

Demand Responsive Dynamic Pricing Framework for Prosumer Dominated Microgrids using Multiagent Reinforcement Learning

Amin Shojaeighadikolaei, Arman Ghasemi, Kailani R. Jones,
Alexandru G. Bardas, Morteza Hashemi, Reza Ahmadi

Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA

E-mails: {amin.shojaei, arman.ghasemi, kailanij, alexbardas, mhashemi, ahmadi}@ku.edu

Abstract—Demand Response (DR) has a widely recognized potential for improving grid stability and reliability while reducing customers' energy bills. However, the conventional DR techniques come with several shortcomings, such as inability to handle operational uncertainties and incurring customer disutility, impeding their wide spread adoption in real-world applications. This paper proposes a new multiagent Reinforcement Learning (RL) based decision-making environment for implementing a Real-Time Pricing (RTP) DR technique in a prosumer dominated microgrid. The proposed technique addresses several shortcomings common to traditional DR methods and provides significant economic benefits to the grid operator and prosumers. To show its better efficacy, the proposed DR method is compared to a baseline traditional operation scenario in a small-scale microgrid system. Finally, investigations on the use of prosumers' energy storage capacity in this microgrid highlight the advantages of the proposed method in establishing a balanced market setup.

Index Terms—Microgrid, Demand Response, Prosumer, Real-Time Pricing, Reinforcement Learning

I. INTRODUCTION

Demand-side management, also known as Demand Response (DR), is one of the most widely studied topics in the context of the smart grid [1]. In order to flatten the demand curve, conventional time-based DR methods such as the Real-Time Pricing (RTP) approach [2] rely on dynamically changing the electricity price to motivate the customers to alter their energy use profile [3]. This potentially improves the grid stability and reliability by shifting the peak demand and decreasing the need for peaking power plants while offering reduced energy bills to residential customers [4].

Nevertheless, the conventional time-based DR methods typically assume that the pricing policies are deterministic and decided ahead of time [5]–[7], or consider that pricing policies follow a random process with known properties [8], [9]. Thus, conventional DR methods are either not able to offer convergence to the optimal solution in presence of uncertainties in the environment or the mathematical formulations become cumbersome to model uncertainties [10], [11], which makes them unsuitable for real-world implementations.

Moreover, the conventional time-based DR approaches almost exclusively focus on altering the preferred electricity consumption pattern of the customers, e.g., by changing the temperature set point on air conditioning systems or delaying the use of major appliances, in order to shift their load to

off-peak periods. These approaches often lead to customer dissatisfaction (disutility), and, as a consequence, have not been widely adopted. Psychologically, the consumers value their comfort much higher than economic savings provided by traditionally proposed DR approaches [12].

Integration of energy storage into residential photovoltaic (PV) systems should provide energy consumers & producers, commonly known as *prosumers*, with more flexibility to participate in DR programs while minimizing their disutility. In other words, prosumers should be able to shape their demand profile in real-time regardless of their consumption profile [13]. As a result, prosumers can potentially receive greater economic benefits by selling their excess energy to the grid at higher prices, while aiding grid support services [14]. Achieving the aforementioned benefits demands for a novel DR approach on the prosumers' side. This DR approach considers various factors such as state-of-charge (SOC) of the storage device, household consumption profile, and real-time electricity price and PV generation levels. Furthermore, taking advantage of the flexibility provided by prosumers requires a modern grid management strategy which can treat the storage capacity of households as a grid asset that can be dispatched by properly incentivizing the households for DR participation, and leverage this asset for grid cost and performance optimization.

This paper proposes a multiagent deep Reinforcement Learning (RL) framework for implementing a new RTP-based DR technique in a prosumer microgrid that provides both, prosumers and the grid operator, with the aforementioned flexibility and greater economic benefits. The main contributions of the proposed framework are summarized as follows: a) *Real-time learning*: Grid and prosumer agents learn the optimal price and DR participation policy by interacting with the environment in real-time, rather than using a complex grid dynamic model for optimization. Therefore, the proposed method is applicable to the high dimensional and non-stationary environment of the grid with much less computational burden than traditional DR methods, allowing for real-world implementations; b) *Altering the households' grid injection patterns*: The goal of the proposed prosumer-side DR algorithm is *not* to alter the consumption pattern of households which can typically lead to customer dissatisfaction. Rather,

the prosumer-side DR algorithm provides cost savings by altering the households' grid injection pattern, using the flexibility provided by the energy storage and PV generation; and c) *Balanced market*: The proposed framework makes better use of prosumers' energy storage capacity, allowing for significant prosumer electricity bill reduction, as well as considerable grid economic benefit improvement, with a reasonably sized battery pack. Our results also show that after a certain threshold value, a larger battery size on the prosumer side does not necessarily yield a much higher profit. This facilitates a balanced market setup where trying to abruptly and unilaterally manipulate the pricing scheme is not in anyone's financial interest.

Related Work: In recent years there has been a growing interest in application of RL to the problem of dynamic pricing and DR. A comprehensive survey of published works in this area is provided in [15]. Among the surveyed publications, the works in [16] and [17] are closely aligned with our work. The authors in [16] propose a RL-based dynamic pricing and energy consumption scheduling framework that can work without priori information and leads to reduced system costs. The proposed framework targets regular customers without grid injection capability and assumes that customer behavior is myopic and deterministic, i.e. each costumer is trying to minimize its cost in every single time slot. In contrast, our work takes advantage of the PV generation and storage capacity of prosumers, leading to greater flexibility for DR participation, and enables the prosumers to make decisions that lead to long-term optimization of their accumulative economic benefit, rather than minimizing their instantaneous cost. On the other hand, the authors in [17] present a RL-based dynamic pricing algorithm which can promote service provider profitability and reduce energy costs for the customers. However, similar to [16], this work only deals with regular electricity consumers, rather than prosumers with generation capability.

The reminder of this paper is structured as follows: Section II details the proposed RL based DR approach. Section III implements the proposed method for a small-scale microgrid as a case study and provides simulation results to verify efficacy of the proposed method. Finally, our concluding thoughts are presented in Section IV.

II. PROPOSED RL BASED DR METHOD

Fig. 1 illustrates the envisioned microgrid system with conventional generation facilities, prosumers, and consumers. Prosumers are entities that can produce energy locally from renewable resources, store the excess energy in their battery, sell energy into the grid, or buy electricity from the grid for local consumption. The prosumers can make a profit by selling electricity to the grid at a dynamic \$/kWh price of (δ^b), referred to as buy price hereinafter. On the other hand, the prosumers incur a cost when buying electricity from the grid at a \$/kWh price referred to as sell price (δ^s), hereinafter. The goal of prosumer agents is to maximize its long term profit by determining an optimal charge/discharge policy for their energy storage devices. Similarly, the grid can buy energy from prosumers at a price of (δ^b) and incur a cost, or sell

electricity to prosumers at a price of (δ^s) to make a profit. The goal of the grid agent is defined as maximizing the long term profit of the grid by determining an optimal buy price (δ^b) policy. Therefore, we define the following optimization problem for grid agents,

$$\begin{cases} \text{maximize} & M(T) \\ \text{subject to:} & \sum_{q=1}^{N_c+N_p} P_q^d(t) = \sum_{i=1}^{N_g} P_i^G(t) + \sum_{j=1}^{N_p} P_j^{inj}(t) \\ & P_i^{G,\min}(t) \leq P_i^G(t) \leq P_i^{G,\max}(t), \end{cases} \quad (1)$$

where $M(T)$ is the grid profit defined as follows:

$$M(T) = V(P^{dm}(T)) - \sum_{i=1}^{N_g} F_i(P_i^G(T)) - \sum_{j=1}^{N_p} C_j(P_j^{inj}(T)).$$

In addition, we have:

$$V(P^{dm}(T)) = \int_0^T P^{dm}(t) \delta^s(t) dt, \quad (2)$$

$$C_j(P_j^{inj}(T)) = \begin{cases} \int_0^T P_j^{inj}(t) \delta^b(t) dt & \text{for } P_j^{inj}(t) > 0 \\ 0 & \text{for } P_j^{inj}(t) \leq 0. \end{cases} \quad (3)$$

In which $V(P^{dm}(T))$ represents the accumulative grid revenue by selling electricity to the households in time horizon T , $C_j(P_j^{inj}(T))$ represents the accumulative grid cost of buying excess energy from j^{th} prosumer in time horizon T , $F_i(P_i^G(T))$ represents the accumulative cost of buying electricity from i^{th} generation facility time horizon T , N_c , N_g and N_p are the number of consumers, generation facilities and prosumers respectively. The first equality constraint in (1) represents the grid's power balance requirement which needs to be maintained at all times, P^{dm} is the total demand of the network, $P_q^d(t)$ is the demand of household q , and P_j^{inj} is the power injection into the grid by j^{th} prosumer.

Similarly we define the following optimization problem for j^{th} prosumer,

$$\begin{cases} \text{maximize} & U_j(T) \\ \text{subject to:} & P_j^{inj}(t) + P_j^{batt}(t) + P_j^c(t) = P_j^{pv}(t) \\ & |P_j^{inj}(t)| \leq P_j^{inj,\max} \\ & |P_j^{batt}(t)| \leq P_j^{batt,\max} \\ & 0 \leq P_j^{pv}(t) \leq P_j^{pv,\max} \\ & SoC_j^{\min} \leq SoC_j(t) \leq SoC_j^{\max}, \end{cases} \quad (4)$$

where $U_j(T) = V_j^p(P_j^{inj}(T)) - C_j^p(P_j^{inj}(T))$ is the j^{th} prosumer accumulative profit in time horizon T , $V_j^p(P_j^{inj}(T))$ represents the j^{th} prosumer accumulative revenue for selling excess electricity to the grid in time horizon T calculated using the same equation as (3) considering $V_j^p(P_j^{inj}(T)) = C_j(P_j^{inj}(T))$, and $C_j^p(P_j^{inj}(T))$ represents the j^{th} prosumer

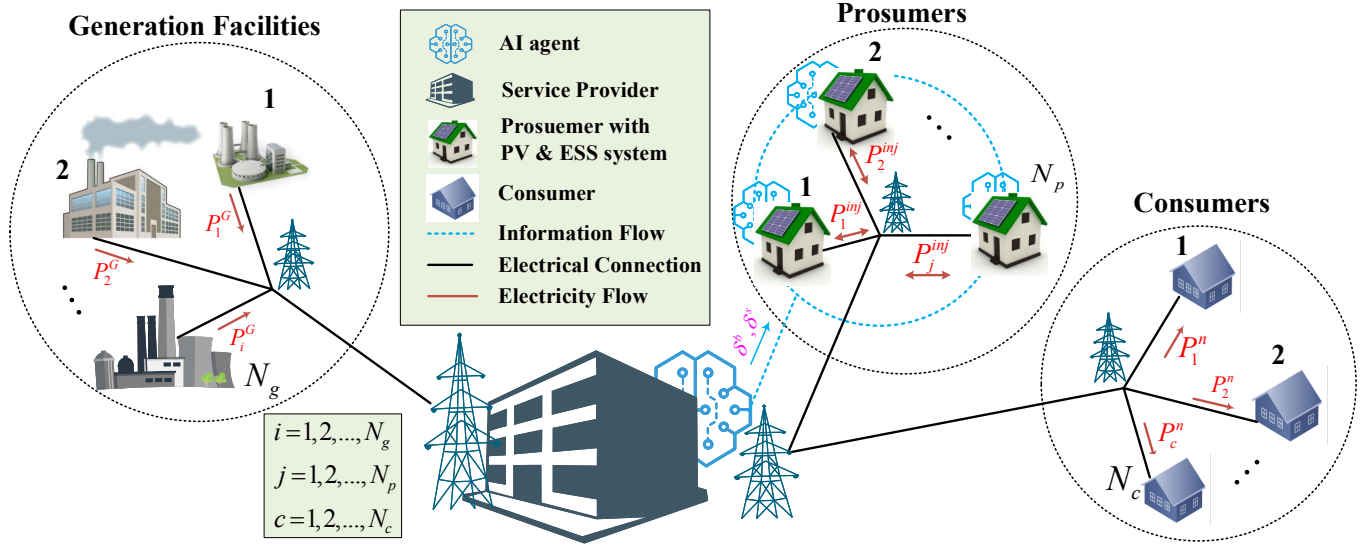


Fig. 1. Small scale microgrid system that consists of generation facilities, traditional consumers, and prosumers. Grid and prosumers are equipped with reinforcement learning (RL) agents to dynamically adjust their policies in terms of buy/sell prices, and power injection.

accumulative cost of buying electricity from the grid in time horizon T calculated as,

$$C_j^P \left(P_j^{inj}(T) \right) = \begin{cases} \int_0^T P_j^{inj}(t) \delta^s(t) dt & \text{for } P_j^{inj}(t) < 0, \\ 0 & \text{for } P_j^{inj}(t) \geq 0. \end{cases} \quad (5)$$

$P_j^{pv}(t)$ is the PV generation with the peak generation of $P_j^{pv,max}$ for j^{th} household, $P_j^c(t)$ is consumption for j^{th} household, $P_j^{inj,max}$ represents the maximum allowable power injection for j^{th} household, $P_j^{batt}(t)$ represents the energy storage charge/discharge power with maximum allowable charge/discharge power of $P_j^{batt,max}$ for j^{th} household, $SoC_j(t)$ is the state of charge for j^{th} prosumer where SoC_j^{min} and SoC_j^{max} are minimum and maximum allowable state of charge. The state of charge is calculated as,

$$SoC_j(t) = SoC_j(0) + \frac{1}{\eta_j} \int_0^t P_j^{batt}(\tau) d\tau, \quad (6)$$

where $SoC_j(0)$ is the initial state of charge and η_j is the energy storage capacity of j^{th} prosumer.

Each agent uses RL for solving the defined optimization problems in real-time. The agents interact with the environment in order to maximize a specified reward. The agents receive positive rewards for taking desirable actions and negative rewards for taking undesirable actions. This learning process is modeled as a Markov Decision Process (MDP) for a Multi-agent Reinforcement Learning (MARL) environment defined by a tuple of $\{N, S, \{A_i\}, P, \{R_i\}, \gamma\}$ where N is the set of agents, S is the set of states of the environment, A_i is the set of actions for i^{th} agent, P is the set of transition function, R_i is the immediate reward function set for i^{th} agent, and $\gamma \in [0, 1]$ is the discount factor. The agent-environment interaction for an MDP is shown in Fig 2. At each discretized time index k , the agent N_i selects an action $a_{i,k} \in A_i$ based on

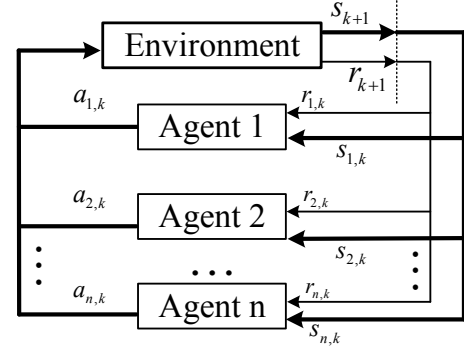


Fig. 2. A typical reinforcement learning framework wherein agents learn their optimal actions through observing the environment states, taking actions, and receiving rewards.

observing the current state of the environment denoted by s_k . Subsequently, the agent receives a numerical feedback signal known as reward, $r_{i,k} \in R_i$, and transitions to a new state s_{k+1} . In the proposed framework, each agent takes actions only based on its own local. In other word, each agent can only observe partial features of the entire environment. Thus, the grid agent can observe the following environment states,

$$s_{1,k} = \{ \mathbf{F}_k, \mathbf{\Omega}_k, P_{1,k}^{dm} \} \in S, \quad (7)$$

where $\mathbf{F}_k = [F_{1,k} \dots F_{i,k}]^T$ and $\mathbf{\Omega}_k = [C_{1,k} \dots C_{j,k}]^T$ are the vectors of the grid cost for buying electricity from the generation facilities and prosumers at time slot k respectively, $P_{1,k}^{dm}$ represents the total grid demand at time slot k .

Similarly, each prosumer agent can observe the following environment states,

$$s_{n,k} = \{ SoC_{n,k}, P_{n,k}^{pv}, P_{n,k}^c, \delta_k^b \} \in S \quad \text{for } n = 2, 3, \dots, N_p + 1, \quad (8)$$

where $SoC_{n,k}$ is the state of charge of energy storage device of n^{th} prosumer at time slot k , $P_{n,k}^{pv}$ and $P_{n,k}^c$ are the PV generation and local consumption of n^{th} prosumer at time slot k , δ_k^b is the buy price at time slot k .

In this work, the grid agent controls the buy price to incentivize the DR participation of the prosumers and maximize its own reward. Therefore, the buy price is the action of the grid agent denoted by $a_{1,k} = \delta_k^b \in A_1$. On the other hand, the prosumer agents control the household's energy storage charge/discharge state in response to the dynamic buy price changes. Therefore, the action for n^{th} prosumer's agent is defined as $a_{n,k} \in A_n$ which denotes the charge/discharge command to the energy storage.

Finally, the immediate reward function for the grid and prosumer agents are defined as,

$$r_{1,k} = P_{1,k}^{dm} \times \delta_k^s - \sum_{i=1}^{N_g} F_{i,k} (P_{i,k}^G) - \sum_{j=1}^{N_p} P_{j,k}^{inj} \times \delta_k^b \text{ for } P_{j,k}^{inj} > 0, \quad (9)$$

$$r_{n,k} = (\rho) P_{n,k}^{inj} \times \delta_k^b + (\rho - 1) P_{n,k}^{inj} \times \delta_k^s, \quad (10)$$

where $\rho \in \{0,1\}$ in which $\rho = 0$ when $P_{n,k}^{inj} \leq 0$ and $\rho = 1$ when $P_{n,k}^{inj} > 0$.

According to the above terminology, the MDP trajectories for the grid agent and n^{th} prosumer's agent begin with $s_{1,1}$, $a_{1,1}$, $r_{1,2}$, $s_{1,2}$, $a_{1,2}$, $r_{1,3}$, ... and $s_{n,1}$, $a_{n,1}$, $r_{n,2}$, $s_{n,2}$, $a_{n,2}$, $r_{n,3}$, ..., respectively. The primary goal of each agent is to maximize the accumulative reward sequence formulated by $G_1 = \sum_{t=0}^{\infty} (\gamma_1)^t r_{1,k+t+1}$ and

$G_n = \sum_{t=0}^{\infty} (\gamma_n)^t \cdot r_{n,k+t+1}$ where $0 \leq \gamma_1 \leq 1$ and $0 \leq \gamma_n \leq 1$ are discount factors and $n = 2, 3, \dots, N_p + 1$. In this work, to find the optimal policy, Deep Q-Network (DQN) [18] is deployed to approximate the optimal action-value function which satisfies the Bellman Equation as,

$$\hat{Q}_{1,k+1}(s_{1,k}, a_{1,k}) = (1 - \alpha_1) \hat{Q}_{1,k}(s_{1,k}, a_{1,k}) + \alpha_1 \left\{ r_{1,k}(s_{1,k}, a_{1,k}) + \gamma_1 \max_{a_{1,k+1}} \hat{Q}_{1,k}(s_{1,k+1}, a_{1,k+1}) \right\}, \quad (11)$$

$$\hat{Q}_{n,k+1}(s_{n,k}, a_{n,k}) = (1 - \alpha_n) \hat{Q}_{n,k}(s_{n,k}, a_{n,k}) + \alpha_n \left\{ r_{n,k}(s_{n,k}, a_{n,k}) + \gamma_n \max_{a_{n,k+1}} \hat{Q}_{n,k}(s_{n,k+1}, a_{n,k+1}) \right\}, \quad (12)$$

where \hat{Q} is the approximation of Q that is estimated by a deep neural network. In this work, we use the updating mechanism provided in [19] for Q-values.

III. CASE STUDY AND NUMERICAL RESULTS

The proposed DR scheme is implemented on a small-scale microgrid similar to Fig. 1. The case study system is comprised

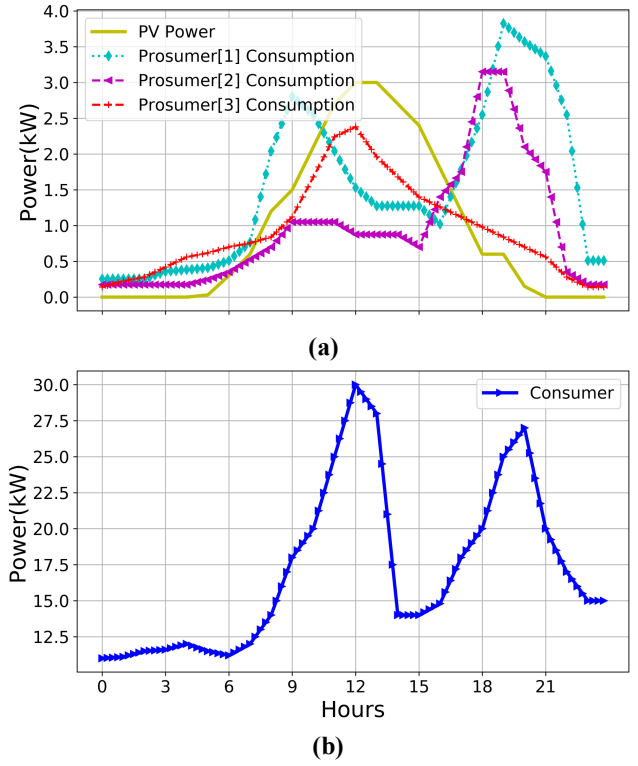


Fig. 3. generation and consumption waveform sample (a) Generation and prosumers' consumption waveform (b) Consumer consumption waveform

of ($N_p = 3$) prosumers equipped with solar rooftop panels, energy storage system and smart agent, a conventional consumer representative of non-generational consumer ($N_c = 1$), and two generation facilities ($N_g = 2$) where one acts as a baseline generation facility and another acts as a reserve generation capacity. The prosumers' details are provided in Table. I. Employed generation and consumption profiles for the prosumers are provided in Fig. 3, where the generation and consumption profiles are derived from California ISO [20].

Two scenarios have been simulated for analyzing the efficacy of the proposed DR scheme. In the first scenario, referred to as conventional scenario hereinafter, no DR scheme is applied to the microgrid. The prosumers simply inject power to grid when there is an excess energy generation in the household and the battery capacity is full. For the second scenario, the proposed DR scheme is implemented on the microgrid system and results are compared with the first scenario to evaluate the effectiveness of the proposed method.

Fig. 4 compares the time domain profiles of the SoC of the prosumers and the buy/sell prices over a 24 hour period between the conventional and agent-based scenarios after fully training the DQN agents. Comparing Fig. 4 (b)-(d) with

Item	Max PV Generation	ESS Capacity	Max Charge/Discharge	Agent
Prosumer 1	4 kW	6 kW	2/-2 kW	Agent1
Prosumer 2	4 kW	12 kW	2/-2 kW	Agent2
Prosumer 3	4 kW	8 kW	2/-2 kW	Agent3

TABLE I. Microgrid Details

Fig. 3 (a), it can be observed that in the conventional scenario the changes in the SoC of the battery of each prosumer is very closely synced with the prosumer's PV generation profile. This is expected since as mentioned above in the conventional scenario prosumers inject power to grid when there is an excess energy generation. On the other hand, in the agent based scenario, the battery SoC is changed as a result of charge/discharge commands issued by the prosumer agents based on the identified optimal charge/discharge policy. To compare the effect of this significant change in battery usage between the two scenarios, prosumers' average daily electricity bill, grid daily profit and reserve power consumption for the two scenarios have been calculated and plotted in Fig.5. According to this figure, the prosumers's average daily electricity bill is reduced significantly in the agent-based scenario. Similarly, the grid profit is considerably higher in the agent-based scenario which can be attributed to the significant drop in the reserve generation usage in this scenario.

According to the findings discussed above, it can be concluded that the grid and prosumer agents are leveraging the battery capacity of the prosumers to maximize their profits and reduce their costs.

In the next experiment, the effect of battery capacity on reduction of daily electricity bill of prosumers and raising the grid profit is investigated by running several agent-based simulations and increasing the battery capacity of the prosumers for each simulation. The results are shown in Fig. 6 and Fig. 7, respectively. As pictured, the battery capacity is increased from 2 kWh to 25 kWh and the daily energy bill of prosumers and the grid profit are measured at the end of simulation and plotted against the battery capacity. Fig. 6 and Fig. 7 show a downward trend in the daily electricity cost of prosumers and an upward trend in the grid profit as a function of battery capacity. However, these trends seem to slow down around 15 kWh battery capacity, meaning that the improvements when using batteries with larger than 15 kWh capacity seem negligible. Therefore, for a given PV generation capacity (i.e., $P_{pv,max}$), it can be concluded that the proposed DR scheme can provide maximum benefits with a reasonable battery pack size of around 15 kWh in the households.

IV. CONCLUSIONS

This paper proposes a new multiagent RL-based decision-making environment for implementing a DR scheme in a microgrid dominated by prosumers. The proposed technique implements a Real-Time Pricing scheme that can mitigate several shortcomings common to traditional DR methods while providing important economic benefits to the grid operator and prosumers. To showcase the better efficacy of the RL-based method, this work includes a comparison to a baseline traditional operation scenario in a small-scale microgrid system. Results showed significant daily bill reductions (e.g., 38% , 46%, and 26%) for prosumers in the proposed RL-based marketplace . Moreover, experiments on the use of prosumers' energy storage capacity in this microgrid setup highlight the

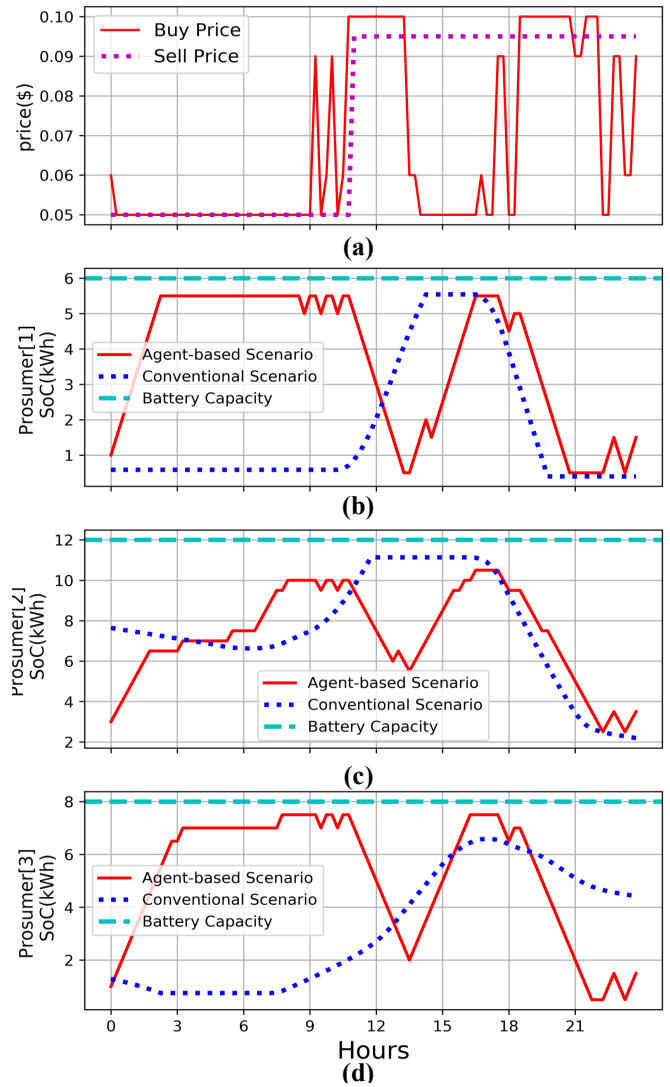


Fig. 4. Simulation results after learning algorithm convergence for a day (24 hours) (a) Grid buying and selling price (b) Prosumer 1 battery SoC (c) Prosumer 2 battery SoC (d) Prosumer 3 battery SoC

advantages of the proposed method in establishing a fair and balanced market setup.

REFERENCES

- [1] "Demand response and smart grids—a survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461 – 478, 2014.
- [2] H. T. Haider, O. H. See, and W. Elmenreich, "A review of residential demand response of smart grid," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 166–178, 2016.
- [3] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A survey on demand response programs in smart grids: Pricing methods and optimization algorithms," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 152–178, 2015.
- [4] J. Morales González, A. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets: Operational Problems*, ser. International Series in Operations Research and Management Science. Springer, 2014.
- [5] H. Roh and J. Lee, "Residential demand response scheduling with multiclass appliances in the smart grid," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 94–104, 2016.

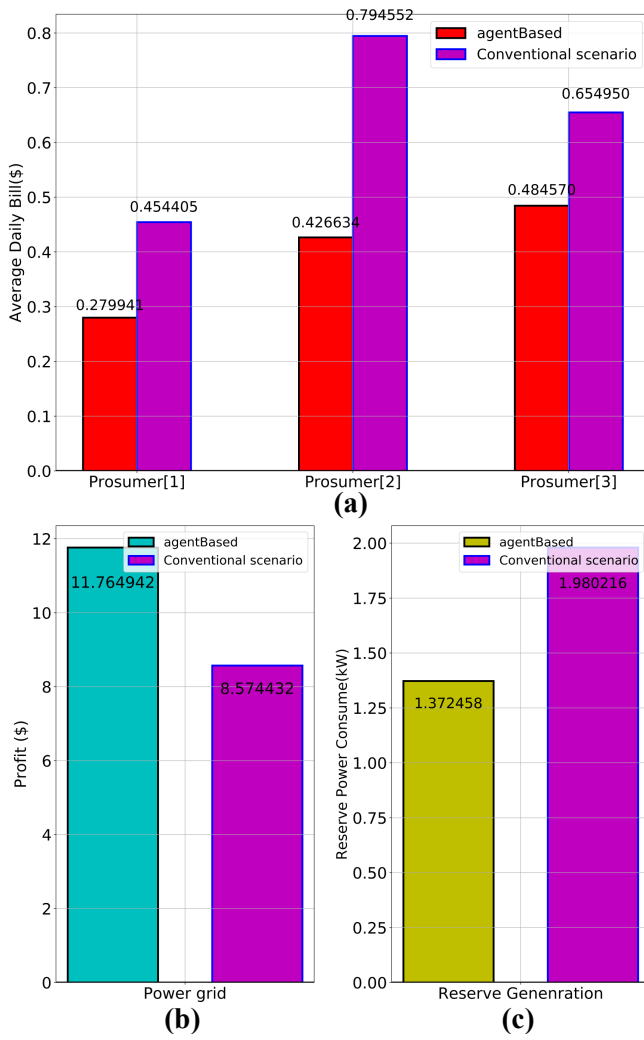


Fig. 5. Daily bill comparison over episodes (a)prosumer 1-3 daily bill (b) grid profit (c) grid reserve power consume

- [6] P. Samadi, R. Schober, and V. W. S. Wong, "Optimal energy consumption scheduling using mechanism design for the future smart grid," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 369–374.
- [7] B. Kim, S. Ren, M. van der Schaar, and J. Lee, "Bidirectional energy trading and residential load scheduling with electric vehicles in the smart grid," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1219–1234, 2013.
- [8] M. H. Shoreh, P. Siano, M. Shafie-khah, V. Loia, and J. P. Catalão, "A survey of industrial applications of demand response," *Electric Power Systems Research*, vol. 141, pp. 31–49, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378779616302632>
- [9] J. Wang, H. Zhong, Z. Ma, Q. Xia, and C. Kang, "Review and prospect of integrated demand response in the multi-energy system," *Applied Energy*, vol. 202, pp. 772–782, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306261917306992>
- [10] M. Yu, R. Lu, and S. H. Hong, *A real-time decision model for industrial load management in a smart grid*, vol. 183, pp. 1488–1497, 2016.
- [11] X. Huang, S. H. Hong, and Y. Li, "Hour-ahead price based energy management scheme for industrial facilities," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2886–2898, 2017.
- [12] S. Althaher, P. Mancarella, and J. Mutale, "Automated demand response from home energy management system under dynamic pricing and power and comfort constraints," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1874–1883, 2015.
- [13] "Optimal operation of des/chp based regional multi-energy prosumer with demand response," *Applied Energy*, vol. 167, pp. 353–365, 2016.

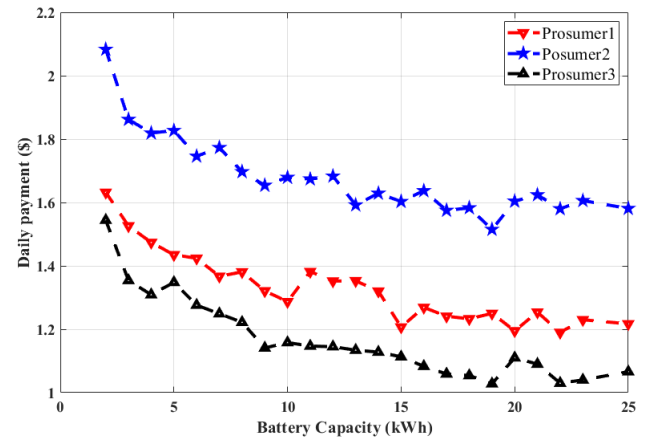


Fig. 6. Daily bill reduction for prosumers with different battery capacity

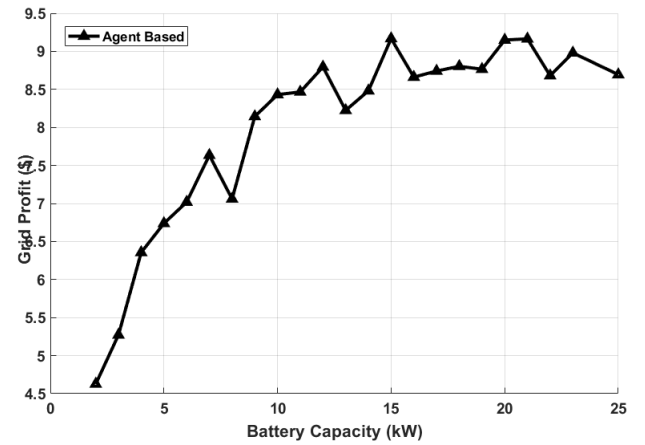


Fig. 7. Monitoring Grid profit change with different battery capacity

- [14] N. Liu, M. Cheng, X. Yu, J. Zhong, and J. Lei, "Energy-sharing provider for pv prosumer clusters: A hybrid approach using stochastic programming and stackelberg game," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6740–6750, 2018.
- [15] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Applied Energy*, vol. 235, pp. 1072–1089, 2019.
- [16] B. Kim, Y. Zhang, M. van der Schaar, and J. Lee, "Dynamic pricing and energy consumption scheduling with reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2187–2198, 2016.
- [17] R. Lu, S. H. Hong, and X. Zhang, "A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Applied Energy*, vol. 220, pp. 220–230, 2018.
- [18] N. Naderializadeh and M. Hashemi, "Energy-aware multi-server mobile edge computing: A deep reinforcement learning approach," in *53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 383–387.
- [19] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," 2017.
- [20] California ISO. Current and forecasted demand. [Online]. Available: <http://www.caiso.com/TodaysOutlook/Pages/default.aspx>