A Reinforced Learning Approach to Dispatch Distributed Generators for Enhanced Resilience

Michael Abdelmalak[1], MD Kamruzzaman[2], Sean Morash[3], Aaron F. Snyder[3], and Mohammed Benidris[1],
[1]Department of Electrical and Biomedical Engineering, University of Nevada-Reno, Reno, NV 89557, USA
[2]Cypress Creek Renewables, Asheville, NC 28801, USA
[3]EnerNex LLC, Knoxville, TN 37932, USA
Emails: {mabdelmalak, mkamruzzaman}@nevada.unr.edu, {smorash, aaron}@enernex.com, mbenidris@unr.edu

2022 IEEE Power & Energy Society General Meeting (PESGM) | 978-1-6654-0823-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/PESGM48719.2022.9916963

***Abstract*—This paper proposes a reinforced learning-based approach for dispatching distributed generators to enhance operational resilience of electric distribution systems against hurricanes. Existing resilience enhancement approaches rely on solving large-scale optimization problems that are computationally expensive and time demanding, which are not suitable for real-time applications. In this paper, a multi-agent framework is developed using a Soft Actor Critic algorithm to dispatch distributed generators for resilience enhancement. The proposed approach provides a fast-acting control algorithm that determines the size and the location of distributed generators to reduce the amount of load curtailment during hurricanes. The problem is formulated as a Markov decision process that consists of system states, an action space, and a reward scheme. A system state represents the system topology and characteristics upon which an action is taken and a reward value is calculated. An iterative Markov decision process is used to train the proposed Soft Actor Critic algorithm using multiple line outages generated from a hurricane fragility model. The trained network dispatches distributed generators whenever there are islanded grids and load curtailments. The proposed method is demonstrated on the IEEE 33-node distribution feeder system. The results show the capability of the proposed algorithm to determine optimal sizes and locations of distributed generators for resilience enhancement.**

***Index Terms*—Distribution system, extreme weather events, reinforced learning, resilience.**

I. Introduction

The frequency and intensity of extreme weather events have increased dramatically in recent years yielding prolonged outages and noticeable economic losses [1], [2]. In 2008, 200 million people in China experienced a severe ice storm resulting in direct losses exceeding 2 billion U.S. dollars [3]. Superstorm Sandy of October 2012 caused over 8 million customers to lose power across 15 states in the United States [4]. Various studies have been conducted to provide proactive, corrective and restorative operational resilience enhancement strategies at both the transmission and distribution levels [5], [6]. Corrective strategies aim to make the proper decisions during an extreme weather event to mitigate or reduce the negative impacts of that event on the system performance [7]. However, determining such decisions in a fast-paced manner is computationally expensive and time consuming. Also, studying large-scale systems considering numerous dynamic constraints and modeling stochastic behavior of component failure increases the complexity of finding an optimal solution and imposes dimensionality limitations [8], [9]. Therefore, implementing a resilience enhancement strategy that considers the aforementioned constraints has become more important than ever before.

Reinforced learning (RL) has been used to provide a fast-acting control algorithm for high-dimension stochastic optimization problems [10]. Several deep reinforced learning (DRL) methods have been proposed to improve resilience of electric power systems [11]. In [12], a hybrid actor-critic algorithm has been used to determine locations and sizes of shunt resources to maintain voltage levels within permissible limits due to multiple line failures. A multi-agent deep reinforcement learning approach has been developed to optimize the control operation of a microgrid after a disaster [13]. A resilience-based protection scheme utilizing reinforcement learning improved the efficiency of microgrid operation considering market participation and stochastic behavior of renewable energy sources [14]. In [15], a DRL method has determined the optimal rescheduling strategy of generation resources for transmission systems impacted by hurricanes. Authors of [16] have developed a DRL method that provides real-time operation decisions to optimally dispatch distributed energy resources installed at specific locations for restoring power to customers after sudden outages. However, the role of integrating continuous and discrete actions through a multi-agent DRL framework is still under investigation.

This paper proposes an DRL-based approach to control distributed generators (DGs) to enhance the operational resilience of power systems during hurricanes. The proposed algorithm is developed based on dispatching DGs to reduce the amount of load curtailments. It also considers proper sizing of DGs to avoid additional operation costs. A Soft Actor Critic (SAC) algorithm is used to develop a multi-agent framework that controls generation dispatch under single or multiple line outage conditions. In the proposed method, the power grid is split into various regions where each region is assigned to an agent. A Markov Decision Process (MDP) is formulated to determine the generation dispatch of each agent at each time instant given a system state. A reward scheme is developed to train the agent for better decision making. The algorithm is trained using a hurricane fragility model

978-1-6654-0823-3/22/$31.00 ©2022 IEEE

of transmission lines. The trained algorithm provides a set of corrective control actions to reduce the amount of load curtailments and to maintain sizes of DGs within a permissible range. The proposed algorithm is tested on the IEEE 33-node distribution feeder for validation.

The rest of the paper is organized as follows: Section II describes the multi-agent Soft Actor Critic, Section III describes the hurricane fragility model and formulates the Markov decision process for minimizing load curtailments, Section IV illustrates the implementation procedures on the IEEE 33-node distribution feeder and discusses the results, and Section V provides concluding remarks.

II. Multi-Agent Soft Actor Critic

A. Policies for Actor-networks of the Proposed Framework

Each agent in the proposed multi-agent framework has one actor network to provide actions, which is developed using a squashed Gaussian distribution function [12]. The policy of the actor network to provide actions is expressed as follows:

$$\alpha_t^{ci} \sim \pi_{\xi^{ci}}(\alpha_t^{ci}|O_t^i), \tag{1}$$

where i represents the i^{th} agent of the multi-agent framework, O_t^i is the observation vector of the i^{th} agent at time t, α_t^{ci} is the provided action by the actor-network of the i^{th} agent, ξ^{ci} is the parameter for policy of the i^{th} agent, and $\pi_{\xi^{ci}}(\alpha_t^{ci}|o_t^i)$ is an unbounded Gaussian policy of the i^{th} agent. A squashing function needs to be applied on $\pi_{\xi^{ci}}(\alpha_t^{ci}|o_t^i)$ to bound actions of the i^{th} agent to a finite value.

B. Policy Training Algorithm for Actors

In the fundamental SAC algorithm, the policy is updated in each iteration to maximize the expected return and entropy (randomness measure of the policy). Following the same convention, policies of the proposed algorithm are also updated in each iteration. A value function, $V_{\psi^i}^{ci}(O_t^i)$, which is used to measure the soft value for policy of the i^{th} agent can be expressed as follows:

$$V_{\psi^i}^{ci}(O_t^i) = \mathop{\mathbb{E}}_{\alpha_t^{ci}\sim\pi_{\xi^{ci}}}\left[Q_\theta(s_t,\alpha_t^{ci},\alpha_t^{-ci}) - \alpha^{ci}\log\left(\pi_{\xi^{ci}}(\alpha_t^{ci}|O_t^i)\right)\right] \tag{2}$$

where ψ^i represents parameter of the value function network for the i^{th} agent, θ represents parameter for the Q value function, $Q_\theta(s_t,\alpha_t^{ci},\alpha_t^{-ci})$ is a critic or centralized policy evaluation function for all the actors, α_t^{-ci} is the action provided by actors of agents except agent i, α^{ci} represents a parameter to determine the relative importance between reward and entropy of the i^{th} agent, and s_t is a set for system states.

The expression provided in (3) is used to minimize the residual squared error of a soft Bellman function to train value functions of the actors.

$$J_v^{ci}(\psi^i) = \mathbb{E}_{s_t^{ci}\sim\mathcal{D}}\left[\frac{1}{2}V_{\psi^i}^{ci}(O_t^i) - \left[Q_\theta(s_t,\alpha_t^{ci},\alpha_t^{-ci}) - \alpha^{ci}\log\left(\pi_{\xi^{ci}}(\alpha_t^{ci}|O_t^i)\right)\right]^2\right] \tag{3}$$

where $\mathcal{D}$ is a replay buffer to store experiences of the actors.

The gradient of (3) using an unbiased estimator is determined as follows to sample actions from the current policy:

$$\hat{\nabla}_{\psi^i}J_v^{ci}(\psi^i) = \nabla_{\psi^i}V_{\psi^i}^{ci}(O_t^i)\left(V_{\psi^i}^{ci}(O_t^i) - Q_\theta(s_t,\alpha_t^{ci},\alpha_t^{-ci}) + \alpha^{ci}\log\left(\pi_{\xi^{ci}}(\alpha_t^{ci}|O_t^i)\right)\right) \tag{4}$$

In this work, we have modified the expression for training the soft-Q parameters of the basic actor given in [17], which can be expressed as follows:

$$J_{Q_\theta}^{ci}(\theta^i) = \mathbb{E}_{(s_t^{ci},\alpha_t^{ci})\sim\mathcal{D}}\left[\frac{1}{2}\left(Q_\theta(s_t,\alpha_t^{ci},\alpha_t^{-ci}) - \hat{Q}(s_t,\alpha_t^{ci},\alpha_t^{-ci})\right)^2\right] \tag{5}$$

where

$$\hat{Q}(s_t,\alpha_t^{ci},\alpha_t^{-ci}) = r(s_t,\alpha_t^{ci},\alpha_t^{-ci}) + \beta\mathbb{E}_{s_{t+1}\sim p}[V_{\bar{\psi}^i}^{ci}(o_{t+1}^i)] \tag{6}$$

with $\beta \in [0,1]$ a discount factor and $\bar{\psi}^i$ an average of the weights for the value network of i^{th} agent.

The value of Q-function (5) is optimized as follows:

$$\hat{\nabla}_{\theta^i}J_{Q_\theta}^{ci}(\theta^i) = \nabla_{\theta^i}(Q_\theta(s_t,\alpha_t^{ci},\alpha_t^{-ci})\left(Q_\theta(s_t,\alpha_t^{ci},\alpha_t^{-ci}) - r(s_t,\alpha_t^{ci},\alpha_t^{-ci}) - \beta V_{\bar{\psi}^i}^{ci}(o_{t+1}^i)\right)^2 \tag{7}$$

The policy needs to be updated in each iteration to maximize the rewards for improving the policy. The authors of [17] have directed the policy update toward exponential of new soft Q-function as they intended to track the policy update. Also, the potential policies are restricted to a parameterized distribution (i.e., Gaussian) family. Following the same convention, we have updated the expression for policy update of basic SAC algorithm for the proposed algorithm as follows:

$$\pi_{\xi^{ci}}^{new} = \arg\min D_{KL}\left(\pi_{\xi^{ci}}(.|O_t^i)\middle|\middle|\frac{Q_\theta(s_t,.)}{Z_\theta(s_t)}\right) \tag{8}$$

where $Z_\theta(s_t)$ is an intractable partition function that does not contribute to the gradient with respect to the new policy.

The policy $\pi_{\xi^{ci}}(.|O_t^i)$ is parameterized for action setting using the policy network of agent i with parameter ξ^{ci}. Finally, the expected KL-divergence of (8) is multiplied by α^{ci} and then minimized, ignoring $Z_\theta(s_t)$ to train the policy parameters of agent i as follows:

$$J_{\pi_{\xi^{ci}}}^{ci}(\xi^{ci}) = \mathbb{E}_{s_t^{ci}\sim\mathcal{D}}\left[\mathbb{E}_{\alpha_t^{ci}\sim\pi_{\xi^{ci}}}\left[\alpha^{ci}\log\left(\pi_{\xi^{ci}}(\alpha_t^{ci}|O_t^i)\right) - Q_\theta(s_t,\alpha_t^{ci},\alpha_t^{-ci})\right]\right] \tag{9}$$

Although several options are available to minimize the objective function $J_{\pi_{\xi^{ci}}}^{ci}(\xi^{ci})$, the authors of [18] have applied the reparameterization trick to achieve target density (the Q-function). The modified expression to reparameterize the policy of agent i is as follows:

$$\alpha_t^{ci} = f_{\xi^{ci}}(\epsilon_t^{ci};o_t^i) \tag{10}$$

978-1-6654-0823-3/22/$31.00 ©2022 IEEE

where ϵ_t^{ci} is a noise vector that uses a spherical Gaussian distribution.

Thus, the new policy objective for agent i is as follows:

$$J^{ci}_{\pi_{\xi^{ci}}}(\xi^{ci}) = \mathbb{E}_{s_t^{ci}\sim\mathcal{D}}, \epsilon_t^{ci} \sim \mathcal{N}\left[\alpha^{ci}\log\left(\pi_{\xi^{ci}}(f_{\xi^{ci}}(\epsilon_t^{ci};o_t^i)|O_t^i)\right) - Q_\theta(s_t, f_{\xi^{ci}}(\epsilon_t^{ci};o_t^i), f_{\phi^{-ci}}(\epsilon_t^{-ci};s_t^{-i}))\right] \quad (11)$$

where $f_\phi^{-ci}(\epsilon_t^{-ci};s_t^{-i})$ is the parameterized policies of other actors.

In [19], the authors have provided a detailed formulation of an alternative approach to obtain the temperature parameter learning objective function, which is not strictly relevant to this work. However, we modify their temperature objective function for the actors of each agent of the proposed framework as follows:

$$J^{ci}(\alpha^{ci}) = \mathbb{E}_{\alpha_t^{ci}} \sim \pi_{\xi^{ci}}\left[-\alpha^{ci}\left(\log\left(\pi_{\xi^{ci}}(\alpha_t^{ci}|O_t^i) + \bar{H}\right)\right)\right] \quad (12)$$

where $\bar{H}$ is an equivalent constant vector of the hyper-parameter to represent target entropy. Equation (12) cannot be minimized directly due to the expectation operator. Therefore, it is minimized using a MC estimator after sampling experiences from a replay buffer based on the procedure from [19]. In the proposed multi-agent algorithm, two soft Q-networks for all agents are trained and then the minimum value among the outputs of the two Q-networks is used in the objective function of (12) to combat state-value overestimation [20].

III. The Proposed Dispatch Algorithm

This section describes the proposed RL-based approach to determine sizes and locations of DGs for resilience enhancement. First, it describes the impacts of hurricanes on system components then it explains the SAC algorithm to formulate and train agents for proper actions.

A. Impacts of Hurricane Progression

The propagation properties due to the spatiotemporal characteristics of hurricanes have unique impacts on the performance of system components. Several resilience-based studies have adopted fragility models to identify the potential impacted components [21], [22]; however, other studies have used actual events or forecasted failure scenarios [23]–[25]. In this work, the hurricane fragility model used in [24] has been adopted to determine the set of potential failures. The failure probability of a transmission corridor can be evaluated as follows:

$$P_i = 1 - \prod_1^M (1 - P_{i,m}) \prod_1^L (1 - P_{i,n}), \quad (13)$$

$$P_{i,n} = 1 - \exp\left\{-\sum_{j=0}^{N-1} \lambda_{i,n}(t_j)\Delta t\right\}, \quad (14)$$

$$P_{i,m} = 1 - \exp\left\{-\sum_{j=0}^{N-1} \lambda_{i,m}(t_j)/(1-\lambda_{i,m}(t_j))\Delta t\right\} \quad (15)$$

where P_i is the cumulative failure probability of i^{th} transmission corridor, $P_{i,m}$ the cumulative failure probability of the m^{th} tower, $P_{i,n}$ the cumulative failure probability of the n^{th} line segment, M and L are the total number of towers and line segments in the same corridor, respectively, $\lambda_{i,n}$ the failure rate of the n^{th} line segment, $\lambda_{i,m}$ the failure rate of the m^{th} tower at time t_j, N the number of time steps, and Δt the time step size.

B. SAC environment

An MDP is used to formulate the problem where a system state represents specific system conditions. A transition to another state is due to taking certain actions yielding a reward that can be defined as a function of desired outcome. The components of the formulated MDP are defined below.

1) States

The state set describes the system conditions and the required information to fully observe the system characteristics. The state set is defined as:

$$s_t = \left\{G_i^l, G_i^s, G_i^r, L_n, Cu_n, u_j\right\} \quad \forall n \in \Omega^N,\ \forall i \in \Omega^G\ \forall j \in \Omega^B \quad (16)$$

where G_i^l is the DG location, G_i^s the DG size, G_i^r the DG generation reserve, L_n the real power load, Cu_n the curtailed load, u_j the line status, Ω^N the set of system nodes, Ω^G the set of DGs, and Ω^B the set of lines.

2) Actions

In the proposed problem, a discrete and a continuous action needs to be taken by each agent. The discrete action signifies the location of the DG whereas the continuous action represents the size of the DG. For each agent, the action is represented as follows:

$$\alpha_t^{ci} = \left\{G_i^l, G_i^s\right\} \quad (17)$$

where α_t^{ci} represents the action specifying the size and location of DG for the i^{th} agents and G_i^l and G_i^s are the location and size of the i^{th} agent, respectively,

3) Rewards

A proper reward value, r_t, should be defined to assess the effectiveness of the actions. Each agent is encouraged to reduce the amount of load curtailment and to maintain enough generation reserve during contingencies. Generally, the reward value increases as the amount of load curtailment decreases. Also, the reward value increases as the amount of generation reserve exceeds a specific threshold. The reward r_t for taking a specific action is calculated as:

$$r_t = -C_{cu}\ .\ \sum_{n\in\Omega^N} Cu_n - C_r\ .\ \sum_{i\in\Omega^G} G_i^r \quad (18)$$

where C_{cu} is the cost of load curtailments, Cu_n the load curtailment at bus n, Ω^N the set of all buses, C_r the cost of additional generation reserve, and Ω^G the set of all generators.

To obtain the amount of load curtailment, an AC optimal power flow (OPF) is formulated and solved by setting the sizes and locations DGs equal to the action taken by each agent. The amount of generation reserve is the difference between the

978-1-6654-0823-3/22/$31.00 ©2022 IEEE

sizes of DGs as determined by the agents and the obtained sizes from solving the AC OPF problem after including 25% generation reserve.

C. Training and Execution Algorithms

The power grid is divided into several regions based on the electrical distance between components such that each region is controlled by one agent. Each agent is responsible for determining the location and size of a DG unit to supply loads within its region. To train all agents, a replay buffer is used as follows:

$$\mathcal{D} \leftarrow \left(s_t, o_t^i, \alpha_t^{ci}, \alpha_t^{-ci}, r_t, s_{t+1}, o_{t+1}^i, \alpha_{t+1}^{ci}, \alpha_{t+1}^{-ci}\right) \quad (19)$$

The training and testing/execution steps for the multi-agent framework are summarized in Algorithm 1 and Algorithm 2.

Algorithm 1 - Training of the Multi-agent Framework

for episode = 1 to M **do**
  Create failure scenario using fragility curve
  Reset the environment to default settings
  Solve AC OPF to determine o_i^t and s_t of each agent
  while load curtailed, additional reserve and step $< N$ **do**
    Evaluate actions, α_t^{ci} for agent i using (17)
    Execute actions α_t^{ci} using AC OPF environment (e.g., Pandapower)
    Observe s_{t+1}, r_t, and d to check terminal conditions.
    Store (s_t, o_t^i, α_t^{ci}, α_t^{-ci}, r_t, s_{t+1}, d) in $\mathcal{D}_i$ using (19)
    If s_{t+1} is terminal, reset the environment
    Update weights of the policies using (11)
    Update the Q-function parameters of local and target networks of each agent using (7)
    Update temperature of actor-networks using (12)
    Update target networks weights of each agent using $\bar{Q}_m \leftarrow \tau Q_m + (1-\tau)\bar{Q}$, where, $m \in \{1, 2\}$ and $m \ll 1$
  end while
end for

Algorithm 2 - Testing of the Multi-agent Framework

for episode = 1 to M **do**
  Create failure scenario using fragility curve
  Reset the environment to default settings
  while load curtailed, additional reserve and step $< N$ **do**
    Evaluate actions, α_t^{ci} for using (17)
    Execute actions α_t^{ci} using power flow solver
    Observe s_{t+1}, r_t, and d to validate terminal conditions
  end while
end for

IV. Implementation and Results

The proposed approach is applied on the IEEE 33-bus distribution feeder. Several failure scenarios are created using the hurricane fragility model provided in section III-A. The vulnerable lines are (1-2), (2-3), (5-6), (9-10), (15-16), (21-22), (26-27), and (31-32), as shown in Fig. 1. To create a more severe condition, the connection to the main feeder is disconnected with the result being the system acting as an islanded microgrid. Also, the impact of load variation is considered by scaling the system nominal load using load demand profile obtained from [26]. The power grid is split into 6 regions. An agent is assigned to each region as shown in Table I. Each DG is assumed to have maximum capacity of 2 MW. The MDP is formulated and algorithm 1 is used for training the multi-agent framework.

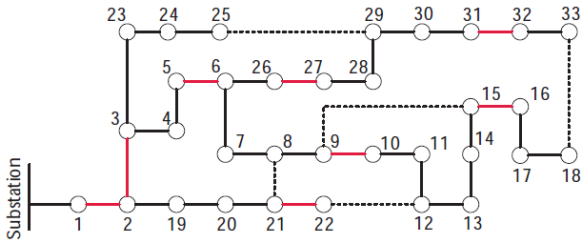

Fig. 1. IEEE 33-bus distribution feeder

TABLE I
Assigned Nodes to Each Agent

Agent	Nodes
A_1	1, 2, 3, 4, 5, 6
A_2	7, 8, 9, 10, 11
A_3	12, 13, 14, 15, 16
A_4	17, 18, 19, 20, 21, 22
A_5	23, 24, 25, 26, 27, 28
A_6	29, 30, 31, 32, 33

The proposed algorithm is implemented for a fixed number of episodes (failure scenarios). A total of 10000 episodes are used for training. The cumulative reward for each episode is plotted as shown in Fig. 2. The learning rate of the agents is improved as more scenarios are simulated. Also, the algorithm explores more situations providing the agents with more experience. Reward values lower than -20 implies significant amount of load curtailments. As reward value approaches zero, the amount of load curtailment is reduced and the capacity of DGs is within permissible limits. Additional training might be required for further improvements.

The trained agents are tested using a set of failure scenarios randomly selected from the training set. Fig. 3 shows the reward obtained for testing data. Agents are able to determine DG sizes and locations that maintain minimal load curtailments; however, in some cases the determined sizes are much larger than the load demand.

V. Conclusion

This paper has proposed a reinforced learning approach to enhance the operational resilience of power grids during hurricanes. The proposed method determines the locations and sizes of DGs to maintain minimal amount of load curtailments. A SAC framework was trained using a formulated MDP and

978-1-6654-0823-3/22/$31.00 ©2022 IEEE

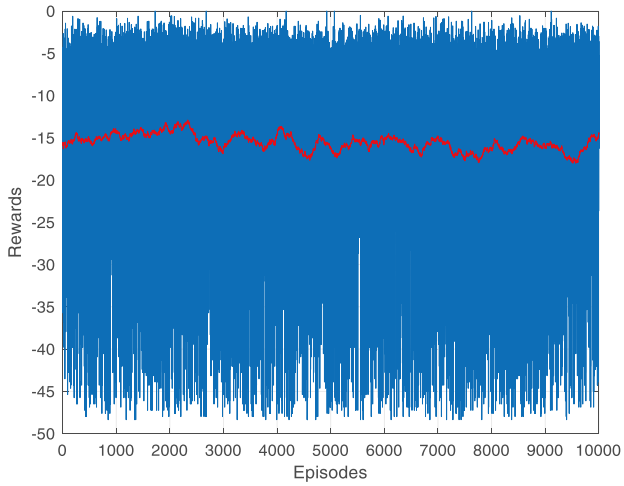

Fig. 2. Reward per Episode for Training set

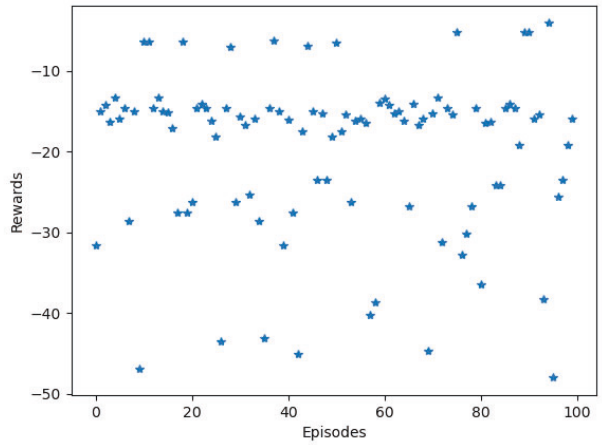

Fig. 3. Reward per Episode for Testing set

hurricane fragility model. The trained algorithm was tested on the IEEE 33-bus distribution feeder. The results showed that the proposed approach could provide proper decisions to maintain reliable operation of an islanded distribution feeder during impacts of hurricane. Feasible sizes and locations of DGs could not be achieved in some cases. The proposed algorithm can be extended to include other resources such as load shedding, reconfiguration, and energy storage for further improvements.

ACKNOWLEDGEMENT

This work was supported by the U.S. National Science Foundation (NSF) under Grant NSF 1847578.

REFERENCES

[1] N. Bhusal, M. Abdelmalak, M. Kamruzzaman, and M. Benidris, "Power system resilience: Current practices, challenges, and future directions," *IEEE Access*, vol. 8, pp. 18 064–18 086, 2020.

[2] R. J. Campbell, "Weather-related power outages and electric system resiliency," Congressional Research Service, Tech. Rep., 2012.

[3] K. Sinan, P. Samuel, and L. Matti, "A summary of the recent weather events and their impact on electricity," *International Review of Electrical Engineering*, vol. 9, no. 4, pp. 821–828, 2014.

[4] W. House, "Economic benefits of increasing electric grid resilience to weather outages," Executive office of the president, Washington, DC, USA, Tech. Rep., Aug 2013.

[5] A. Kavousi-Fard, M. Wang, and W. Su, "Stochastic resilient post-hurricane power system recovery based on mobile emergency resources and reconfigurable networked microgrids," *IEEE Access*, vol. 6, pp. 72 311–72 326, 2018.

[6] A. Gholami, T. Shekari, and S. Grijalva, "Proactive management of microgrids for resiliency enhancement: An adaptive robust approach," *IEEE Trans. on Sust. Energy*, vol. 10, no. 1, pp. 470–480, Jan 2019.

[7] "Severe impact resilience: Considerations and recommendations," NERC, Tech. Rep., May 2012. [Online]. Available: http://www.nerc.com

[8] M. Abdelmalak and M. Benidris, "A Markov decision process to enhance power system operation resilience during hurricanes," in *IEEE Power Energy Society General Meeting (PESGM)*, July 2021, pp. 1–5.

[9] M. Abdelmalak and B. Mohammed, "A Markov decision process to enhance power system operation resilience during wildfires," in *IEEE Industrial Applications Society Annual Meeting*, Vancouver, BC, Canada, October 2021.

[10] M. M. Hosseini and M. Parvania, "Artificial intelligence for resilience enhancement of power distribution systems," *The Electricity Journal*, vol. 34, no. 1, p. 106880, 2021.

[11] J. Xie, I. Alvarez-Fernandez, and W. Sun, "A review of machine learning applications in power system resilience," in *2020 IEEE Power Energy Society General Meeting (PESGM)*, 2020, pp. 1–5.

[12] M. Kamruzzaman, J. Duan, D. Shi, and M. Benidris, "A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5525–5536, 2021.

[13] H. Nie, Y. Chen, Y. Xia, S. Huang, and B. Liu, "Optimizing the post-disaster control of islanded microgrid: A multi-agent deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 153 455–153 469, 2020.

[14] L. Tightiz and H. Yang, "Resilience microgrid as power system integrity protection scheme element with reinforcement learning based management," *IEEE Access*, vol. 9, pp. 83 963–83 975, 2021.

[15] Z.-c. Zhou, Z. Wu, and T. Jin, "Deep reinforcement learning framework for resilience enhancement of distribution systems under extreme weather events," *International Journal of Electrical Power & Energy Systems*, vol. 128, p. 106676, 2021.

[16] M. M. Hosseini and M. Parvania, "Resilient operation of distribution grids using deep reinforcement learning," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2021.

[17] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," *CoRR*, vol. abs/1801.01290, 2018. [Online]. Available: http://arxiv.org/abs/1801.01290

[18] Z. Fan, R. Su, W. Zhang, and Y. Yu, "Hybrid actor-critic reinforcement learning in parameterized action space," *CoRR*, vol. abs/1903.01344, 2019. [Online]. Available: http://arxiv.org/abs/1903.01344

[19] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, "Soft actor-critic algorithms and applications," *CoRR*, vol. abs/1812.05905, 2018.

[20] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," *CoRR*, vol. abs/1802.09477, 2018.

[21] A. Hussain, A. Oulis Rousis, I. Konstantelos, G. Strbac, J. Jeon, and H. Kim, "Impact of uncertainties on resilient operation of microgrids: A data-driven approach," *IEEE Access*, vol. 7, pp. 14 924–14 937, Jan. 2019.

[22] H. Nazaripouya, "Power grid resilience under wildfire: A review on challenges and solutions," in *2020 IEEE Power Energy Society General Meeting (PESGM)*, 2020, pp. 1–5.

[23] M. Abdelmalak and M. Benidris, "Proactive generation redispatch to enhance power system operation resilience during hurricanes," in *2020 52^{nd} North American Power Symposium (NAPS)*, 2021, pp. 1–6.

[24] X. Liu, K. Hou, H. Jia, J. Zhao, L. Mili, X. Jin, and D. Wang, "A planning-oriented resilience assessment framework for transmission systems under typhoon disasters," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5431–5441, 2020.

[25] C. Wang, Y. Hou, F. Qiu, S. Lei, and K. Liu, "Resilience enhancement with sequentially proactive operation strategies," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2847–2857, 2017.

[26] NYISO. Load data. [Online]. Available: https://www.nyiso.com/load-data

978-1-6654-0823-3/22/$31.00 ©2022 IEEE