An Efficient Distributed Reinforcement Learning for Enhanced Multi-Microgrid Management

Avijit Das*, Zhen Ni†, and Di Wu*
*Pacific Northwest National Laboratory, Richland, WA 99352 USA
Email: avijit.das@pnnl.gov and di.wu@pnnl.gov
†Florida Atlantic University, Boca Raton, FL 33431 USA
Email: zhenni@fau.edu

2022 International Joint Conference on Neural Networks (IJCNN) | 978-1-7281-8671-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IJCNN55064.2022.9892754

***Abstract*—Economic dispatch in a multi-microgrid (MMG) system involves an increasing number of states from distributed energy resources (DERs) compared to a single microgrid. In these cases, traditional reinforcement learning (RL) approaches may become computationally expensive or less effective in finding the least-cost solution. This paper presents a novel RL approach that employs local learning agents to interact with individual microgrid environments in a distributed manner and a global agent to search for actions to minimize system cost at the MMG system level. The proposed distributed RL framework is more efficient in learning the dispatch policy compared to conventional approaches. Case studies are performed on a 3-microgrid system with different types of DERs. Results substantiate the effectiveness of the proposed approach in comparison with conventional methods in terms of operation costs, computation time, and peak-to-average ratio.**

***Index Terms*—Aggregating knowledge, distributed energy resources, distributed learning, multi-microgrid dispatch, and reinforcement learning.**

I. Introduction

Recent advances and achievements of reinforcement learning (RL) have opened the door for its applications to a broad range of power system problems [1], including multi-microgrid (MMG) energy management. The development and deployment of distributed energy resources (DERs) along with advanced metering, communication, and control technologies at the distribution level have transformed many conventional distribution systems into modern MMG systems [2]. Energy management in MMG systems is crucial to harvest the potential benefits of DERs and has attracted a lot of attention in recent years.

There are two kinds of approaches for energy management in an MMG system: competitive and collaborative [2], [3]. In a competitive approach, a dedicated control center is designed and used to minimize the cost of each microgrid. On the other hand, a collaborative approach aims to minimize the total cost of an MMG system with or without a global energy management center. Examples of each kind of approaches are provided as follows:

- Competitive approaches
  An RL-based bi-level coordinated energy management framework is proposed in [4] for an MMG system where a game-theoretic interactive mechanism between the operator and the microgrids is applied to minimize their individual costs. In [5], an interactive dispatching model of virtual microgrids and a distribution network is proposed with a bi-level bidding and market clearing strategy using the multi-agent RL. An auction-based microgrid market is proposed in [6] using multi-agent RL to determine the equilibrium of all agents' benefits to maximize the average rewards. A multi-agent RL framework is proposed in [7] to solve a distributed energy management problem by maintaining a benefits balance during agents' interactive learning. RL approaches are also proposed in [8] and [9] where DER agents interact with each other in a distributed manner to find an optimal operation strategy in competitive environments. Most of these approaches look for equilibrium points like the Nash equilibrium point strategy that may not always exist or guarantee the optimal dispatch [3]. In addition, since all agents aim to maximize their own benefits, it may create an extra burden to balance energy in the distribution network [2].
- Cooperative approaches
  Cooperative control frameworks are proposed in [10] and [11] to solve energy imbalance and energy storage dispatch problems in MMG systems. An RL-based cooperative multi-agent system is designed in [12] for MMG dispatch using fuzzy Q-learning methods. An RL approach with a distributed cooperative mechanism is proposed in [13] to avoid a centralized controller that involves a large number of states. Nevertheless, it is quite challenging to design distributed control without any centralized coordination to achieve the desired performance. A centralized controller is therefore recommended for creating a more efficient generation-demand balance both inside and between microgrids [2], [3].

Distributed RL with a global agent [14]–[17] has the po-

This work was supported in part by the U.S. Department of Energy (DOE), Office of Electricity through the Energy Storage program and in part by the National Science Foundation under Grants 1949921 and 2047064. Pacific Northwest National Laboratory is operated for the U.S. DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830.

978-1-7281-8671-9/22/$31.00 ©2022 IEEE
Authorized licensed use limited to: Florida Atlantic University. Downloaded on January 13,2023 at 19:42:49 UTC from IEEE Xplore. Restrictions apply.

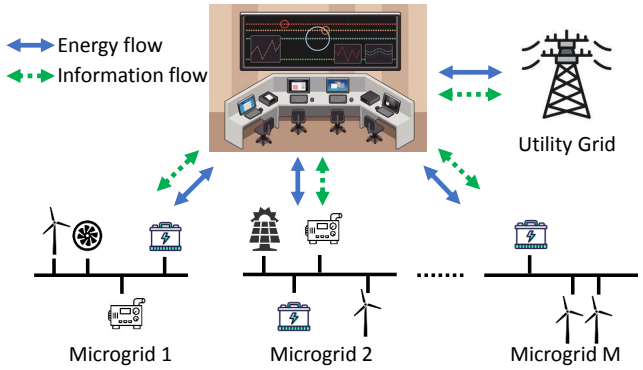

Fig. 1. Multi-microgrid energy management with a global control agent.

tential to more efficiently solve the MMG dispatch problem. This paper proposes an innovative RL for the MMG dispatch problem, where a global agent employs local agents in a distributed manner to learn a policy to minimize the total operation cost. In particular, local agents are used for distributed exploration based on local states within individual microgrid. On the other hand, the global agent serves two roles: i) aggregation and Q-table update during exploration and ii) taking actions to minimize the total cost at the system level during exploitation. The proposed hybrid learning approach leads to effective exploration of the solution space and guides the global agent to learn the dispatch policy efficiently. To evaluate the performance of the proposed approach, case studies are performed on a system with three microgrids with different types of DERs. The results in comparison with the conventional approaches validate the effectiveness of the proposed approach in terms of operation costs, computation time, and peak-to-average ratio (PAR).

The rest of this paper is organized as follows. The model description and problem formulation are presented in Section II. Section III presents the proposed RL approach for MMG energy management. The proposed RL is evaluated through case studies in Section IV. Finally, the conclusions are drawn in Section V.

II. Multi-Microgrid System Model and Problem Formulation

This section presents the mathematical models used for the MMG energy management problem. The interaction between the individual microgrid controllers and system operator is illustrated in Fig. 1. The microgrid data and dispatch decisions are exchanged between the microgrids and the operator through a bi-directional communication channel. The central operator can purchase energy from the main grid or use DERs to meet the microgrid demand. DERs considered in this paper include distributed renewable generation (RG) such as photovoltaic (PV) and wind, battery energy storage system (BESS) assets, and dispatchable distributed generators (DGs) such as diesel engines (DEs) and fuel cells (FCs).

The MMG energy management problem is to make hourly dispatch decisions to minimize the total operation cost. The objective function is the sum of the operation cost of M microgrids over a time period of T (typically 24 hours), as expressed in (1):

$$\min \sum_{t=1}^{T} \sum_{m=1}^{M} \left[\sum_{k \in \mathcal{K}^m} C_k^{\text{dg}}(p_{t,k}^{\text{dg}}) + C^{\text{ex}}(p_{t,m}^{\text{grid}}) \right], \quad (1)$$

where m is the microgrid index, t is the hour index, $\mathcal{K}^m$ is a set that contains all DGs in microgrid m, and $C_k^{\text{dg}}(p_{t,k}^{\text{dg}})$ is the operational cost of DG k, which can be expressed as

$$C_k^{\text{dg}}(p_{t,k}^{\text{dg}}) = d_{t,k} \left[(a_k (p_{t,k}^{\text{dg}})^2 + b_k p_{t,k}^{\text{dg}} + c_k \right], \quad (2)$$

where $p_{t,k}^{\text{dg}}$ is DG k power output, a_k, b_k, and c_k are the coefficients of the quadratic function, and $d_{t,k}$ is a binary variable that indicates the ON/OFF status of DG k. The power purchase cost or sell revenue $C^{\text{ex}}(p_{t,m}^{\text{grid}})$ is expressed as

$$C^{\text{ex}}(p_{t,m}^{\text{grid}}) = \begin{cases} p_{t,m}^{\text{grid}} \lambda_t \eta_m \Delta t, & \text{if } p_{t,m}^{\text{grid}} \geq 0 \\ \frac{p_{t,m}^{\text{grid}} \lambda_t \Delta t}{\eta_m}, & \text{if } p_{t,m}^{\text{grid}} < 0 \end{cases} \quad (3)$$

where λ_t is the retail price at the point of common coupling, $p_{t,m}^{\text{grid}}$ is the power purchased from (positive) or sold to (negative) the main grid, Δt is the time step size, and η_m is used to capture network losses, which can differ among microgrids depending on their locations within the distribution network. In this way, each microgrid receives different retail electricity prices due to different losses, which is known as distribution locational marginal price [18].

The microgrid resources must be dispatched considering both microgrid- and component-level constraints [19]. First, the power balance of microgrid m must be satisfied all the time:

$$\sum_{b \in \mathcal{B}^m} p_{t,b}^{\text{batt}} + \sum_{k \in \mathcal{K}^m} p_{t,k}^{\text{dg}} + p_{t,m}^{\text{grid}} + w_{t,m} = l_{t,m}, \quad (4)$$

where $\mathcal{B}^m$ is a set that contains all BESSs in microgrid m, $w_{t,m}$ is the RG output assuming RG is operated in the maximum power point tracking mode to maximize the environmental and economic benefits, and $l_{t,m}$ is the load of the m-th microgrid. The DG power output constraint is expressed in (5):

$$\phi_k^{\text{dg}} p_k^{\text{rated}} d_{t,k} \leq p_{t,k}^{\text{dg}} \leq p_k^{\text{rated}} d_{t,k}, \quad (5)$$

where ϕ_k^{dg} is the minimum power generation limit as a percentage of rated power p_k^{rated} of DG k.

The BESS operation follows the linear model described in [20]. In particular, the BESS power limits are provided in (6):

$$-P_b^- \leq p_{t,b}^{\text{batt}} \leq P_b^+, \quad (6)$$

where $p_{t,b}^{\text{batt}}$ is the charging/discharging power (positive when discharging) of BESS b, P_b^+ is the maximum discharging

Authorized licensed use limited to: Florida Atlantic University. Downloaded on January 13,2023 at 19:42:49 UTC from IEEE Xplore. Restrictions apply.

power, and P_b^- is the maximum charging power. The BESS state of charge (SOC) transition function is given in (7).

$$s_{t+1,b} = s_{t,b} - \frac{\Delta e_{t,b}}{E_b}, \tag{7}$$

where $s_{t,b}$ is the SOC of BESS b at the end of hour t, E_b is the rated energy capacity, and $\Delta e_{t,b}$ is the change of energy in BESS in hour t, which can be expressed as:

$$\Delta e_{t,b} = \begin{cases} \eta_b^- p_{t,b}^{\text{batt}} \Delta t, & \text{if } \ p_{t,b}^{\text{batt}} \leq 0 \\ \frac{p_{t,b}^{\text{batt}}}{\eta_b^+} \Delta t, & \text{otherwise} \end{cases} \tag{8}$$

where η_b^- and η_b^+ are the charging and discharging efficiencies, respectively. The SOC of the battery is constrained by (9):

$$\underline{s}_b \leq s_{t+1,b} \leq \overline{s}_b, \tag{9}$$

where $\underline{s}_b$ and $\overline{s}_b$ are the lower and upper bounds of SOC, respectively.

III. Proposed RL for Multi-Microgrid Energy Management

RL is a type of machine learning approach focusing on how agents interact with the problem environment with the goal of achieving the optimal policy in a sequential decision-making process [21]. In RL research, the sequential decision-making problem is formalized as a Markov decision process [22]. In this process, the RL agent observes the environment state, takes action, and receives an immediate reward from the environment. Then, the agent moves to the next state and iteratively repeats this process to find the optimal policy that minimizes the total cost [23].

The fundamental elements of a conventional centralized RL approach include:

- Agent: Operator
- State: $S_t = (S_t^1, ..., S_t^m, ..., S_t^M)$
- Action: $a_t = (a_t^1, ..., a_t^m, ..., a_t^M)$
- Reward/Cost: The cost r_t can be defined as the summation of the cost functions presented in (1).

Herein, S_t is the state information for the entire MMG system. The set of state variables of m-th microgrid is $S_t^m = (s_{t,b}, d_{t,k})$, which is a part of S_t. Similarly, a_t is the action taken by the operator at time t, with $a_t^m = (p_{t,b}^{\text{batt}}, p_{t,k}^{\text{dg}}, p_{t,m}^{\text{grid}})$ as a part of the action. The corresponding cost r_t can be obtained for taking action a_t from state S_t. For determining the next-state S_{t+1}, the SOC dynamics in (7) can be used for the battery, and the DG status can be updated using the DG status transitions presented in [24].

Q-learning is a model-free RL approach to learn the value of an action in a particular state through an iterative process and inform the RL agent what action to execute under certain circumstances [25], [26]. The Q-learning algorithm uses a function that calculates the value of a state-action combination:

$$Q \ : \ S \ \times \ a \rightarrow \mathbb{R}. \tag{10}$$

A Q-value is assigned for each state-action pair, and it indicates the quality of the action for the given state. The core of the Q-learning algorithm is the Bellman equation with the value iteration process, and the Q-value is usually updated using the weighted average of the old value and the new information. For example, when the agent receives the cost r_t and next-state S_{t+1} for taking action a_t from state S_t, the Q-value can be updated using the Bellman equation as follows:

$$Q^n(S_t, a_t) \leftarrow Q^{n-1}(S_t, a_t) + \alpha\left(r_t + \gamma \min_{a_{t+1}} Q^{n-1}(S_{t+1}, a_{t+1}) - Q^{n-1}(S_t, a_t)\right), \tag{11}$$

where $Q^n(S_t, a_t)$ is the Q-value for the given state-action pair at iteration n, α is the learning rate, and γ is the discount factor.

In the given problem formulation, the number of states increases exponentially with the number of microgrids in the system and downgrades the solution quality due to increased computation burden. To tackle the challenge, this paper proposes an innovative RL approach that follows the exploration and exploitation strategy and involves local agents within individual microgrids and a global agent at the system level. During exploration, local agents explore a random action considering the states within the microgrid and collect rewards. The global agent updates the value table based on the states, actions, and rewards received from local agents. During exploitation, only global agent is used to search for a least-cost action based on the Q-table and update Q-table based on the actions taken and rewards received. With this proposed hybrid exploration and exploitation process, the proposed approach can effectively explore the solution space and learn the MMG dispatch policy more efficiently compared to a conventional centralized approach. Additional descriptions of functions of local and global agents are provided as follows.

- *Local agents*
  The local agents are only active during exploration process. They interact with the microgrid environments in a distributed manner to collect states, randomly explore actions, and calculate the corresponding rewards, making them ready for the global agent for aggregation and value table update. The fundamental elements of local agents include:
  - Agent: Microgrid local agents Q^m
  - State: $S_t^m = (s_{t,b}, d_{t,k})$, $\forall m$
  - Action: $a_t^m = (p_{t,b}^{\text{batt}}, p_{t,k}^{\text{dg}}, p_{t,m}^{\text{grid}})$, $\forall m$
  - Cost: $r_t^m = \sum_{k \in \mathcal{K}^m} C_k^{\text{dg}}(p_{t,k}^{\text{dg}}) + C^{\text{ex}}(p_{t,m}^{\text{grid}})$
  - Next-state: $S_{t+1}^m = (s_{t+1,b}, d_{t+1,k})$, $\forall m$
- *Global agent*
  The global agent is active during both exploration and exploitation. During exploration, upon information received from local agents, aggregation is performed and value table is updated. During exploitation, the minimum cost policy is used to take actions and update Q-table using (11). The fundamental elements of the global agent include:

Authorized licensed use limited to: Florida Atlantic University. Downloaded on January 13,2023 at 19:42:49 UTC from IEEE Xplore. Restrictions apply.

TABLE I
TYPES OF DERS BY MICROGRID.

No.	Compositions
1	Wind, BESS
2	Wind, PV, BESS, DE
3	Wind, BESS, FC

- Agent: Operator Q
- State: $S_t = (S_t^1, ..., S_t^M)$
- Action: $a_t = (a_t^1, ..., a_t^M)$
- Cost: $r_t = \sum_{m=1}^{M} r_t^m$
- Next-state: $S_{t+1} = (S_{t+1}^1, ..., S_{t+1}^M)$

This process continues until the iteration number exceeds the limit. Once the training is completed, for any MMG state S_t, the corresponding action a_t can be determined as

$$a_t = \min_{a \epsilon \mathcal{A}} Q(S_t, a), \tag{12}$$

where $\mathcal{A}$ is the feasible action space that satisfies operational constraints of the microgrids presented in (4)–(9).

IV. SIMULATION ANALYSIS

In this section, the parameters of the MMG system and proposed RL approach used in the simulation studies are presented first. Next, case studies are presented to validate and evaluate the proposed RL approach in terms of operation costs, computation time, and PAR. The simulation results are evaluated and compared with two existing methods: i) cooperative Q-learning and ii) Monte Carlo method. In cooperative Q-learning (hereinafter referred to as "cooperative RL" for simplicity), the microgrid agents share their information in a cooperative mechanism and take the dispatch decisions based on the feasible action space of the MMG system [11], [13]. Thus, the microgrid agents learn their cooperative dispatch policy by interacting with the MMG environment in an iterative process. The Monte Carlo method is a widely used mathematical technique that generates random variables to treat optimization problems [27]. In the Monte Carlo method, random actions are generated based on the feasible action space $\mathcal{A}$ for solving the given MMG dispatch problem, and the solution is determined after N iterations based on the minimum MMG operation cost.

A. Simulation Setup

The test system is an MMG system with three microgrids connected to the main grid. Each microgrid consists of different DERs, and the compositions of DERs in the microgrids are summarized in Table I. Table II lists dispatchable DG and BESS parameters, which were adopted from [18], [28], [29] and slightly modified to align with the microgrid design.

For RG, the rated power of wind is assumed to be 100 kW for all microgrids for simplicity, and the power output profiles are generated using the System Advisory Model (SAM) [30] considering different manufacturers to add some variations. Similarly, the output profile of PV is obtained using SAM with

TABLE II
PARAMETERS OF DISTRIBUTED ENERGY RESOURCES.

DG Type	Range (kW)	a_k ($/kW^2$)	b_k ($/kW)	c_k ($)
Diesel Engine	[30 200]	0.00042	0.0185	0.4
Fuel Cell	[0 60]	0.00024	0.0267	0.38
Energy Storage	**Energy Capacity (kWh)**	**Rated Power (kW)**	**Char. Efficiency (%)**	**Dischar. Efficiency (%)**
Lithium-ion	160	40	91.5	91.5

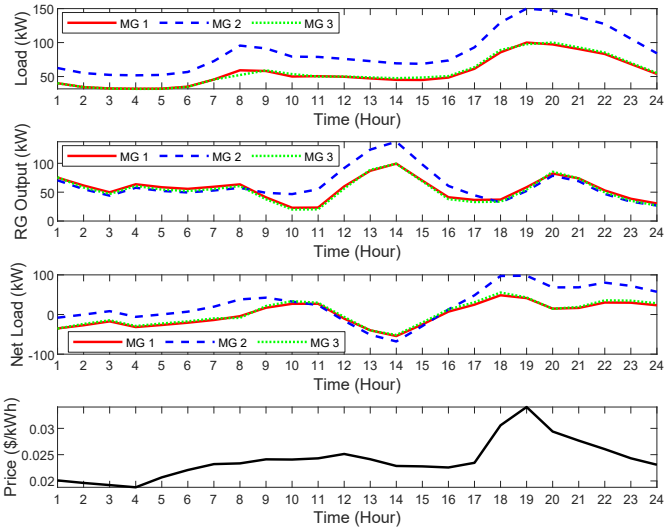

Fig. 2. Load, renewable generation, and net load of different microgrids in the MMG system. MG denotes microgrid.

a rated power of 50 kW for the city of Phoenix, Arizona. The residential load profiles in Phoenix from [31] were scaled to represent the microgrid loads with peaks of 100 kW, 150 kW, and 100 kW, respectively. The microgrid loads, total power outputs obtained from PV and wind, and the corresponding net loads on a typical winter day are plotted in Fig. 2, used for numerical experiments in this paper.

It is assumed that the distance of the microgrid from the point of common coupling increases with the serial number. The higher-distance microgrid suffers from more network losses, hence receiving a higher retail price. The η_m is set for the microgrids in the range of 1.01-1.05, with an interval of 0.02. The maximum and minimum battery SOCs ($\underline{s}_b$ and $\overline{s}_b$) are set to be 0.1 and 0.9, respectively.

A decayed ϵ-greedy strategy is used to balance exploration and exploitation for the RL approaches [32]. In this strategy, the ϵ initializes as 0.8 and divides the value by 1.3 at every 50 iterations until it reaches 0.01. The maximum iteration is set to be $N = 1000$. A fixed step-size of $\alpha = 0.5$ is used for updating the Q-values using (11). The Q-tables are defined as a cell array with respect to states, action, and time steps and initialize them with all zeros. For defining the states of microgrids, discretized battery SOC and DG ON/OFF status

Authorized licensed use limited to: Florida Atlantic University. Downloaded on January 13,2023 at 19:42:49 UTC from IEEE Xplore. Restrictions apply.

TABLE III
PERFORMANCE COMPARISON FOR MMG DISPATCH.

Approach	Cost ($)	% of Improvement	Time (minutes)
Proposed RL	22.4	70.5	1.2
Cooperative RL	27.2	40.4	3.1
Monte Carlo	38.2	-	3.3

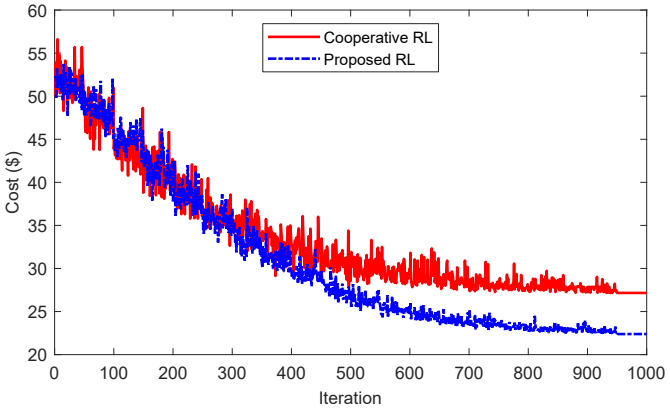

Fig. 3. Convergence performance of the proposed versus conventional cooperative RL.

TABLE IV
EVALUATION IN TERMS OF PEAK-TO-AVERAGE RATIO.

Approach	PAR		
	MG 1	MG 2	MG 3
Proposed RL	1.79	1.79	1.82
Cooperative RL	1.98	1.85	1.93
Monte Carlo	2.05	1.85	2.11

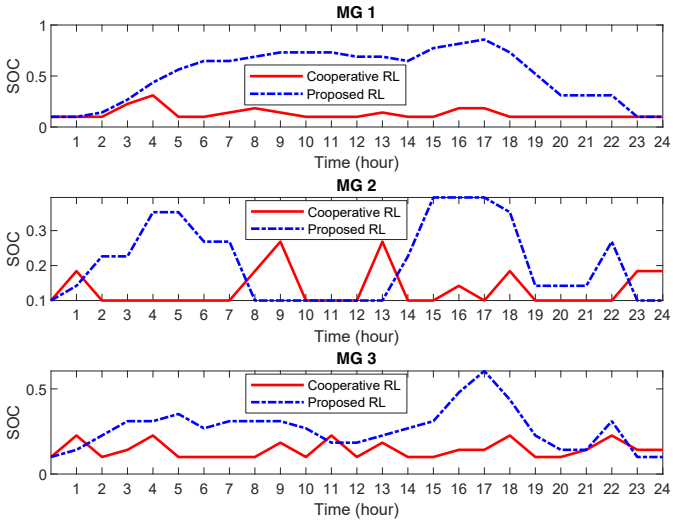

Fig. 4. Battery SOC profiles of different microgrids, obtained using the proposed and conventional RL approaches. MG denotes microgrid.

are used. If a microgrid does not have DGs, the DG state is defined with zero. All numerical experiments were performed in MATLAB $R2020b$ on a computer with an Intel Core i7-8665U 1.90 GHz CPU and 16 GB RAM.

B. Result Analysis

This section presents the results obtained from the simulation studies for evaluating the performance of the proposed RL approach in the MMG environment. Table III summarizes the performance results in terms of operation cost and computation time obtained using the proposed RL approach and the other conventional approaches.

As can be seen, the proposed RL solves the MMG dispatch problem in 1.2 minutes with the minimum operation cost, showing 70% improvement compared to the conventional Monte Carlo method. On the other side, with the same simulation settings, the conventional cooperative RL approach takes more than double the computation time to solve the problem and incurs $\$4.8$ in extra operating costs. The conventional Monte Carlo method shows the maximum operating cost during the experiment.

To better see the comparison, the cost versus number of iterations is plotted in Fig. 3 to show the training process of the two RL approaches and how they converge to the given solutions. As can be seen, due to the high initial exploration rate, both approaches involve exploring the solution space showing excessive cost fluctuations at the beginning. As the exploitation rate increases with the iteration, the improved performance of the proposed RL approach is revealed. As shown in the figure, the proposed RL approach shows a 5.9% improvement after 500 iterations, and the improvement increases to 21.4% at the end of 1000 iterations.

The microgrids' energy storage systems play a critical role in minimizing the operation cost by shaving the system peak and shifting energy to reduce PAR. The results in terms of PAR are summarized in Table IV. The results show that the proposed RL approach effectively dispatches the microgrid BESS and achieves the minimum PAR value for all microgrids compared to the conventional approaches. The performance of the RL approaches can also be evaluated through the BESS dispatch decisions made after the training. To better show the dispatch performance, the battery SOC profiles are plotted in Fig. 4 for both RL approaches.

As can be seen, the proposed RL approach efficiently charges the BESS of the microgrids with the negative net loads and is discharged to shave the loads, mainly around the evening time when the net load is the highest, to maximize the benefits of energy shifting. BESSs are poorly dispatched with the conventional RL approach, as indicated by the high operation cost. For example, in some hours, the BESSs are charged or on standby when discharging would be beneficial, and vice versa, adding extra operating costs to the system. Again, some RG is sold to the grid instead of being used to charge BESSs in some hours, which provided an instant benefit; however, it increased the total operation cost of the MMG system. Overall, the proposed RL approach is a

Authorized licensed use limited to: Florida Atlantic University. Downloaded on January 13,2023 at 19:42:49 UTC from IEEE Xplore. Restrictions apply.

promising technique for solving the MMG dispatch problems, and the learning performance is validated through simulation studies.

V. Conclusion

This paper proposes an innovative RL approach for solving an MMG dispatch problem. The proposed learning approach consists of both a global agent for system-level coordination and local agents for individual microgrids. The RL agent employs local learning agents to interact with microgrid environments in a distributed manner with a common goal and aggregates the outcomes to learn the dispatch policy for the MMG system. Through the distributed learning and the proposed aggregation process, the proposed RL effectively explores the solution space to learn the MMG dispatch policy with minimum operation cost. The proposed approach is evaluated and validated through case studies on an MMG system of three microgrids with different types of DERs. The results showed that the proposed RL approach outperforms the conventional approaches in terms of operation costs, computation time, and PAR value. An interesting research direction is to develop a distributed RL approach based on the policy iteration strategy for MMG dispatch considering distribution power flow.

References

[1] M. Glavic, R. Fonteneau, and D. Ernst, "Reinforcement learning for electric power system decision and control: Past considerations and perspectives," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6918–6927, 2017.

[2] S. A. Arefifar, M. Ordonez, and Y. A.-R. I. Mohamed, "Energy management in multi-microgrid systems—development and assessment," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 910–922, 2016.

[3] H. Karimi and S. Jadid, "Optimal energy management for multi-microgrid considering demand response programs: A stochastic multi-objective framework," *Energy*, vol. 195, p. 116992, 2020.

[4] C. Guo, X. Wang, Y. Zheng, and F. Zhang, "Optimal energy management of multi-microgrids connected to distribution system based on deep reinforcement learning," *International Journal of Electrical Power & Energy Systems*, vol. 131, p. 107048, 2021.

[5] Z. Zhu, K. W. Chan, S. Bu, B. Zhou, and S. Xia, "Real-time interaction of active distribution network and virtual microgrids: Market paradigm and data-driven stakeholder behavior analysis," *Applied Energy*, vol. 297, p. 117107, 2021.

[6] X. Fang, J. Wang, G. Song, Y. Han, Q. Zhao, and Z. Cao, "Multi-agent reinforcement learning approach for residential microgrid energy scheduling," *Energies*, vol. 13, no. 1, p. 123, 2020.

[7] X. Fang, Q. Zhao, J. Wang, Y. Han, and Y. Li, "Multi-agent deep reinforcement learning for distributed energy management and strategy optimization of microgrid market," *Sustainable Cities and Society*, vol. 74, p. 103163, 2021.

[8] E. Samadi, A. Badri, and R. Ebrahimpour, "Decentralized multi-agent based energy management of microgrid using reinforcement learning," *International Journal of Electrical Power & Energy Systems*, vol. 122, p. 106211, 2020.

[9] E. Foruzan, L.-K. Soh, and S. Asgarpoor, "Reinforcement learning approach for optimal distributed energy management in a microgrid," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5749–5758, 2018.

[10] F. Luo, Y. Chen, Z. Xu, G. Liang, Y. Zheng, and J. Qiu, "Multiagent-based cooperative control framework for microgrids' energy imbalance," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1046–1056, 2016.

[11] X. Liu, B. Gao, Z. Zhu, and Y. Tang, "Non-cooperative and cooperative optimisation of battery energy storage system for energy management in multi-microgrid," *IET Generation, Transmission & Distribution*, vol. 12, no. 10, pp. 2369–2377, 2018.

[12] P. Kofinas, A. Dounis, and G. Vouros, "Fuzzy q-learning for multi-agent decentralized energy management in microgrids," *Applied Energy*, vol. 219, pp. 53–67, 2018.

[13] W. Liu, P. Zhuang, H. Liang, J. Peng, and Z. Huang, "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2192–2203, 2018.

[14] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2016, pp. 1928–1937.

[15] G. Sartoretti, Y. Wu, W. Paivine, T. S. Kumar, S. Koenig, and H. Choset, "Distributed reinforcement learning for multi-robot decentralized collective construction," in *Distributed Autonomous Robotic Systems*. Springer, 2019, pp. 35–49.

[16] T. A. Nakabi and P. Toivanen, "Deep reinforcement learning for energy management in a microgrid with flexible demand," *Sustainable Energy, Grids and Networks*, vol. 25, p. 100413, 2021.

[17] T. Pu, X. Wang, Y. Cao, Z. Liu, C. Qiu, J. Qiao, and S. Zhang, "Power flow adjustment for smart microgrid based on edge computing and multi-agent deep reinforcement learning," *Journal of Cloud Computing*, vol. 10, no. 1, pp. 1–13, 2021.

[18] Y. Du and F. Li, "Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1066–1076, 2019.

[19] D. Wu, X. Ma, S. Huang, T. Fu, and P. Balducci, "Stochastic optimal sizing of distributed energy resources for a cost-effective and resilient microgrid," *Energy*, vol. 198, May 2020, 117284.

[20] D. Wu and X. Ma, "Modeling and optimization methods for controlling and sizing grid-connected energy storage: A review," vol. 8, pp. 123–130, 2021.

[21] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[22] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 2014.

[23] Y. Du and D. Wu, "Deep reinforcement learning from demonstrations to assist service restoration in islanded microgrids," *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 1062–1072, Apr. 2022.

[24] H. Shuai, J. Fang, X. Ai, J. Wen, and H. He, "Optimal real-time operation strategy for microgrid: An ADP-based stochastic nonlinear optimization approach," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 931–942, 2018.

[25] V.-H. Bui, A. Hussain, and H.-M. Kim, "Double deep q-learning-based distributed operation of battery energy storage system considering uncertainties," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 457–469, 2019.

[26] S. Paul, F. Ding, U. Kumar, W. Liu, and Z. Ni, "Q-learning-based impact assessment of propagating extreme weather on distribution grids," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.

[27] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2016, vol. 10.

[28] D. Wu, T. Yang, A. A. Stoorvogel, and J. Stoustrup, "Distributed optimal coordination for distributed energy resources in power systems," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 2, pp. 414–424, Apr. 2017.

[29] K. Mongird, V. Viswanathan, J. Alam, C. Vartanian, and V. Sprenkle, "Grid energy storage technology cost and performance assessment," Pacific Northwest National Laboratory, Tech. Rep. DOE/PA-0204, 2020.

[30] J. Freeman, N. Blair, D. Guittet, M. Boyd, B. Mirletz *et al.*, "System Advisor Model," Available: https://sam.nrel.gov/.

[31] National Renewable Energy Laboratory, *Commercial and residential hourly load profiles for all TMY3 locations in the United States*, 2014, http://data.openei.org/submissions/153.

[32] I. Sajedian, H. Lee, and J. Rho, "Double-deep Q-learning to increase the efficiency of metasurface holograms," *Scientific Reports*, vol. 9, 2019, 10899.

Authorized licensed use limited to: Florida Atlantic University. Downloaded on January 13,2023 at 19:42:49 UTC from IEEE Xplore. Restrictions apply.