{t*}$, under (i) low, (ii) medium, and (iii) high prosumers' dissatisfaction sensitivity.

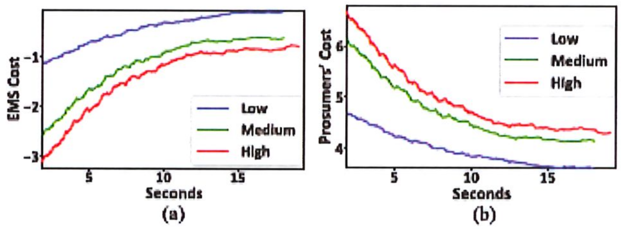
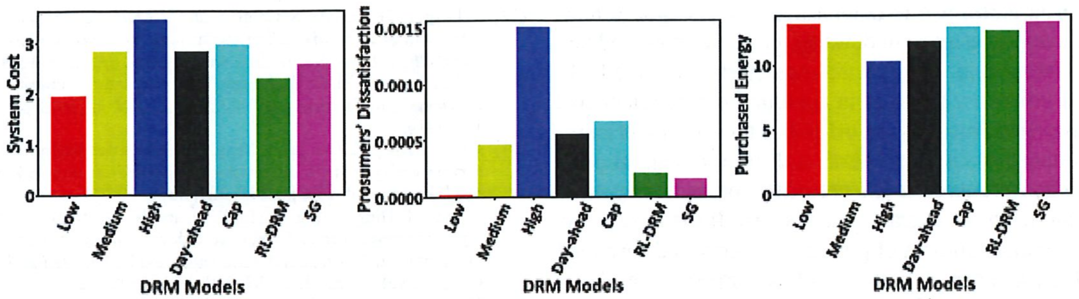


Fig. 7. (a) EMS cost EC^t, and (b) Prosumers' cost $\sum_{\forall n \in N} PC_n^t$ under (i) low, (ii) medium, and (iii) high prosumers' dissatisfaction sensitivity.

D. Comparative Evaluation

In the following, a detailed comparative evaluation of the proposed reinforcement learning price-based DRM mechanism against other price-based DRM models and alternative strategies, is presented. Specifically, throughout evaluation we consider six alternative comparative DRM models: (a)–(c) Low, Medium, High: the EMS announces a low, medium, and high fixed price during the day, respectively; (d) Day-ahead hourly pricing: the EMS announces from the previous day the hourly pricing for the day ahead; (e) Cap-based pricing: the prosumers are charged in a cap-based manner based on their consumption; and (f) Stackelberg Game (SG): the EMS announces a price in order to minimize its cost (7), and the prosumers minimize their cost function in order to determine their optimal amount of purchased energy (5) [7].

Figs. 9(a)–(c) and 10(a) and (b) present the system's cost, the prosumers' dissatisfaction, the total amount of purchased energy, the EMS cost, and the total prosumers' cost, respectively, for all the state of the abovementioned alternative scenarios, as well as for our proposed approach (referred to as RL-DRM). The results clearly show that our proposed DRM model incentivizes the prosumers to purchase energy [Fig. 9(c)], thus, keeping their experienced dissatisfaction low [Fig. 9(b)], while at the same time providing for a relatively low cost [Fig. 10(b)] compared to the other models. Also, the achieved system cost remains low as well [Fig. 9(a)]. It is highlighted that the cap-based pricing and the day-ahead hourly pricing experience higher system cost [Fig. 9(a)] compared to our proposed DRM model, as the prosumers' cost is higher [Fig. 10(b)], thus, those models mainly

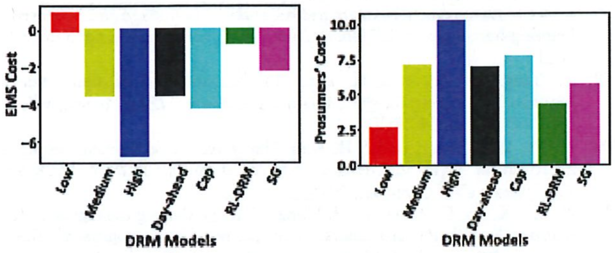


(a)

(b)

(c)

Fig. 9. Comparative evaluation: (a) System cost SC^t, (b) prosumers' dissatisfaction $\sum_{\forall n \in N} \mathcal{D}_n$, and (c) prosumers' total amount of purchased energy $\sum_{\forall n \in N} e_n^{t*}$ considering different pricing mechanisms.



(a)

(b)

Fig. 10. Comparative evaluation: (a) EMS cost EC^t, and (b) prosumers' cost $\sum_{\forall n \in N} PC_n^t$ considering different pricing mechanisms.

favor the EMS that achieves a higher profit [Fig. 10(a)]. The Stackelberg game-theoretic model also achieves higher system cost, mainly due to the higher cost experienced by the prosumers [Fig. 10(b)], who are incentivized to purchase more energy [Fig. 9(c)], resulting in lower levels of their personal dissatisfaction [Fig. 9(b)], and higher profit for the EMS [Fig. 10(a)]. The medium and high pricing scenarios are not favoring the prosumers due to the static price, thus the prosumers experience a high cost [Fig. 10(b)], resulting in a high cost for the overall system [Fig. 9(a)]. The low pricing DRM model acts unfavorably for the EMS, which experiences the highest cost compared to all the other scenarios.

E. Guidelines and Discussion

Based on the results we obtained, we conclude to the following outcomes and general guidelines regarding the reinforcement learning-based price-based DRM in smart grid systems. These observations account for the prosumer behavioral characteristics and the involved interactions among the involved actors—utility companies, the EMS, and the prosumers—while reflecting both the prosumer and the system point of view and tradeoffs.

1) A heterogeneous prosumers' population regarding their behavioral characteristics in purchasing energy, decreases the overall system cost, as the EMS collects the prosumers' budget surplus in a holistic/generic—and not personalized—manner. However, this comes with a cost for the prosumers, who either need to pay more to buy energy or they decide to postpone part of their energy needs, as the announced price is not considered beneficial and affordable to buy the whole amount of necessary energy.
2) A high system cost is experienced when the prosumers are characterized by high dissatisfaction sensitivity. The latter phenomenon is mainly observed due to the fact that the EMS has less flexibility to vary the energy price and collect higher revenue.
3) The proposed reinforcement learning price-based DRM mechanism bridges the gap between the EMS and the prosumers, who have competitive goals among each other in terms of minimizing their experienced cost. Therefore, the proposed DRM mechanism achieves a relatively low cost for the overall system while keeping the prosumers' dissatisfaction and cost at low levels as well. Also, our DRM mechanism incentivizes the prosumers to purchase energy when possible, thus enabling the EMS to still make profit.
4) From an operational and implementation point of view, the proposed reinforcement learning price-based DRM mechanism enables the EMS to adapt in a real-time manner to the prosumers' energy demand and announce a retail price that minimizes the overall system's long-term cost. The latter outcome cannot be achieved with the existing cap-based and day-ahead pricing models that are characterized by a more static retail price announcement. Also, the proposed approach alleviates the common drawback of the Stackelberg game-theoretic pricing models, which is that the game and the cost functions both for the EMS and the prosumers must be carefully designed in order to conclude to a Stackelberg equilibrium point. Moreover, the latter one is not always guaranteed that it exists in large-scale systems of heterogeneous nature (i.e., the prosumers and the EMS have competing interests).

VI. Conclusion

This article introduces a novel reinforcement learning price-based DRM model and mechanism. This research work is motivated by the observation that the majority of the existing literature mainly adopts the Stackelberg game-theoretic approaches in order to deal with the DRM problem, while limited research work has been performed to demonstrate the benefits of reinforcement learning in terms of devising an autonomous and distributed DRM method. Initially, the characteristics of the EMS and the prosumers are reflected in appropriately designed cost and dissatisfaction functions. An off-policy reinforcement learning is introduced enabling the EMS to learn the optimal price that should be announced to the prosumers on an hourly-basis toward minimizing the overall system's cost. Respectively, the prosumers' optimal amount of purchased energy is determined in a real-time manner. A detailed set of numerical results is presented demonstrating not only the pure operation and performance of the proposed mechanism but also its effectiveness in accommodating prosumers' populations of different behavioral characteristics in terms of purchasing energy patterns. Moreover, a detailed comparative evaluation against other price-based DRM models is presented showing the key benefits and tradeoffs of our proposed model.

Part of our current and future work includes the extension of the proposed DRM model in order to study the prosumers' risk-aware behavior in terms of purchasing energy under different pricing models, as the ones presented in the comparative evaluation. Toward this direction, we will explore the principles of prospect theory and contract theory, respectively, to analyze how the prosumers act and decide in terms of energy purchase, under different customers' behavioral patterns and personas, while also provide an economic/labor-driven methodology to capture the involved actors interactions.

REFERENCES

[1] M. Ghorbanian, S. H. Dolatabadi, and P. Siano, "Game theory-based energy-management method considering autonomous demand response and distributed generation interactions in smart distribution systems," *IEEE Syst. J.*, vol. 15, no. 1, pp. 905–914, Mar. 2021.

[2] M. Park, J. Lee, and D.-J. Won, "Demand response strategy of energy prosumer based on robust optimization through aggregator," *IEEE Access*, vol. 8, pp. 202969–202979, 2020.

[3] S. Belhaiza, U. Baroudi, and I. Elhallaoui, "A game theoretic model for the multiperiodic smart grid demand response problem," *IEEE Syst. J.*, vol. 14, no. 1, pp. 1147–1158, Mar. 2020.

[4] S. L. Arun and M. P. Selvan, "Intelligent residential energy management system for dynamic demand response in smart buildings," *IEEE Syst. J.*, vol. 12, no. 2, pp. 1329–1340, Jun. 2018.

[5] H. A. U. Muqeet and A. Ahmad, "Optimal scheduling for campus prosumer microgrid considering price based demand response," *IEEE Access*, vol. 8, pp. 71378–71394, 2020.

[6] N. Irtija, F. Sangoleye, and E. E. Tsiropoulou, "Contract-theoretic demand response management in smart grid systems," *IEEE Access*, vol. 8, pp. 184976–184987, 2020.

[7] M. Yu and S. H. Hong, "A real-time demand-response algorithm for smart grids: A stackelberg game approach," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 879–888, Mar. 2016.

[8] G. Hafeez et al., "An innovative optimization strategy for efficient energy management with day-ahead demand response signal and energy consumption forecasting in smart grid using artificial neural network," *IEEE Access*, vol. 8, pp. 84415–84433, 2020.

[9] O. Samuel et al., "Towards real-time energy management of multi-microgrid using a deep convolution neural network and cooperative game approach," *IEEE Access*, vol. 8, pp. 161377–161395, 2020.

[10] J. Zeng, Q. Wang, J. Liu, J. Chen, and H. Chen, "A potential game approach to distributed operational optimization for microgrid energy management with renewable energy and demand response," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4479–4489, Jun. 2019.

[11] M. Ye and G. Hu, "Game design and analysis for price-based demand response: An aggregate game approach," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 720–730, Mar. 2017.

[12] P. Jacquot, O. Beaude, S. Gaubert, and N. Oudjane, "Analysis and implementation of an hourly billing mechanism for demand response management," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4265–4278, Apr. 2019.

[13] B. Barabadi and M. H. Yaghmaee, "A new pricing mechanism for optimal load scheduling in smart grid," *IEEE Syst. J.*, vol. 13, no. 2, pp. 1737–1746, Jun. 2019.

[14] P. Liu, T. Ding, Z. Zou, and Y. Yang, "Integrated demand response for a load serving entity in multi-energy market considering network constraints," *Appl. Energy*, vol. 250, pp. 512–529, 2019.

[15] O. Han, T. Ding, L. Bai, Y. He, F. Li, and M. Shahidehpour, "Evolutionary game based demand response bidding strategy for end-users using Q-learning and compound differential evolution," *IEEE Trans. Cloud Comput.*, vol. 10, no. 1, pp. 97–110, Jan. Mar. 2022.

[16] Y. Yu, S. Chen, and Z. Luo, "Residential microgrids energy trading with plug-in electric vehicle battery via stochastic games," *IEEE Access*, vol. 7, pp. 174507–174516, 2019.

[17] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Başar, "Demand response management in the smart grid in a large population regime," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 189–199, Jan. 2016.

[18] P. A. Apostolopoulos, E. E. Tsiropoulou, and S. Papavassiliou, "Demand response management in smart grid networks: A two-stage game-theoretic learning-based approach," *Mobile Netw. Appl.*, vol. 26, no. 2, pp. 548–561, 2021.

[19] Z. Yang, M. Ni, and H. Liu, "Pricing strategy of multi-energy provider considering integrated demand response," *IEEE Access*, vol. 8, pp. 149041–149051, 2020.

[20] M. Yu, S. H. Hong, and J. B. Kim, "Incentive-based demand response approach for aggregated demand side participation," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2016, pp. 51–56.

[21] N. Liu, X. Yu, C. Wang, and J. Wang, "Energy sharing management for microgrids with PV prosumers: A Stackelberg game approach," *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1088–1098, Mar. 2017.

[22] X. Dong, X. Li, and S. Cheng, "Energy management optimization of microgrid cluster based on multi-agent-system and hierarchical Stackelberg game theory," *IEEE Access*, vol. 8, pp. 206183–206197, 2020.

[23] M. Yu, S. H. Hong, Y. Ding, and X. Ye, "An incentive-based demand response (DR) model considering composited DR resources," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1488–1498, Feb. 2019.

[24] B. Wang, Y. Li, W. Ming, and S. Wang, "Deep reinforcement learning method for demand response management of interruptible load," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3146–3155, Jul. 2020.

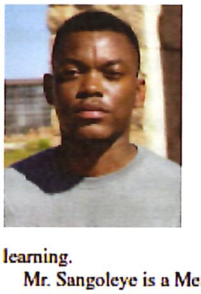
[25] D. Zhang, S. Li, M. Sun, and Z. O'Neill, "An optimal and learning-based demand response and home energy management system," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1790–1801, Jul. 2016.

[26] Y. Liang, Z. Ding, T. Ding, and W.-J. Lee, "Mobility-aware charging scheduling for shared on-demand electric vehicle fleet using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1380–1393, Mar. 2021.

[27] S. Cui, Y.-W. Wang, J.-W. Xiao, and N. Liu, "A two-stage robust energy sharing management for prosumer microgrid," *IEEE Trans. Ind. Inform.*, vol. 15, no. 5, pp. 2741–2752, May 2019.

[28] T. Ding, M. Qu, N. Amjady, F. Wang, R. Bo, and M. Shahidehpour, "Tracking equilibrium point under real-time price-based residential demand response," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2736–2740, May 2021.

[29] USA EIA, Energy Information Administration. 2021. [Online]. Available: https://www.eia.gov/



Fisayo Sangoleye (Student Member, IEEE) received thet bachelor's degree in electrical and electronics engineering from the University of Lagos, Nigeria, in 2016. He is currently working toward the Ph.D. degree in computer engineering with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA.

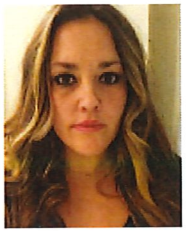
His research interests include optimization in wireless networks, artificial intelligence, and demand response management in smart grid systems based on game theory, contract theory, and reinforcement learning.

Mr. Sangoleye is a Member of the National Society of Black Engineers.

Jenilee Jao (Student Member, IEEE) is currently working toward the senior year with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA.

Her main research interests include machine learning, artificial intelligence, and visualization in 3-D space.

Ms. Jao is a former officer for the Institute of Electrical and Electronics Engineers student branch, Tau Beta Pi, and an inductee for Eta Kappa Nu.



Kimberly Faris is currently working toward the senior year with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA.

She is also a Research Assistant with the Aerospace Research Division, COSMIAC, Albuquerque, NM, USA, and the Department of Electrical and Computer Engineering, University of New Mexico. Her research interests include applied pulsed energy, pulse generators, power flow in conical/coaxially magnetically induced transmission lines, and plasma.

Eirini Eleni Tsiropoulou (Senior Member, IEEE) is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA.

Her research interests include cyber-physical social systems and wireless heterogeneous networks, with emphasis on network modeling and optimization, resource orchestration in interdependent systems, reinforcement learning, game theory, and network economics.

Dr. Tsiropoulou was selected by the IEEE Communication Society - N2Women - as one of the top ten Rising Stars of 2017 in the communications and networking field, and the recipient of the Early Career Award by the IEEE Communications Society Internet Technical Committee, in 2019.

Symeon Papavassiliou (Senior Member, IEEE) is currently a Professor with the School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece. From 1995 to 1999, he was a Senior Technical Staff Member with AT&T Laboratories, Middletown Township, NJ, USA. In August 1999, he joined the ECE Department, New Jersey Institute of Technology, Newark, NJ, USA, where he was an Associate Professor, until 2004. He has an established record of publications in his field of expertise, with more than 350 technical journal and conference published papers.

His research interests include modeling, optimization and performance evaluation of distributed complex systems and social networks.