Multi-Agent Reinforcement Learning for Distribution System Critical Load Restoration

Preprint

Yiyun Yao, Xiangyu Zhang, Jiyu Wang, and Fei Ding

National Renewable Energy Laboratory

Presented at the 2023 IEEE Power and Energy Society General Meeting
Orlando, Florida
July 16–20, 2023

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-84636
July 2023

Multi-Agent Reinforcement Learning for Distribution System Critical Load Restoration

Preprint

Yiyun Yao, Xiangyu Zhang, Jiyu Wang, and Fei Ding

National Renewable Energy Laboratory

Suggested Citation
Yao, Yiyun, Xiangyu Zhang, Jiyu Wang, and Fei Ding. 2023. *Multi-Agent Reinforcement Learning for Distribution System Critical Load Restoration: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5D00-84636. https://www.nrel.gov/docs/fy23osti/84636.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5D00-84636
July 2023

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office Agreement Number 37770. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Multi-Agent Reinforcement Learning for Distribution System Critical Load Restoration

Yiyun Yao, Xiangyu Zhang, Jiyu Wang, Fei Ding
National Renewable Energy Laboratory
Golden, CO, U.S.A
{yiyun.yao; xiangyu.zhang; jiyu.wang; fei.ding}@nrel.gov

***Abstract*—Grid resilience has become a critical topic recently because of the increasing occurrence of extreme events and the growing integration of intermittent renewable energy sources. To build a resilient distribution system, this paper develops a multi-agent reinforcement learning-based (MARL) method to coordinate distribution energy resources (DERs) dispatch, load pickup, and network reconfiguration for load restoration after a system outage. With the help of two types of control agents, namely critical load restoration (CLR) and coordination (COR) agents, system loads can be restored efficiently, given available resources. The effectiveness and superiority of the proposed algorithm are demonstrated through simulations and comparative studies on a real distribution feeder in Western Colorado.**

***Index Terms*-- distribution system, grid resilience, load restoration, multi-agent reinforcement learning.**

I. INTRODUCTION

Although efforts have been made to enhance the grid infrastructure and heighten power supply reliability for years, we have witnessed an increasing trend in outages caused by extreme weather and natural disasters. After a power outage, loads need to be restored as soon as possible to satisfy basic societal needs. To this end, the increasing penetration of distributed energy resources (DERs) brings the capabilities of providing emergency power and assisting grid load restoration. However, their volatile and inconsistent behavior makes control and strategy-making more complex and challenging, especially during catastrophic incidents. As a result, it is of great significance to develop load restoration solutions [1], which allow coordination among DERs dispatch, load pickup, and network reconfiguration to support post-event load restoration.

To achieve a coordinated control scheme, methods based on optimal power flow (OPF) have been developed to obtain optimal restoration. Reference [2] investigated the collaboration of various DERs and legacy devices in distribution system service restoration, which is formulated as a mixed-integer, second-order cone programming problem. In [3], the authors formulated an islanding strategy in the event of line failures in distribution systems, and they propose a decentralized, multi-agent system to control the DERs. Recognizing that OPF is nonconvex and nondeterministic polynomial-time hard (NP-hard), the solutions generally rely on convex surrogates [4]. Besides, many of the required data (e.g., system models that include secondary, accurate load profiles on each node) might not be available, and the time required to solve the OPF models might not be consistent with distribution system dynamics.

As an effective alternative, reinforcement learning (RL) has been introduced to implement sequential decision-making. Different from OPF-based approaches, RL trains the control policy based on historical data prior to implementation, which lessens the reliance on the distribution system power flow model. Once the control policy is trained, it enables near real-time decision-making to match advanced sensing/ communications rates and hence enable the online application. Reference [5] considered the asynchronous data arrival using a deep Q-network (DQN). [1] explored the merits of curriculum leaning on facilitating the controller's training and enabling convergence to a better control policy. These strategies consider the scheme of a single centralized control agent with perfect bidirectional communications with all controllable components, which is reasonable and suitable for normal operating conditions. However, the limited communication capability after extreme events and the impacts of potential event propagation may jeopardize the foundation of the centralized scheme.

To overcome the aforementioned drawbacks, a multi-agent reinforcement learning-based (MARL) load restoration approach is developed. After a system outage, the distribution system is partitioned into multiple cells, which is a group of interconnected components that makes up the smallest subset of the grid. During the post-event restoration phase, each cell can disconnect from the main grid and be operated by a critical load restoration (CLR) agent. Additionally, cells can connect with each other (to be clustered) by a coordination (COR) agent to respond to event propagation and achieve advanced load restoration. This paper contributes the following:

1) A MARL-based control scheme is developed to coordinate DER control, load pickup, and network reconfiguration for load restoration in distribution grids.
2) Two types of RL control agents, namely CLR and COR, are developed to facilitate DERs dispatch and cell clustering, respectively.
3) The effectiveness of the proposed solution is demonstrated by conducting reliable load restoration on the model of a real distribution feeder in Western Colorado.

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office Agreement Number 37770. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.

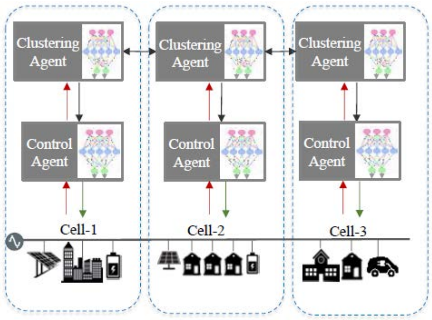

Figure. 1. Cell-based operational strategy for resilience enhancement.

II. Cell-Based Load Restoration for Distribution System Resilience Enhancement

A. Resilience Concept and Cell-Based Operational Strategy

Although with various emphases, two concepts are commonly shared when describing the perception of resilience [6]:

1) The capability of a system to resist, withstand, and adapt to a major disruption, albeit with reduced performance.
2) The capability of a system to recover to a normal state after a major disturbance.

To improve distribution grids resilience, a cell-based operational strategy [7] is introduced in this paper. The distribution network is partitioned into multiple cells. A cell is a group of interconnected DERs and buildings that makes up the smallest subset of the grid capable of operating using its own generation resources. After extreme events happen,

1) Each cell can disconnect from the main grid and operate in island mode to sustain the load supply with local DERs.
2) Cells can connect with each other to increase total load pickup and assist for recovery at the system level.

The cell-based scheme introduces two core functionalities (agent) for resilience enhancement, as shown in Fig. 1. One is the *control agent* that dispatches local DERs for load restoration at the cell level. The other is the *clustering agent* that configures cells connection at the system level.

B. Formulation of Cell-Based Load Restoration Problem

Based on the cell-based operation strategy, we consider a multi-step prioritized critical load restoration problem after cells are islanded from the main grid due to an extreme event. The goal is to restore as many loads as possible in the outage duration, denoted by discrete steps $t \in T = \{1, \ldots, T\}$, using local DERs in each cell $c \in \mathcal{C} = \{1, \ldots, N^c\}$. Loads $i \in \mathcal{L}$ are prioritized by importance factors $\mathscr{b}^i$, and $\boldsymbol{z} = [\mathscr{b}^1, \ldots, \mathscr{b}^N]$, $N = |\mathcal{L}|_1$. DERs include PV ($\mathcal{H}$), (mobile) battery ($\mathcal{B}$), and diesel generator ($\mathcal{D}$), and all together denoted as $\mathcal{G} = \mathcal{H} \cup \mathcal{B} \cup \mathcal{D}$. At each time step t, the CLR agent determines the power setpoints for all DERs (i.e., $p_t^{\mathcal{G}}, q_t^{\mathcal{G}}$) and the value of restored load (i.e., $p_t^{\mathcal{L}}, q_t^{\mathcal{L}}$). The COR agent determines the switcher operation and power exchange among connected cells (i.e., $w_t^{\mathcal{C}}\ p_t^{\mathcal{C}}, q_t^{\mathcal{C}}$). Let $x_t := (p_t^{\mathcal{G}}, q_t^{\mathcal{G}}, p_t^{\mathcal{L}}, q_t^{\mathcal{L}}, w_t^{\mathcal{C}}\ p_t^{\mathcal{C}}, q_t^{\mathcal{C}})$, the load restoration problem can be formulated as:

$$max_{x_t}: \quad \textstyle\sum_{t\in\mathcal{T}} \sum_{c\in\mathcal{C}} (r_{t,c}^{\mathrm{CLR}} + \mathscr{v}_{t,c}) \tag{1}$$

$$s.t.: \quad r_{t,c}^{CLR} = z_c{}^{\top} p_{t,c}^{\mathcal{L}} - z_c{}^{\top} diag\{\epsilon_c\} \left[p_{t-1,c}^{\mathcal{L}} - p_{t,c}^{\mathcal{L}}\right]^{+}, \forall c \in \mathcal{C}, \tag{2}$$

$$\mathscr{v}_{t,c} = -\lambda 1_{N_{b,c}}^{\top} \mathrm{diag}\{\delta \mathrm{v}_{t,c}\} \delta \mathrm{v}_{t,c}, \forall c \in \mathcal{C}, \tag{3}$$

$$\delta \mathrm{v}_{t,c} = \left[\mathrm{v}_{t,c} - \overline{\mathrm{v}}\right]^{+} + \left[\underline{\mathrm{v}} - \mathrm{v}_{t,c}\right]^{+}, \forall c \in \mathcal{C}, \tag{4}$$

$$f(p_t^{\mathcal{G}}, q_t^{\mathcal{G}}, p_t^{\mathcal{L}}, q_t^{\mathcal{L}}, w_t^{\mathcal{C}}, p_t^{\mathcal{C}}, q_t^{\mathcal{C}}, \mathrm{v}_{t,c}) = 0, \tag{5}$$

$$g\left(w_t^{\mathcal{C}}\right) = 0, \tag{6}$$

$$\underline{p_t^{\mathcal{G}}} \leq p_t^{\mathcal{G}} \leq \overline{p_t^{\mathcal{G}}}, \underline{q_t^{\mathcal{G}}} \leq q_t^{\mathcal{G}} \leq \overline{q_t^{\mathcal{G}}}, \tag{7}$$

where $r_{t,c}^{\mathrm{CLR}}$ and $\mathscr{v}_{t,c}$ represent the single-step load restoration (CLR) reward and voltage violation penalty in cell c. The c in some variables' subscript indicates they are subsets of the system-level variables (e.g., z_c vs. z). $N_{b,c}$ indicates the number of buses in cell c. $\mathrm{v}_{t,c}$ denotes the voltage magnitude, and $\underline{\mathrm{v}}$ and $\overline{\mathrm{v}}$ are the lower and upper limits for the voltage magnitude, i.e., 0.95 and 1.05, respectively. In (2), the first term encourages load restoration, and the second term penalizes shedding previously restored by factors of ϵ_c. This penalty facilitates a reliable and monotonic restoration and thus minimizes the impact of intermittent renewable generation. The value of ϵ can be adjusted to manage the strictness of the monotonic load restoration requirement. Specifically, the CLR controller should only restore load i if it can be sustained for the next $\epsilon_i + 1$ steps to obtain a positive reward. Eq (5)-(7) represents the power flow constraints, system topology requirements, and DER operational constraints [8, Section. II].

Two major concerns of the OPF (1)-(7) are discussed here:

1) *Nonconvex formulation*. Because constraints (5) are nonlinear and constraints (6) will introduce integer variables, OPF (1)-(7) is a nonconvex, nonlinear, NP-hard programming problem, which suffers from high computational complexity.
2) *Unavailable information*. Constraints (5) also require an accurate system model and load request profile along the whole scheduling horizon. Whereas in practical, system model with secondary (where most DERs are connected) and precise load request profile might not even exist.

III. An MARL-Based Load Restoration Approach

A. Bi-Level MARL Approach

To address these challenges, this paper proposes a novel bi-level MARL-based solution, as shown in Fig.2. In the upper-level design, the COR agent coordinates the cells for reconfiguration to be adapted to system changes and to achieve higher

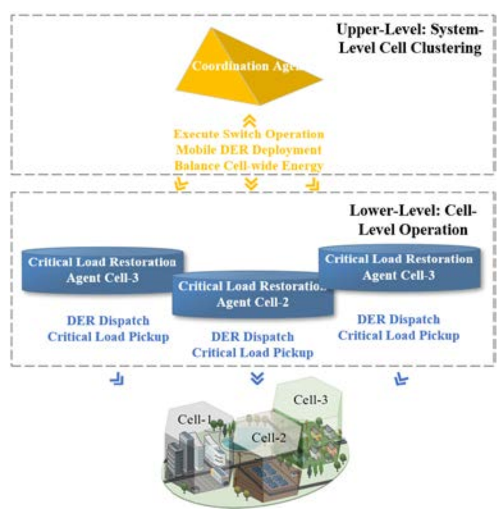

Figure. 2. Scheme of the bi-level MARL-based approach.

service restoration at system level. Specifically, the COR agent will (1) manage each cell with its aggregated generation resources and load requests, (2) decide the switch operation to connect cells by balancing total generation and load requests at the system level, (3) deploy mobile battery to cells based on the gap between generation resources and critical load request.

In the lower-level design, a CLR agent is deployed in each cell to dispatch DERs and determine the restored load. CLR agents operate the assigned cells independently—no mutual communications are required. Each of the CLR agents only reports the cell-level aggregated generation and load requests to COR. Based on the cells status (generation-consumption balance), COR agent will merge cells if the connection can increase system-level total load restoration, then deploy switcher actions and send load pickup adjustment to CLR agents.

To this end, we formulate this multi-cell load restoration problem as a cooperative multi-agent problem with $N^c + 1$ agents. Specifically, there are N^c CLR agents (one for each cell) and one COR agents coordinates the power transfer among cells.For CLR and COR agents, two key elements of the Markov decision process (MDP) are:

- **State $\boldsymbol{s}_{t,c}$**: In cell c, the state at step t includes the observation of available PV power, load requests, supplied load value, the battery's state of charge (SOC), the remaining fuel of the diesel generators, and a natural time index.
- **Action $\boldsymbol{a}_{t,c}$**: The action at step t includes active and reactive power dispatch of PV inverters, diesel generators, batteries, and load pickup decisions in the cell.

For the COR agent, the MDP state and action are:

- **State $\boldsymbol{s}_{t,N^c+1}$**: For the COR agent, the state set at step t includes the observation of the total DERs generation decision made by each of CLR agents, the total critical load request of each cell, and the switcher decision made in the upper-level clustering process.
- **Action $\boldsymbol{a}_{t,N^c+1}$**: The action at step t is the adjustment of the total load pickup decision for each cell.

As the problem is formulated as purely cooperative, a single reward r_t signal is provided to all agents to evaluate the current step of cooperative control. At each step $t \in \mathcal{T}$, the reward received by each agent is given:

$$r_t = \sum_{c \in C} (r_{t,c}^{\text{CLR}} + \mathcal{v}_{t,c}) / (N^c + 1). \tag{8}$$

By formulating the problem into MDP, the constraints (5) are enforced by OpenDSS simulator, which is a part of OpenAI Gym environment. Constraints (6) are enforced by COR agent. Constraints (7) are handled by the design of CLR action space.

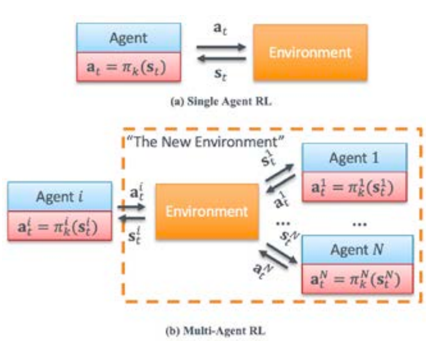

Figure. 3. Comparison between single agent RL and multi-agent RL.

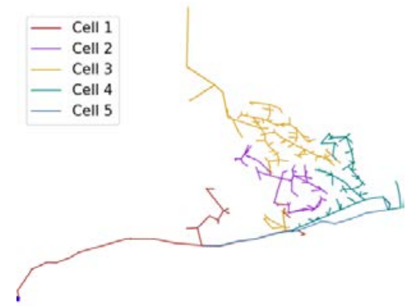

Figure. 4. A real feeder located in Western Colorodo

The violations will be reflected as penalty during the (MA)RL training [1].

B. Multi-agent Reinforcement Learning Algorithm

Compared with single-agent RL, one key challenge in multi-agent RL is the *non-stationarity*, as illustrated in Fig. 3. Specifically, for single-agent learning, the agent collects experience by interacting with the environment, and then using the experience to improve its policy. During training, the environment stays the same. In contrast, in the MARL case, from the perspective of any single agent, the environment is no longer stationary since other agents impact the environment and their policies evolve at the same time, leading to a non-stationary "new environment" as shown in Fig 3(b). Such non-stationarity will result in performance being not solely explainable by changes in a single agent's policy. Therefore, directly using algorithms from the single agent RL realm can lead to unstable learning. As a result, this paper utilizes a MARL specialized algorithm called *multi-agent deep deterministic policy gradient (MADDPG)* method [9]. The MADDPG method addresses the non-stationarity issue by conducting a coordinated training, making agents aware of other agents' behavior during the training. Specifically, an implementation framework called "*centralized training and decentralized execution*" (CTDE) is proposed. The decentralized execution aspect ensures the trained policy can still make decision in a distributed manner using only local data as inputs, i.e., follows $\boldsymbol{a}_{t,c} = \pi_c(\boldsymbol{s}_{t,c})$. The centralized training aspect aims at coordinating the training and is achieved by learning a centralized action-value approximator, i.e., the Q-function. Specifically, for agent c, instead of learning $Q_c(\boldsymbol{s}_{t,c}, \boldsymbol{a}_{t,c})$ as done in the single agent RL algorithm, the Q-func takes $Q_c(\boldsymbol{s}_{t,1}, \boldsymbol{a}_{t,1}, \dots, \boldsymbol{s}_{t,c}, \boldsymbol{a}_{t,c}, \dots, \boldsymbol{s}_{t,N^c+1}, \boldsymbol{a}_{t,N^c+1})$.

IV. Numerical Tests

To demonstrate how the proposed load restoration approach improves the distribution grid resilience by leveraging DERs, we implement tests on the model of a real feeder located in

TABLE I.
Cell-Wise Generation and Load Information

cell	gen. cap. (kW)	ld. cap. (kW)	cri-ld. cap. (kW)	ld. #
1	196	211	77	16
2	416	418	187	42
3	1110	838	411	98
4	534	534	250	64
5	120	120	2	4

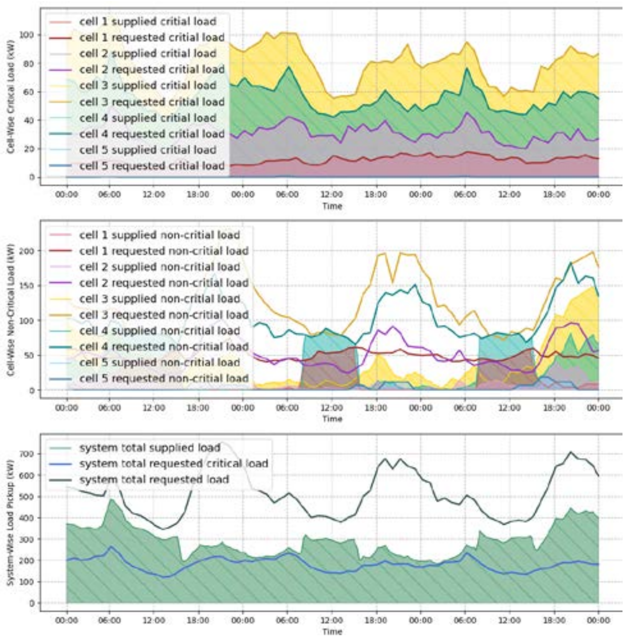

Figure. 5. Implement CLR at cell level only (case 1)

Western Colorado. The feeder consists of 759 nodes, including both the primary and secondary nodes, and it has a peak load of 2,121 kW, with 40% of this (927kW) as critical load.

To evaluate the impacts of DERs, load nodes are randomly picked to deploy PVs, batteries, or diesel generators. The total DERs generation capacity is 2,376 kW, which includes 112 PVs, 10 diesel generators, 95 batteries and 1 mobile battery. Batteries are fully charged and can support discharging at rated power for 2.7 hours.

For network partition, the feeder is divided into 5 cells, as shown in Fig.4. Cells 1 and 5 are three-phase cells along the backbone. Cell 2 is a single-phase (C-phase) cell. Cells 3 and 4 are single-phase (B-phase) cells with the most customer load. Table I reports the DERs and load information in each cell.

A 3-day load restoration was performed by using the data from November 28th – 30th, 2019. This is a 72-hour timeframe with 15-min time resolution, which collects 288 steps. For upper level clustering, the dispatch on switchers and deployment of mobile battery are executed every 6 hours (i.e., at 00:00, 6:00, 12:00, and 18:00). For load categorization, all critical loads are given the importance factor as 1, while all noncritical loads have importance factor of 0.1.

A. *Benefits of Implementing CLR Agents at Cell Level*

To validate the effectiveness and superiority of the proposed solutions, two comparative cases are also simulated:

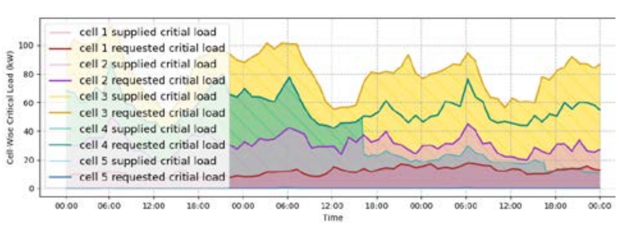

Figure. 6. Implement CLR at cell level with further event (case 2)

TABLE II.
COMPARISON OF LOAD RESTORATION UNDER CASE 0 AND 1

	critical load pickup (kWh)		noncritical load pickup (kWh)		total load pickup (kWh)	
	case0	case1	case0	case1	case0	case1
cell 1	242	841	840	1,268	1,082	2,109
cell 2	305	2,238	683	628	988	2,866
cell 3	1,445	5,880	1,593	3,020	3,038	8,900
cell 4	827	4,081	3,038	2,744	1,596	6,825
cell 5	4	11	42	230	46	241
system	2,823	13,051	3,937	7,890	6,750	20,941

Case 0: Baseline case where no control will be implemented. DER (if exists) can only supply power to local loads.

Case 1: CLR agents are deployed to operate each of cells independently. There is no coordination among cells.

Fig. 5 reports the supply of critical load, noncritical load, and system total load during the 3-day restoration for case 1. The top figure reports the restoration of critical load for all five cells. For example, the dark yellow curve is the critical load requests during the 3-day in cell 3, while the light-yellow-colored area is the restored critical load in cell 3. Similarly, the middle figure shows restoration of noncritical load. The bottom figure depicts the system-wise load restoration. The dark blue curve represents the system's total load request, the blue curve represents the system's total critical load request, and the light-green-colored area represents the system's total restored load. It can be observed that the CLR agents can dispatch DERs within each cell to supply 100% critical load during the 3-day restoration, as shown in Fig. 5 (top). Besides, the partial noncritical load can also be restored, shown in Fig. 5 (middle).

Table II compares the load restoration results between case 0 and 1. For case 0, since there is no control/coordination, all the load with local battery can be supplied till around 5:00 on day 1. Besides, all the load with local PV can be supplied during solar hours. Across the 3-day restoration, there are 2,823 kWh of critical load, 3,937 kWh of noncritical load, and 6,750 kWh of total load get restored. While for case 1, there are 13,051 kWh of critical load, 7,890 kWh of noncritical load, and 20,941 kWh of total load get restored.

The results shows that cells are formed by balancing the load and generation capability. CLR agents can effectively restore all the critical load by dispatching local DERs at cell level.

B. *Benefits of Implementing Bi-Level Strategy*

Since the COR agent is developed to enhance load restoration at the system level and provide the responding capability to prolonged events, further tests with event propagation,

TABLE III.
COMPARISON OF LOAD RESTORATION UNDER CASE 1 AND 2

	critical load pickup (kWh)		noncritical load pickup (kWh)		total load pickup (kWh)	
	case2	case3	case2	case3	case2	case3
cell 1	841	840	1,268	1,040	2,109	1,880
cell 2	2,238	2,238	628	970	2,866	3,208
cell 3	5,880	5,878	3,020	2,592	8,900	8,470
cell 4	2,980	4,024	1,616	1,615	4,596	5,638
cell 5	11	11	230	150	241	161
system	11,950	12,990	6,762	6,367	18,712	19,357

which cause unbalance between generation and load, is required to justify the benefits of bi-level CLR and COR design. In this case, we simulate an interruption event that will cause 80% generation DERs (PV, battery, diesel generators) in cell 4 to be offline from 16:00, day 2 to the end of day 3; the disaster that causes the interruption can vary from the winter storms, tropical cyclones, etc. However, the request of the load remains during the 3-day period, which causes the generation-consumption unbalance in cell 4. To validate the effectiveness and benefits of the proposed bi-level CLR and COR strategy, two more comparative cases are simulated:

Case 2: A simulated event interruptus DERs in cell 4. CLR agents are operating independently (no COR agent).

Case 3: A simulated event interruptus DER in cell 4. CLR agents are deployed to operate each of cells. COR agent is deployed to coordinate the reconfiguration among cells.

Fig. 6 depicts the 3-day critical load restoration under case 2. It can be observed that the interruption starting at 16:00 causes all noncritical load and more than 60% of the critical load to be curtailed in cell 4, while all the other cells are operating without any interruptions.

Fig. 7 depicts the 3-day load restoration after introducing COR agent as case 3. After the interruption at 16:00 in cell 4, the upper level COR agent closes the switcher to connect cell 3 and 4 and deploys the 150 kWh mobile battery to cell 4 at 18:00. Since then, all the noncritical load in cell 3 get curtailed, and the corresponding generation capability is transferred to cell 4 to supply the critical load, which can be observed as the recovering cliff in Fig. 7 (upper). However, not all the critical load in cells 3 and 4 are 100% satisfied; the upper level further closes switchers to connect cell 3 and 1 at 0:00 day 3, and connect cells 1, 2, and 5 at 6:00 day 3. From Fig. 7 (lower), It can be observed that from 0:00 to 6:00 on day 3, most of the noncritical load in cells 1, 3, and 4 are curtailed, and the corresponding power is transferred to cell 4 to maintain the critical load supper within it. The cells reconfiguration and deployment of mobile battery are reported in Fig. 8. The same color indicates that cells are merged to one larger cell by COR agent. Table III compares the load restoration between cases 2 and 3. After introducing COR agent, case 3 restored 1,040 kWh more in critical load than case 2 did at system level, but only restored 395 kWh less in noncritical load. Consequently, case 3 restored

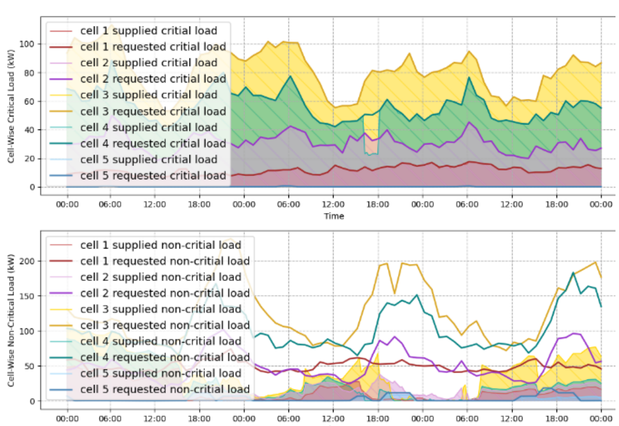

Figure. 7. Implement bilevel scheme with further event (case 3)

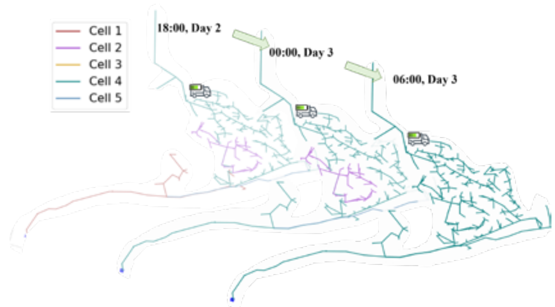

Figure. 8. MARL cell reconfiguration. Once two cells are connected, only one color is used to illustrate the larger merged cell.

19,357 kWh during the 3-day restoration, 645 kWh more than case 2 restored, which is 18,712 kWh.

Different to CLR, COR agent aims to balance the DERs generation and load request at system level. Whenever there is unbalance between generation and critical load (which could be caused by event propagation), the COR agent can connect the cells to transfer the power of supplying noncritical load in other cells to the unbalanced cell for supplying critical load, and thus achieve an advanced load restoration at the system level.

V. Conclusions

This paper proposes a novel MARL-based solution to solve load restoration in distribution grids. Numerical results show that the proposed CLR agents can effectively restore all the critical load by dispatching local DERs at cell level. Further more, the COR agent can connect the cells to transfer the power of supplying noncritical load in other cells to the unbalanced cell for the supply of critical load, and thus achieve an advanced load restoration at the system level.

References

[1]. X. Zhang, A. T. Eseye, B. Knueven, W. Liu, M. Reynolds, and W. Jones, "Curriculum-based Reinforcement Learning for Distribution System Critical Load Restoration," *IEEE Transactions on Power Systems*, pp. 1–13, Sep. 2022.

[2]. N. C. Koutsoukis, P. Georgilakis, and N. D. Hatziargyriou, "Service restoration of active distribution systems with increasing penetration of renewable distribution generation," *IET Gener. Transm. Distrib*, vol. 13, no. 14, pp. 3177–3187, 2019.

[3]. W. Li *et al.*, "A Full Decentralized Multi-Agent Service Restoration for Distribution Network with DGs," *IEEE Trans Smart Grid*, vol. 11, no. 2, pp. 1100–1111, 2020.

[4]. G. Huang, J. Wang, C. Chen, J. Qi, and C. Guo, "Integration of Preventive and Emergency Responses for Power Grid Resilience Enhancement," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4451–4463, 2017.

[5]. J. C. Bedoya, Y. Wang, and C. C. Liu, "Distribution System Resilience under Asynchronous Information Using Deep Reinforcement Learning," *IEEE Transactions on Power Systems*, vol. 36, no. 5, pp. 4235–4245, Sep. 2021.

[6]. J.-P. Watson, et al., "Conceptual framework for developing resilience metrics for the electricity, oil, and gas sectors in the United States," Sandia National Lab. (SNL-NM), Albuquerque, NM, SAND-2014-18019, 2014.

[7]. A. Sahu, K. Utkarsh, and F. Ding, "A Fast and Scalable Genetic Algorithm-Based Approach for Planning of Microgrids in Distribution Networks," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*, Jul. 2022, pp. 1–5.

[8]. W. Liu and F. Ding, "Hierarchical Distribution System Adaptive Restoration With Diverse Distributed Energy Resources," in *IEEE Trans. on Sustainable Energy*, vol. 12, no. 2, pp. 1347-1359, April 2021.

[9]. Lowe, Ryan, et al., "Multi-agent actor-critic for mixed cooperative-competitive environments." in *Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 1–12.