Network Reconfiguration for Enhanced Operational Resilience using Reinforcement Learning

Michael Abdelmalak[1], Mukesh Gautam[1], Sean Morash[2], Aaron F. Snyder[2],
Eliza Hotchkiss[3], and Mohammed Benidris[1]
[1]Department of Electrical and Biomedical Engineering, University of Nevada-Reno, Reno, NV 89557, USA
[2]EnerNex LLC, Knoxville, TN 37932, USA
[3]National Renewable Energy Laboratory, Golden, CO 80401, USA
Emails: {mabdelmalak, mukesh.gautam}@nevada.unr.edu, {smorash, aaron}@enernex.com,
eliza.hotchkiss@nrel.gov, mbenidris@unr.edu

***Abstract*—This paper proposes a reinforcement learning-based approach for distribution network reconfiguration (DNR) to enhance the resilience of the electric power supply. Resilience enhancements usually require solving large-scale stochastic optimization problems that are computationally expensive and sometimes infeasible. The exceptional performance of reinforcement learning techniques has encouraged their adoption in various power system control studies, specifically resilience-based real-time applications. In this paper, a single agent framework is developed using an Actor-Critic algorithm (ACA) to determine statuses of tie-switches in a distribution feeder impacted by an extreme weather event. The proposed approach provides a fast-acting control algorithm that reconfigures the feeder topology to reduce or even avoid load shedding. The problem is formulated as a discrete Markov decision process in such a way that a system state captures the system topology and its operational characteristics. An action is made to open or close a specific set of tie-switches after which a reward is calculated to evaluate the practicality and advantage of that action. The iterative Markov process is used to train the proposed ACA under diverse failure scenarios and is demonstrated on the 33-node distribution feeder system. Results show the capability of the proposed ACA to determine proper switching action of tie-switches with accuracy exceeding 93%.**

***Index Terms*—Actor critic, Markov decision process, network reconfiguration, reinforcement learning, resilience.**

NOMENCLATURE

A	Node-branch incidence matrix
E_{tn}, E_{tt}	Number of training/testing episodes
e, n	Number of graph edges/nodes
$\mathcal{E}_\mathcal{P}$, $\mathcal{N}_\mathcal{P}$	Set of graph edges/nodes
i, j	Index of distribution line/tie-switches
O	Agent observation
N_l, N_s	Number of distribution lines/tie-switches
R	Reward value
s	State value
t	Index of trial iteration
$V_\psi(O_t)$	policy value function
α	Action by the actor-network
β	Discount factor
ξ	Policy network parameter
$\pi_\xi(\alpha_t\|O_t)$	Unbounded Gaussian policy
ψ	Value function network parameter
θ	Q-value function parameter
$Q_\theta(s_t, \alpha_t)$	Critic policy evaluation function

I. INTRODUCTION

A. Motivation and Background

The frequency and intensity of extreme weather events have increased in recent years, yielding prolonged outages and significant 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 eight million customers to lose power across 15 states in the United States [4]. Fast and efficient restoration of lost loads due to extreme events is one of the most important attributes to achieve resilient operation of power systems and reduce their economic and community impact [5], [6]. Because of the increased vulnerability of power distribution systems to extreme weather events, a proper restoration strategy is required to enhance distribution system resilience. This can be achieved via microgrid formation [7], network reconfiguration [8], and utilization of distributed energy resources (DERs) [9]. However, determining proper restoration decisions in a fast-paced manner is computationally expensive and time consuming, specifically for large-scale systems. Therefore, implementing a restoration resilience enhancement strategy that provides proper decisions considering the system operational constraints has become important.

B. Relevant Literature

Several studies have been conducted to improve the restoration performance of distribution systems against extreme weather events. A spectral clustering algorithm has been employed to determine optimal network partitions under tight potential $N - k$ (i.e., $k > 1$) contingencies [10]. A risk-based defensive islanding approach has been studied in [11] to reduce the impact of cascading failures on transmission systems for enhanced restoration against hurricanes. Resilience-based microgrid formation frameworks have been proposed to enhance the restoration of critical loads in both radial and meshed networks [12]. An evolutionary algorithm approach has been proposed in [13]

978-1-6654-0557-7/22/$31.00 ©2022 IEEE

to improve restoration behavior of lost loads via dispatching tie-switches in distribution feeders. In [14], a heuristic approach integrated with a fuzzy multi-objective function has been proposed to determine the sequence of line energizing for enhanced restoration. A mixed-integer linear programming optimization-based formulation has been used to retain critical loads through microgrid formation after an extreme event [15]. Most of these studies have leveraged analytical and heuristic-based techniques to enhance the restoration of distribution systems. Despite the significant contributions of these methods, their efficacy depends mainly on the accuracy of the system models and degree of approximations. Also, the computational complexity of these methods increases dramatically with the system size imposing scalability challenges.

Reinforcement learning (RL) approaches have been used to provide a fast-acting control algorithm for high-dimension stochastic optimization problems [16]. Several deep reinforcement learning (DRL) methods have been proposed to improve resilience of electric power systems [17]. A soft actor-critic algorithm could potentially improve voltage stability of transmission systems during a hurricane based on dispatching shunt resources [18]. In [19], a DRL-based protection scheme has been used to improve the operational efficiency of microgrids integrated with market participation constraints. An optimal rescheduling strategy has been used to train an RL-based network for improved resilience during hurricanes [20]. RL-based optimal control algorithms have been used to improve the operational performance of microgrids after a disaster [21]. RL-based approaches provide a pathway to overcome some of the challenges of analytical and population-based search methods. In addition, learning-driven models have the capability to apply lessons from experiences during online operations [22]. Also, RL-based methods can be easily integrated into online decision-making process once fully trained and implemented. However, the role of RL in DNR for improved resilience is still under investigation.

C. Contributions and Organization

This paper proposes a DRL-based approach to control tie-switches of distribution circuits to enhance the operational resilience of the power supply during disruptive events. The proposed algorithm is developed leveraging a DNR strategy to reduce/eliminate the amount of load curtailment. A single-agent Actor-Critic Algorithm (ACA) is used to train an RL-based model under multiple line outages in a distribution system. A Markov Decision Process (MDP) is used to formulate the sequential iterative learning process for the agent. An action implies connecting tie-switches to modify the system topology, while a system state provides information about the system operating conditions and availability of system components. A reward function is used to assess the appropriateness of the executed action. A proper action should satisfy the traverse constraint and radiality constraint of the distribution system. The sequential MDP is repeated for numerous failure scenarios until the agent is fully-trained. The trained ACA provides a set of tie-switches to be reconnected for enhanced resilient operation after an extreme event. The proposed algorithm provides a corrective and restorative resilience enhancement strategy that can be adopted for real-time applications. The ACA is demonstrated on the 33-node distribution feeder for validation. The contributions of this paper are: (1) Develop a RL-based model to control tie-switches of distribution power systems; (2) Provide a resilience enhancement strategy leveraging network reconfiguration approach; and (3) Validate the capabilities of ACA to improve system performance during an extreme event.

The remainder of the paper is organized as follows. The mathematical formulation of the ACA is explained in Section II. Section III describes the ACA for the DNR problem. A case study on the 33-node system is used to validate the proposed work in Section IV. Finally, in section V, there are some concluding remarks.

II. Actor Critic Algorithm

RL-based approaches rely mainly on estimating optimal value functions and discovering the optimal policy for a given problem environment. Various methods have been used to estimate the value functions, including dynamic programming and backward induction methods [23]. Reinforcement learning involves a repetitive sequential Markov decision process from samples of states, actions, and rewards. The Markov game consists of an uncertain *environment* where an agent makes an *action* to maximize cumulative *reward*. The *state* representing a specific condition of the environment changes based on the executed action. In some problems where the action space is significantly large, or the problem environment is highly non-linear, temporal difference approaches have been used to overcome these challenges including Q-learning, deep Q-networks, and ACAs [24].

In ACA, a single- or multi-agent framework is formulated as a Markov game where it is required to maximize the discounted returns of the agents. The ACA includes an actor network and a critic network. The former is trained to determine the proper actions, whereas the latter is trained to determine the optimal policy upon which the actor makes proper actions. A policy is defined to be the mapping process from the environment state to the action space. The goal of each agent is to find a policy that maximizes its total rewards. A single agent has one actor network to provide appropriate actions with a policy that can be expressed as follows.

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

In each iteration, the policy is updated to maximize the expected return of an agent in the fundamental ACA model. A policy is evaluated as follows.

$$V_\psi(O_t) = \mathop{\mathbb{E}}_{\alpha_t \sim \pi_\xi}[Q_\theta(s_t, \alpha_t)] \tag{2}$$

$$Q_\theta(s_t, \alpha_t) = r(s_t, \alpha_t) + \beta \mathbb{E}_{s_{t+1} \sim p}[V_\psi(o_{t+1})] \tag{3}$$

The expression provided in (4) is used to minimize the residual squared error of a soft Bellman function to train value functions of the actor network.

$$J_v(\psi) = \mathbb{E}_{s_t}\left[\frac{1}{2}\left(V_\psi(O_t) - Q_\theta(s_t, \alpha_t)\right)^2\right] \tag{4}$$

The gradient of (4) to sample actions from the current policy is determined as follows.

$$\hat{\nabla}_\psi J_v(\psi) = \nabla_\psi V_\psi(O_t)\left[V_\psi(O_t) - Q_\theta(s_t, \alpha_t)\right] \tag{5}$$

To update the Q-parameters of the basic actor, the following expression can be used.

$$J_{Q_\theta}(\theta) = \mathbb{E}_{(s_t,\alpha_t)}\left[\frac{1}{2}\left(Q_\theta(s_t, \alpha_t) - \hat{Q}(s_t, \alpha_t)\right)^2\right] \tag{6}$$

The value of Q-function (6) is optimized as follows:

$$\hat{\nabla}_\theta J_{Q_\theta}(\theta) = \nabla_\theta Q_\theta(s_t, \alpha_t)\left[Q_\theta(s_t, \alpha_t) - \hat{Q}(s_t, \alpha_t)\right] \tag{7}$$

The presented algorithm leverage A2C framework explained in [22]. Detailed explanation of the proposed ACA is provided in our previous work in [18] for further illustration.

III. The proposed ACA-DNR

This section explains the adoption of ACA for resilience enhancement of distribution systems. First, it discusses the graph theoretic representation of a distribution network for topology reconfiguration. Then, it describes the ACA-DNR environment and the algorithm execution procedure.

A. Formulation of DNR

A distribution power system can be represented as a undirected graph $\mathcal{G}_\mathcal{P} = (\mathcal{N}_\mathcal{P}, \mathcal{E}_\mathcal{P})$, where $\mathcal{N}_\mathcal{P}$ is a set of vertices corresponding to buses or nodes in the power system and $\mathcal{E}_\mathcal{P}$ is a set of edges referring to distribution line segments, transformers, sectionalizers, and tie-switches [25]. Changing the status of sectionalizers and tie-switches provides different topologies of a distribution feeder. For enhanced resilience, minimal amount of load curtailment should be achieved. Also, node traversing and radiality constraints should be fulfilled for feasible operation of the distribution system.

1) Traversing Constraint

In the absence of DERs, only the main substation can supply energy to load nodes. There should be at least one path from the source to the load node. In other words, all system nodes should be connected together without the existence of islanded nodes.

2) Radiality Constraint

Radiality requirements should be satisfied in distribution systems to align with the existing protection coordination schemes and voltage regulation fundamentals. A node-branch incidence matrix A can be constructed using (8) for a distribution network, such that $A \in \mathbb{R}^{n\times e}$. The radiality constraint is satisfied if matrix A is a full rank matrix.

$$a_{x,y} = \begin{cases} +1 & \text{if branch } y \text{ starts at node } x \\ -1 & \text{if branch } y \text{ ends at node } x \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

B. ACA environment

This section formulates the DNR problem as an MDP representing the ACA approach. The MDP is a sequential process where a reward value is calculated based on a specific action to change the problem environment from one state to another. The better the action is, the higher the reward will be. The states, the actions, and the rewards are formulated as follows.

1) States:

The state set describes the system conditions and the required information to fully observe the system characteristics. In this paper, the state is represented by a vector of on/off status of network branches, including both distribution lines and tie-switches, as follows.

$$s_i = \begin{cases} 1 & \text{if line is connected} \\ 0 & \text{if line is not connected} \end{cases}, \forall i \in N_l + N_s \tag{9}$$

2) Action:

In the proposed problem, a discrete action represents changing the status of a specific tie-switch. A vector of on/off status of network tie-switches is fed into the problem environment at each iteration. The action vector is formulated as follows.

$$\alpha_j = \begin{cases} 1 & \text{if tie-switch is connected} \\ 0 & \text{if tie-switch is not connected} \end{cases}, \forall j \in N_s \tag{10}$$

3) Reward:

A proper reward value should be defined to assess the effectiveness of the actions. An agent is encouraged to determine the best set of tie-switches to be turned on for a specific failure scenario. A discrete reward function is formulated where a value of -1 is given for each *wrong* action and a value of 10 when reaching a feasible solution.

The total reward at time step t is computed as follows.

$$R = \begin{cases} 10 & \text{if all constraints are satisfied} \\ -1 & \text{if any constraint is violated} \end{cases} \tag{11}$$

C. Training and Execution Algorithms

The proposed ACA agent is trained to determine the set of tie-switches to be connected for improved resilience. The agent is subjected to different failure scenarios from a list of potential failures. For each failure scenario, the agent takes an action and a reward is calculated. The process is repeated till the ACA converges. The training and testing steps for the ACA are summarized in Algorithm 1 and Algorithm 2 below.

IV. Implementation and Results

The proposed approach is applied on the 33-node distribution feeder for validation. The proposed ACA model is formulated to control tie-switches of the distribution feeder for enhanced resilience leveraging DNR approach.

Algorithm 1 - Training of the ACA Framework

Define hyper-parameters of ACA
for episode = 1 to E_{tr} **do**
  Create failure scenario
  Reset the environment to default settings
  while Constraints not fulfilled and step $< N$ **do**
    Generate an action (set of connected tie-switches) using the actor network
    Evaluate the value of the current state using the critic network
    Execute the action on the environment
    Compute the reward value
    Observe the new state
    Reset the environment, if terminal reached.
    Update the weights of the actor network using (5)
    Update the weights of the critic network using (7)
  end while
end for

Algorithm 2 - Testing of the ACA Framework

for episode = 1 to E_{tt} **do**
  Create failure scenario
  Reset the environment to default settings
  Generate an action using the Actor network
  Execute the action on the environment
  Count success if terminal condition is fulfilled
end for

A. System under Study

The 33-node distribution test system is a radial distribution system with 33 nodes, 32 branches, and 5 tie-lines (37 branches) with total load of 3.72 MW [26]. The proposed algorithm is implemented on the original system to validate the effectiveness of the proposed algorithm to adapt to existing system characteristics. A list of tie-switches and vulnerable lines is summarized in Table I and shown in Fig. 1.

B. Case Studies

The proposed ACA model is trained for three cases based on the number of failed lines as follows: (a) Case C_1: single line failure, (b) Case C_2: two-line failures, and (c) Case C_3: randomly selected failures between one and four lines. The training is performed for 30,000 episodes with a maximum of ten iterations per episode. Also, a stopping criterion is adopted to terminate the training process if the average reward value exceeds a specific threshold for 100 consecutive episodes. This is due to the high step impact from one episode to another

TABLE I
LIST OF VULNERABLE LINES AND TIE-SWITCHES

Tie-Switch	Connecting nodes	Vulnerable lines	Connecting nodes
SW_1	21-8	L_1	3-23
SW_2	9-15	L_2	5-6
SW_3	12-22	L_3	21-22
SW_4	18-33	L_4	10-11
SW_5	25-29	L_5	29-30

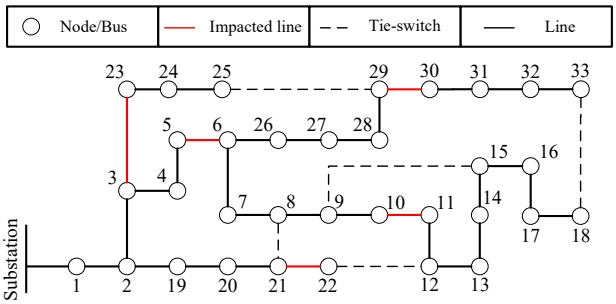

Fig. 1. Schematic diagram of the 33-node distribution feeder

TABLE II
HYPER-PARAMETER SETTINGS OF THE ACA FOR 33-NODE SYSTEM

Hyper-parameters	Values
Number of hidden layers	3
No. of neurons in hidden layers	100, 100, 100
Learning rate	10^{-3}
Reward discount factor	0.99
Activation function of output layer	Sigmoid
Activation function of hidden layers	ReLU
Optimizer	Adam

causing potential instability in the ACA networks [18]. The hyper-parameter settings of the actor and critic networks of the proposed framework are shown in Table II.

1) Training: In each training episode, the system is initialized with a random state representing the status of all the system lines. A set of failed lines is selected randomly from the set of vulnerable lines. An action is generated using the actor network and a corresponding value is computed using the critic network for the given system state. A reward value is calculated based on the obtained new system state. The process is repeated for all aforementioned cases. For evaluation, the running mean of the episodic rewards and the number of iterations per episode are calculated using a window of 100 episodes. The proposed DRL-based approach takes approximately 4 milliseconds to execute on a PC with a 64-bit Intel i7 core processor running at 3.15 GHz, 16 GB RAM, and Windows OS.

Fig. 2 and Fig. 3 show the running mean and number of iterations per episode for cases C_1 and C_2. The average reward value increases as the number of training episodes

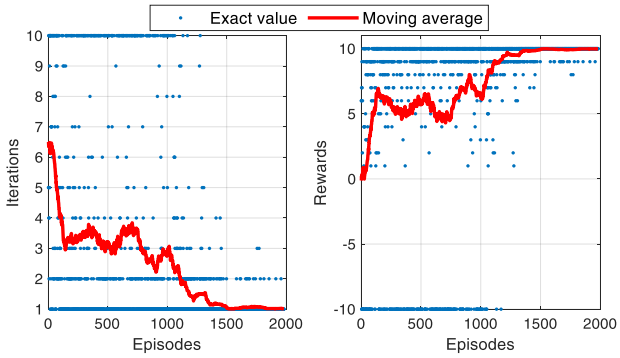

Fig. 2. Reward and iterations per episode for C_1

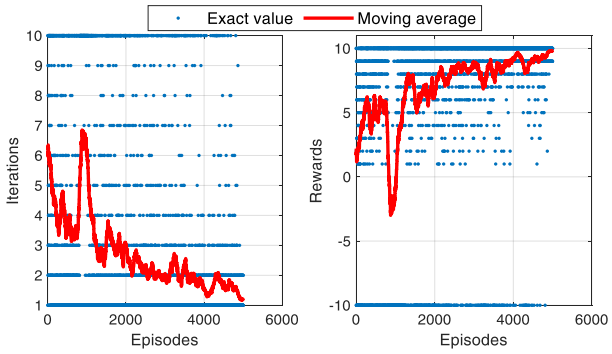

Fig. 3. Reward and iterations per episode for C_2

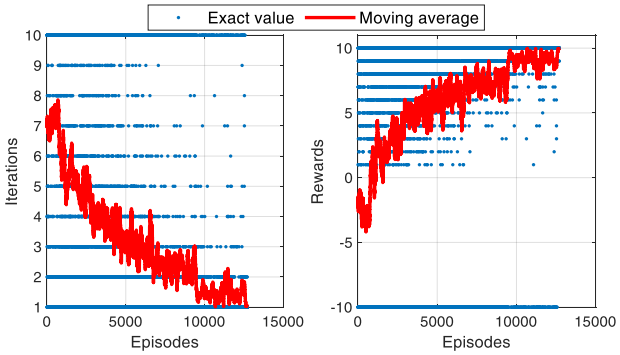

Fig. 5. Reward and iterations per episode for C_3

increases, as anticipated. The average reward value reaches the saturation level in less than 1,500 episodes in C_1 yielding the effectiveness of the proposed algorithm to turn on a proper tie-switch to maintain system constraints. In C_2, the average reward reaches saturation around 5,000 episodes. This is due to the existence of more than one possible action for a specific failure scenario. For instance, failure of L_4 (nodes 10-11) and L_5 (nodes 29-30) can be mitigated by turning on either SW_4 and SW_2 (nodes 18-33 and 9-15) or SW_4 and SW_3 (nodes 18-33 and 12-22). On the other hand, the number of iterations per episode decreases as the ACA networks are trained. As the average value of iterations per episode reaches one, the trained ACA is capable of determining the set of tie-switches that maintain radiality constraints and eliminate the amount of load curtailments within one decision iteration.

The training performance of the ACA might change due to the random initialization of the weights of the NN models. Fig. 4 shows the mean of rewards and number of iterations per episode for case C_1 for 50 different runs. It is noticeable that the proposed ACA have almost the same performance and converges after 1,700 episodes. This implies the high robustness level of the ACA.

Fig. 5 shows the running mean and number of iterations per episode for cases C_3. The average reward converges in around 12,000 episodes. The proposed ACA has the capability to learn and make proper decisions as more training episodes are executed. In C_3, the average reward converges in a much slower rate due to the high variability in the environment behavior. In other words, for a specific failure scenario, more than one set of tie-switches is considered a feasible solution. Also, the random failure scenario generation creates further challenges to train the ACA networks.

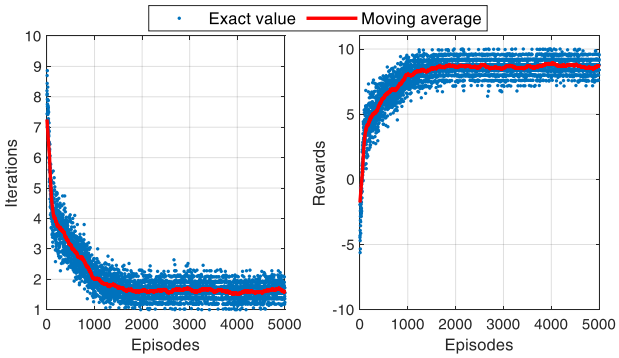

Fig. 4. Reward and iterations per episode for C_1 for 50 different runs

2) Testing: To validate the efficiency of the trained models, a total of 1000 failure scenarios are tested for each case. For each episode, the model is required to provide a feasible set of tie-switches to reconfigure the distribution feeder for enhanced resilience. A successful decision is counted if the provided decision is a proper solution. The success rates are as follows: 99.7% for case C_1, 96% for case C_2, and 93.5% for case C_3.

The trained ACA models are capable of providing a proper reconfiguration of the 33-node feeder with relatively high success rate. Though the efficiency rate can be improved through various modifications of the ACA networks and hyper-parameters tuning procedure, this paper focuses on the capability of the proposed ACA to reconfigure a given distribution system under a specific failure scenario. It is worth noting that all trained ACA models are able to achieve 100% accuracy when two iterations of decisions are allowed. In other words, if the maximum number of iterations per episode is two, a 100% success rate is achieved.

3) Validation: The ACA is trained to determine the set of tie-switches to avoid load shedding. In this case, a failure scenario is provided to visualize the impact on network reconfiguration using the ACA model. Lines L_2 and L_4 are selected to fail resulting in two islands as shown in Fig. 6.

The trained ACA provides two possible network reconfigurations, as shown in Fig. 7 and Fig. 8. Both solutions satisfy the traversing constraint—no islands, and radiality constraint—no circulating loops. In Fig. 7, both SW_3 and SW_4 are connected, whereas switches SW_3 and SW_5 are connected in Fig. 8. Though other possible feasible

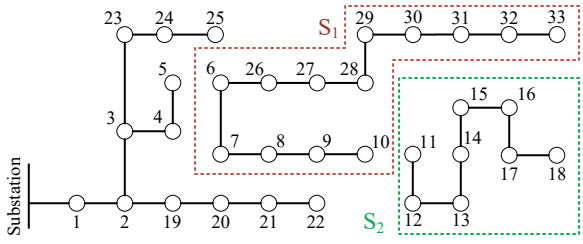

Fig. 6. Failure of L_3 and L_4

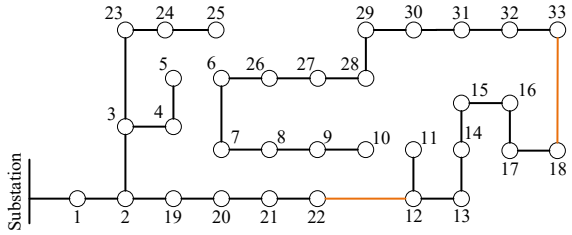

Fig. 7. First possible network reconfiguration due to failure of L_3 and L_4

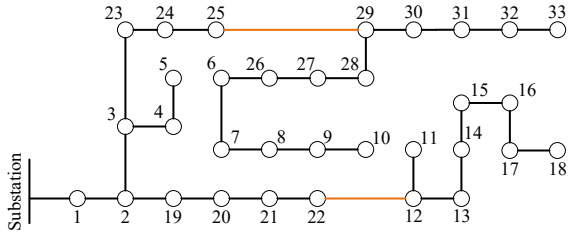

Fig. 8. Second possible network reconfiguration due to failure of L_3 and L_4

reconfigurations might exist, the ACA selects the decision based on their corresponding probability of success. In other words, connecting SW_1 and SW_3 will result in feasible reconfiguration solutions. However, this decision is associated with less probability value within the trained ACA model.

V. Conclusion

This paper has proposed a distribution network reconfiguration approach to control tie-switches of distribution systems for enhanced operational resilience. The proposed method leverages ACA to determine set of tie-switches to be connected due to multiple line outages. An MDP was formulated to train a single agent actor-critic model. The process was repeated for diverse failure scenarios to train the ACA networks. The proposed methodology was tested on the 33-node distribution feeder. The results showed the effectiveness of the proposed ACA to determine the set of tie-switches that allow feasible network reconfiguration maintaining traverse and radiality constraints. The trained ACA was tested against single, double, and multiple failure scenarios and showed accuracy of almost 97%. The proposed algorithm provides the system operators with a fast-acting algorithm to restore curtailed loads in distribution networks after an extreme event. In the future, the characteristics of power systems including loads, voltages, and currents will be considered as well as scalability to large-scale systems.

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] Summary of january 2022 winter storm. [Online]. Available: https://www.weather.gov/akq/Jan22-2022Snow

[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] B. Chen, J. Wang, X. Lu, C. Chen, and S. Zhao, "Networked microgrids for grid resilience, robustness, and efficiency: A review," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 18–32, 2021.

[8] J. Liu, Y. Yu, and C. Qin, "Unified two-stage reconfiguration method for resilience enhancement of distribution systems," *IET Generation, Transmission & Distribution*, vol. 13, no. 9, pp. 1734–1745, 2019.

[9] J. Wu, X. Chen, S. Badakhshan, J. Zhang, and P. Wang, "Spectral graph clustering for intentional islanding operations in resilient hybrid energy systems," *arXiv preprint arXiv:2203.06579*, 2022.

[10] R. Rocchetta, "Enhancing the resilience of critical infrastructures: Statistical analysis of power grid spectral clustering and post-contingency vulnerability metrics," *Renewable and Sustainable Energy Reviews*, vol. 159, p. 112185, 2022.

[11] M. Panteli, D. N. Trakas, P. Mancarella, and N. D. Hatziargyriou, "Boosting the power grid resilience to extreme weather events using defensive islanding," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2913–2922, 2016.

[12] K. S. A. Sedzro, A. J. Lamadrid, and L. F. Zuluaga, "Allocation of resources using a microgrid formation approach for resilient electric grids," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2633–2643, 2018.

[13] L. T. Marques, A. C. B. Delbem, and J. B. A. London, "Service restoration with prioritization of customers and switches and determination of switching sequence," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2359–2370, 2018.

[14] N. Xia, J. Deng, T. Zheng, H. Zhang, J. Wang, S. Peng, and L. Cheng, "Fuzzy logic based network reconfiguration strategy during power system restoration," *IEEE Systems Journal*, pp. 1–9, 2021.

[15] C. Chen, J. Wang, F. Qiu, and D. Zhao, "Resilient distribution system by microgrids formation after natural disasters," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 958–966, 2016.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] A. Zai and B. Brown, *Deep reinforcement learning in action*. Manning Publications, 2020.

[23] M. L. Puterman, "Markov decision processes," *Handbooks in operations research and management science*, vol. 2, pp. 331–434, 1990.

[24] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[25] B. Ti, G. Li, M. Zhou, and J. Wang, "Resilience assessment and improvement for cyber-physical power systems under typhoon disasters," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 783–794, 2022.

[26] Distribution System Analysis Subcommittee, "1992 test feeder cases," IEEE, PES, Tech. Rep., 1992. [Online]. Available: http://sites.ieee.org/pestestfeeders/resources/