



21, rue d'Artois, F-75008 PARIS

<http://www.cigre.org>

**CIGRE US National Committee
2021 Grid of the Future Symposium**

Reinforced Learning Approach Based on Topology Reconfiguration to Enhance Operational Resilience against Hurricanes

**M. ABDELMALAK¹, S. MORASH², M. BENIDRIS¹, A.F. SNYDER²
University of Nevada Reno¹, EnerNex LLC²
USA**

SUMMARY

This paper proposes a reinforced learning approach to control distribution circuit switches to enhance the operational resilience of power systems against extreme weather events. The undergoing developments in machine learning techniques and their exceptional performance has encouraged their adoption in various electric power system applications, specifically to grid modernization and resilience enhancements. Various resilience enhancement strategies rely on solving mathematically involved optimization problems that are not suitable to real-time applications. In real-time operations during extreme events, a set of sequential corrective decisions should be made fast enough to avoid further cascading behavior. The developed data-driven framework is trained using multiple line outages based on various hurricane simulations. Proper fragility curves integrated with spatiotemporal failure models are utilized to assess system characteristics and determine potential component failures. The proposed algorithm integrates a fast-acting machine learning approach to enhance the operational performance of distribution grids during rapidly growing extreme weather events.

KEYWORDS

Distribution system, extreme weather events, reinforced learning, resilience.

I. Introduction

The frequency and intensity of extreme weather events have increased dramatically in the recent decade yielding prolonged outages and noticeable economic losses [1]. In 2008, 200 million people in China experienced a severe ice storm resulting in direct losses exceeding 2 billion US dollars [2]. Superstorm Sandy of October 2012 caused over 8 million customers to lose power across 15 states in the United States [3]. Various resilience enhancement strategies have been developed to reduce, mitigate, and prevent the catastrophic impacts of extreme weather events on overall system performance. Many of the developed strategies highlight the importance of Machine Learning (ML) for a fast and accurate resilience-based solution.

Resilience enhancement strategies vary based on the study period: prior to, during, and after an event [4]. Operational resilience enhancements aim to provide an immediate solution—a single action or a set of sequential actions—through utilization of available assets for improved performance during adverse events. Authors of [5] have developed a mixed-integer linear programming method to enhance power supply resilience via generator re-dispatching, network topology reconfiguration, and load shedding. Moreover, several preventive action-based strategies have been proposed to enhance power system resilience, such as a multi-sensor prediction-based wide-area monitoring and control [6], a linear-programming-based optimal siting and sizing of energy storage devices [7], a Monte-Carlo simulation (MC)-based proactive unit commitment framework [8], and an MC-based crew preparation and network reconfiguration technique [9]. Although the aforementioned strategies have been demonstrated to enhance the resilience of power systems, they do not fully explore the potential of Reinforced Learning (RL) algorithms to resilience assessment and enhancement.

Due to the stochastic behavior and spatiotemporal characteristics of extreme weather events, fast-acting control algorithms are required to determine optimal decisions to maintain an acceptable performance level considering system constraints [1]. In [10], a predictive statistical ML algorithm was developed to evaluate the resilience of power systems in terms of the number of outages, outage duration, and the number of interrupted customers. The training data is obtained through various hurricane simulations based on weather characteristics, geographical location, and network topology. Authors of [11] have proposed a data-driven RL-based algorithm for locating shunt compensators to enhance the voltage regulation of distribution systems during windstorms. RL-based methods do not directly require handling numerous system parameters once trained to provide an accurate and reliable decision in short time. In other words, once the RL-algorithm is trained, it can provide accurate solution based on a few input features.

This paper proposes an RL-based approach to control distribution circuit switches for enhancing the operational resilience of power systems during hurricanes. The proposed algorithm is developed based on minimal load curtailment, and thereby better restoration and correction performance is maintained in case of single or multiple outages. The fundamental Soft Actor Critic (SAC) algorithm is used to develop a single agent framework to control distribution circuit switches. In the proposed method, a state vector (observation) is used to represent the network topology including the status of distribution switches as well as the locations of slack buses. An agent is trained to take actions which are either turn a circuit switch on/off or change the type of specific bus into a slack bus. As a result of taking an action, the whole distribution feeder might split into smaller microgrids where the overall load curtailments should be kept minimal. The SAC algorithm is trained using various failure scenarios which are generated through hurricane spatiotemporal propagation model and fragility curve model. The trained algorithm is then used to provide proactively corrective

control actions when a power system is impacted by a hurricane. The proposed algorithm is tested on a modified version of the IEEE 33-node distribution test feeder [12].

Adopting an RL-based framework enhances the situational awareness. System operators do not need to understand every piece of information on how the system operates and the impact of each decision on the performance of each system component, specifically during fast evolving events. The proposed algorithm will be able to assess the current system state and determine a set of optimal decisions considering current and future system variations. This work reflects the reality of many distribution system operators: they just need to know the proper actions in severe situations.

The rest of the paper is organized as follows. Section II describes the proposed algorithm. Section III illustrates the implementation procedure. Section IV shows the demonstration of the proposed solution. Section V provides concluding remarks.

II. Soft Actor-Critic learning approach

Reinforced learning relies mainly on estimating optimal value functions and discovering the optimal policy. In well-known environments and feasibly solvable problems, dynamic programming and Markov decision process (MDP) can be utilized to calculate the value functions of each state and hence, optimal policies can be derived [13]. Reinforced learning involves a repetitive episodic approach to estimate value functions from sample sequence of states, actions, and rewards. Temporal difference learning methods can learn directly from previous experience data without complete awareness of the problem's environment. In some problems, the state space and action space are significantly large yielding infeasibility of MDP. Various approaches have been proposed to overcome such challenges including Q-learning, Deep Q Networks, and Actor Critic algorithms.

In the Soft Actor Critic algorithm, a single or multi-agent framework is utilized to formulate a Markov game [14], where it is required to maximize the discounted returns. An actor network is used to provide proper actions whereas a critic network learns the optimal policy for the actor network to take decisions. A state (observation) represents a specific condition of the environment. An agent has its individual policy, which is a mapping process from the observation to an action. When each agent takes its individual action, the environment changes as a result of the joint action according to the state of transition model. Each agent obtains rewards as a function of the state and joint actions, and receives a private observation condition on the observation model. The goal of each agent is to find a policy that maximizes its expected discounted return.

Each agent in the proposed multi-agent framework has one actor network to provide actions. The policy of the actor network to provide actions is expressed as follows.

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

where i represents the i^{th} agent of the multi-agent framework. In the context of our problem statement, a system operator is an "agent". O_t^i is the observation vector of the i^{th} agent at time t . The observation vector, representing the system topology, is the result of the agent decisions, so in this work that means the impact on network topology. α_t^{ci} is the provided action by the actor-network of the i^{th} agent. Within the context of this work, this is the action that an operator takes resulting in a specific impact (or observation vector) on network configuration. ξ^{ci} is the parameter for the policy of the i^{th} agent. In other words, the ξ^{ci} parameter is tuned through training of the model and is used to capture the optimal policy

associated with each agent. Finally, when these factors are put together we get $\pi_{\xi ci}(\alpha_t^{ci} | O_t^i)$, which is a policy of the i^{th} agent to select optimal actions.

The policy of the actor network is updated in each iteration to maximize the expected return and entropy. Since the main goal is to determine the optimal policy for each agent, a value function is used to measure the value for each agent's policy. The objective function to be minimized is represented by the soft Bellman residual, which expresses the squared error between the optimal value and the output value. The gradient of the objective error function using an unbiased estimator is determined and used to reflect the performance of training agents. The policy is updated in each iteration to maximize the rewards for each agent. Various approaches might be useful to track the policy update as provided in [14].

III. Implementation Procedure

This section describes the proposed RL-based approach to enhance the resilience of distribution systems through a circuit re-switching strategy. First, it describes the impacts of hurricane propagation on power grid components. Then, it explains the SAC algorithm to formulate and train agent for proper actions.

A. Hurricane Spatiotemporal Impacts

The performance and the status of each system component is directly impacted by the propagation behavior and spatiotemporal characteristics of extreme weather events. Fragility curves are usually used to model the probabilistic behavior of component failure and identify potentially impacted components. Due to the temporal and geographical progression of a hurricane, various components in the system can be impacted at sequential time intervals [15]. Hence, the power grid may reside in different states at each time instant. Based on the available components, a state representing a unique system topology can be identified. The set of impacted components at t includes the current and previously impacted components resulting in 2^{N_t} states. As the system size increases, the list of potential failure components increases dramatically resulting in very large state space. The spatiotemporal characteristics of hurricanes are governed by key parameters that identify their uncertainties. Such parameters are either weather-related (wind speed) or geographical-related (landing site). During a hurricane, a component might fail because of high wind speed. Various models have been proposed to model the propagation behavior of hurricanes. In this paper, a hurricane model is adopted from [16] to calculate the wind speed on transmission lines.

B. Fragility Model

As a hurricane propagates through system components, various transmission corridors are impacted. A transmission corridor is defined to be the set of transmission lines and towers that connect between two terminal components. Since transmission corridors are usually very long, it is divided into smaller segments such as one segment comprises two corresponding transmission towers and the part of transmission line connecting between them. Failure of one segment will result in the failure of the whole corridor since they are connected in series. Consequently, the equivalent failure probability of a specific transmission corridor can be evaluated as a series combination of components. The fragility model quantifies the failure probability of each component in terms of weather parameters from both temporal and spatial perspectives. In this work, a spatiotemporal fragility model from [16] has been integrated to calculate the cumulative failure probability of each transmission corridor during the hurricane duration.

C. SAC Algorithm

In the proposed algorithm, a single agent framework is utilized to train SAC networks for network reconfiguration in distribution system. A state represents a specific system topology based on the available transmission lines and slack buses. For a specific state, operational constraints might not be fully satisfied such as the absence of a slack bus for each microgrid or lost connectivity to load spots. Even fulfilling all system constraints might lead to existence of load curtailments due to insufficient generation supply. When the agent takes an action, the environment changes according to the state of transition model. For each action, a specific reward is obtained as a function of the system state and a new system state is observed. The goal of the agent is to find a policy that maximizes its expected return. Algorithm I describes the detailed SAC implementation procedure.

Algorithm I. Detailed SAC algorithm for network reconfiguration

```

Initialize system characteristics
for Episode = 1 to M do
    Create failure scenario using fragility model
    Initialize starting state (switches status and slack buses)
    While non-terminal state and step < N do
        Inject state into Actor and Critic networks
        Execute an action using DC optimal power flow
        Observe and store new state, reward, and state type
        If state type is terminal, reset the environment
        Update weights of both Actor and Critic networks
    
```

IV. Case Studies

The proposed approach is applied on a modified IEEE 33-node distribution feeder system. The proposed SAC is formulated using PandaPower package integrated with python.

A. System under Study

Various DERs are connected to arbitrarily chosen locations as shown in Figure 1. Although the location of DERs plays a vital role in the overall resilience performance, the scope of this work is to assess the network reconfiguration approach based on predefined system component and characteristics. In this work, the DERs are assumed to be identical with maximum power capacity of 1 MW each.

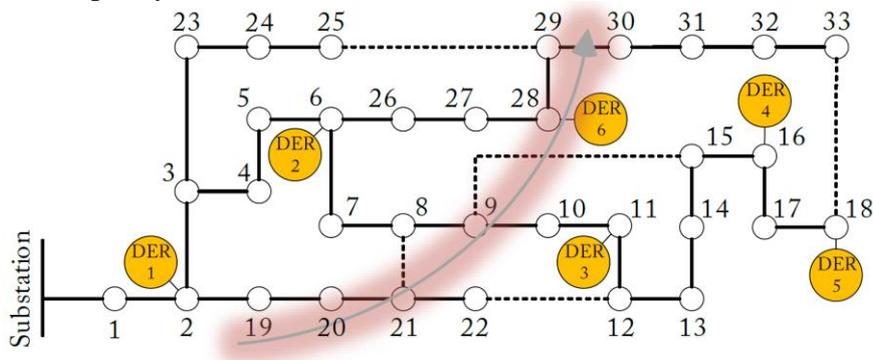


Figure 1. Modified IEEE 33-node feeder with Hurricane propagation

B. Hurricane Scenarios

Various hurricanes are simulated using the provided spatiotemporal in section C.1. Since the behavior of weather parameters vary based on geographical locations, statistical analysis is conducted on the North-eastern region of USA to determine the proper probability distribution function (PDF) for each parameter as summarized in Table I. Although these parameters might have different distribution functions, the main scope of this work is the resilience enhancement rather than the statistical behavior of such parameters. Each failure scenario is injected in the fragility model to determine the list of potential failure components. Figure 1 shows the propagation behavior of a single hurricane event across the modified IEEE 33-node distribution feeder.

Table 1. Parameters of Distributions for Hurricane Events

Key parameter	PDF type	Parameters
Wind speed	Lognormal	$\mu = 2.668 \text{ m/s}, \sigma = 0.5185$
Wind direction	Binormal	$\mu_1 = -73.3, \mu_2 = -7.2,$ $\sigma_1 = 22.6, \sigma_2 = 70.35, \alpha = 0.5$
Central pressure	Uniform	$H_p \in [1.5, 3] \text{ hPa}$
Translational speed	Uniform	$V_T \in [0, 15] \text{ m/s.}$

C. Simulation Results

Several test cases are considered to assess the role of network reconfiguration on resilience enhancements. The simulated failure scenarios are injected into the RL algorithm to train the SAC networks. In this paper, the same network configuration has been assumed for both Actor and Critic networks. The network consists of one single hidden layer of 32 neurons and an ReLu activation function. For the Actor network, the output layer has 44 outputs—total number of actions—and softmax activation function, whereas in the critic network, the output layer has single output representing the Q-value.

To show the capability of the SAC algorithm to learn from failure scenarios, the algorithm is run for 1000 episodes with maximum of 200 step per episode. Figure 2 shows the total number of steps taken at each episode. At earlier stages, the SAC algorithm usually takes less steps compared to later episodes due to reaching terminal state very quickly. The SAC algorithm was not efficiently able to recognize difference between terminal and non-terminal states at earlier stages. Due to randomness of failure scenarios, the SAC learns to capture the features of the system states.

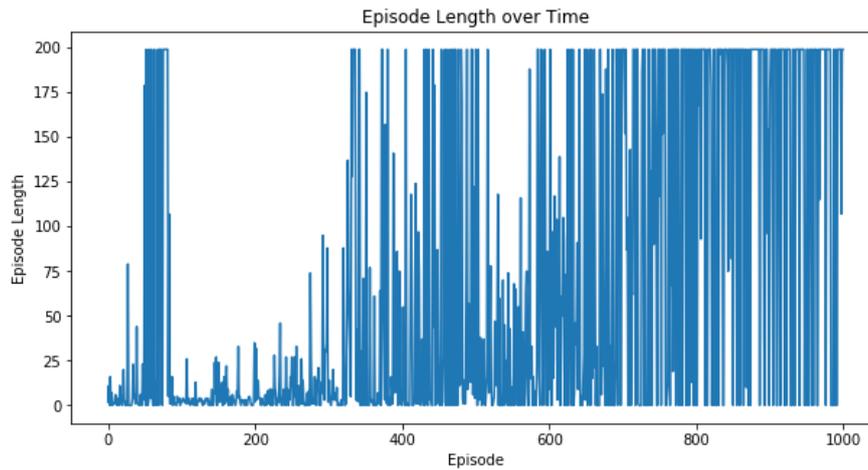


Figure 2. Number of steps per Episode

In Figure 3, the SAC algorithm is able to maximize its reward as it learns. After almost 700 episodes, the SAC networks are able to avoid taking actions that reactivate failed components resulting in higher overall rewards. After around 850 episodes, the SAC algorithm has learned enough experience to capture the features of the system state resulting in rewards that are very close to zero.

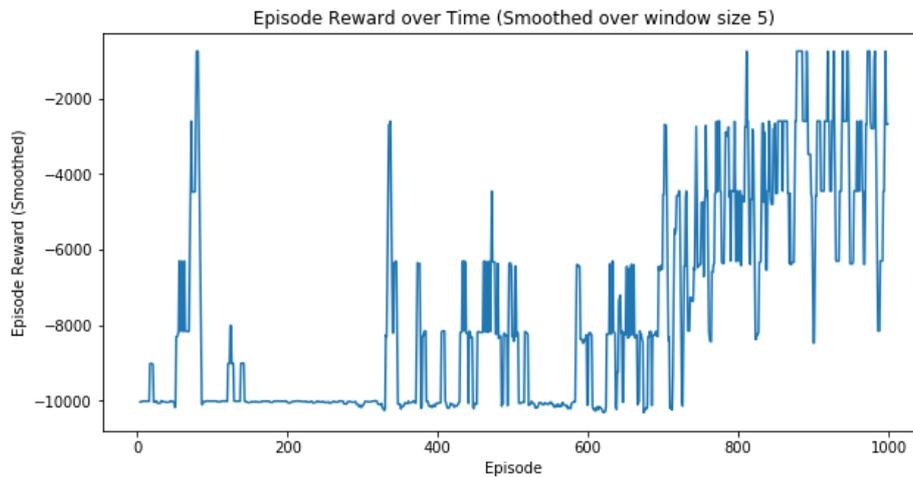


Figure 3. Reward per Episode

Figure 4 shows relationship between the episode number and total number of time steps taken. It shows low exploratory behavior of the SAC algorithm at earlier episodes since the number of steps are not increasing significantly. As SAC algorithm improves its learning capabilities, a constant behavior is noticed between episode number and steps taken showing its capability to balance between exploratory and exploitation actions.

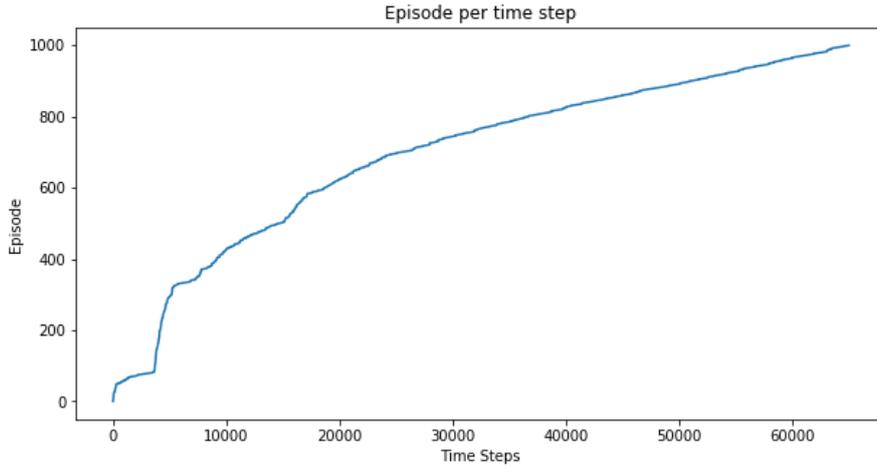


Figure 4. Relationship between Episodes and Steps taken

The simulation is run for 2500 episodes and the obtained results are shown in Figure 5. It is obvious that as the larger number of episodes are simulated, the SAC algorithm attains better learning behavior. Due to the large number of possible actions to be taken, 200 steps limit per episode are not properly sufficient to further improvements. Various learning parameters as well as SAC tuning parameters can be adjusted to achieve higher efficiency.

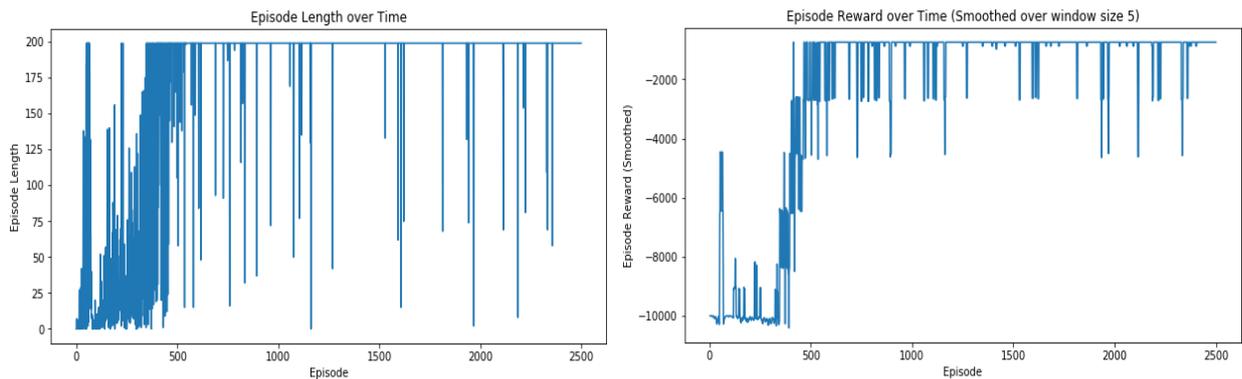


Figure 5. Results for 2500 Episodes

V. Conclusion

This paper has proposed an RL-based resilience enhancement approach to improve the resilience of electric distribution systems against hurricanes. The proposed method trains the SAC network to determine the best set of actions to reconfigure the network feeder into a new set of networks that maintain operational constraints and achieve minimal load curtailments. A hurricane model is used to simulate numerous scenarios based on predefined weather parameters that are governed by calculated PDFs. Fragilities of system components to hurricanes were calculated using a spatiotemporal fragility model against wind speeds. The proposed algorithm was demonstrated on a modified version of the IEEE 33-node distribution feeder. The results showed the effectiveness of the SAC to learn from failure scenarios and determine a new network topology based on sequential set of actions. In the future, the proposed algorithm will be tested on larger systems to validate its scalability capabilities. Also, further development on SAC to encounter continuous decision environment will be studied.

BIBLIOGRAPHY

- [1] N. Bhusal, M. Abdelmalak, M. Kamruzzaman, and M. Benidris, "Power System Resilience: Current Practices, Challenges, and Future Direction," *IEEE Access*, pp. 18064 - 18086, 2020.
- [2] 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, pp. 821 - 828, 2014.
- [3] House, W., "Economic Benefits of Increasing Electric Grid Resilience to Weather Outage," Washington, DC, 2013.
- [4] NERC, "Severe Impact Resilience: Considerations and Recommendations," North American Electric Reliability Corporation, 2012.
- [5] G. Huang, J. Wang, C. Chen, J. Qi, and C. Guo, "Integration of Preventive and Emergency Responses for Power Grid Resilience Enhancement," *IEEE Trans. on Power Systems*, vol. 32, no. 6, pp. 4451 - 4463, 2017.
- [6] Al-Durra, A. S. Musleh, H. M. Khalid, and S. M. Muyeen , "A Prediction Algorithm to Enhance Grid Resilience Toward Cyber Attacks in WAMCS Applications," *IEEE Systems Journal*, vol. 13, no. 1, pp. 710 - 719, 2019.
- [7] Dehghanian, M. Nazemi, M. Moeini-Aghaie, and M. Fotuhi-Firuzabad, "Energy Storage Planning for Enhanced Resilience of Power Distribution Networks Against Earthquakes," *IEEE Trans. on Sustainable Energy*, vol. 11, no. 2, pp. 795 - 806, 2020.
- [8] Y. Wang and, L. Huang, M. Shahidehpour, L. L. Lai, H. Yuan, and F. Y. Xu, "Resilience-Constrained Hourly Unit Commitment in Electricity Grids," *IEEE Trans. on Power Systems*, vol. 33, no. 5, pp. 5604 - 5614, 2018.
- [9] B. Taheri, A. Safdarian, M. Moeini-Aghaie, and M. Lehtonen, "Enhancing Resilience Level of Power Distribution Systems Using Proactive Operational Actions," *IEEE Access*, vol. 7, no. 1, pp. 137378 - 137389, 2019.
- [10] R. Nateghi, "Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems," *IEEE Access*, vol. 6, pp. 13478 - 13489, 2018.
- [11] MD Kamruzzaman, M. Abdelmalak, and M. Benidris, "A Data-driven Shunt Dispatch Approach to Enhance Power System Resilience against Windstorms," in *IEEE Power System and Energy General Meeting (PESGM)*, 2021.
- [12] D. S. A. Subcommittee, "2010 Test Feeder Cases," IEEE, PES, 2010. [Online]. Available: <http://sites.ieee.org/pestestfeeders/resources/>.
- [13] Richard Sutton, and Andrew Barto, Reinforcement learning, MIT press, 2018.
- [14] S. Wang, J. Duan, D. Shi, C. Xu, H. Li, R. Diao, and Z. Wang, "A Data-driven Multi-agent Autonomous Voltage Control Framework Using Deep Reinforcement Learning," *IEEE Trans. on Power Systems*, pp. 1 - 11, 2020.
- [15] M. Abdelmalak and M. Benidris, "A Markov Decision Process to Enhance Power System Operation Resilience during Hurricanes," in *IEEE PES GM*, 2021.
- [16] 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 Transc. on Smart Grid*, vol. 11, no. 6, pp. 5431 - 5442, 2020.