Deep Reinforcement Learning for Distribution System Cyber Attack Defense with DERs

Alaa Selim, *Student Member, IEEE*, Junbo Zhao, *Senior Member, IEEE*, Fei Ding, *Senior Member, IEEE*, Fei Miao, *Member, IEEE*, Sung-Yeul Park, *Member, IEEE*

2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) | 978-1-6654-5355-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISGT51731.2023.10066375

***Abstract*—The use of smart inverter capabilities of distributed energy resources (DERs) enhances the grid reliability but in the meanwhile exhibits more vulnerabilities to cyber-attacks. This paper proposes a deep reinforcement learning (DRL)-based defense approach. The defense problem is reformulated as a Markov decision making process to control DERs and minimizing load shedding to address the voltage violations caused by cyber-attacks. The original soft actor-critic (SAC) method for continuous actions has been extended to handle discrete and continuous actions for controlling DERs' setpoints and load-shedding scenarios. Numerical comparison results with other control approaches, such as Volt-VAR and Volt-Watt on the modified IEEE 33-node, show that the proposed method can achieve better voltage regulation and have less power losses in the presence of cyber-attacks.**

***Index Terms*—Cyber attack, Active distribution systems, Renewable generation, Deep reinforcement learning.**

NOMENCLATURE

Constants

γ_{loss} Weight constant for power losses
γ_{ES} Weight constant for energy storage
γ_v Weight constant for voltage violations
$p_t^{losses}, q_t^{losses}$ Total active/reactive power losses per time step
p_t^{ES} Total active power dispatched by energy storage per time step
p_t^{pv}, q_t^{pv} Total active/reactive power dispatched by solar PV per time step
$p_t^{DER,K}, q_t^{DER,K}$ Total active/reactive power supplied by K distributed energy resources (DER) per time step
p_t^{grid}, q_t^{grid} Total active/reactive power supplied by the grid per time step
$p_t^{unc.}, q_t^{unc.}$ Uncertainties for the active/reactive power generation per time step
p_t^{ch} Total active power charged to the energy storage per time step
p_t^{dis} Total active power discharged by the energy storage per time step
p_t^d, q_t^d Total load active/reactive power per time step
$p_t^{d,sh}, q^{d,sh}$ Total active/reactive load shedding per time step
E_t^{ES} Remaining energy of energy storage at each time step
α, β, ρ, μ penalty and reward constants for the reward function
SW_t Tie switching status at each time step t
i node index
N Total number nodes
T Time period considered
$S_t^{ES}, S_{min}, S_{max}$ State of charge boundaries of energy storage system
v_i^t Voltage of ith node per time step

Abbreviations

DER Distributed Energy Resource
ES Energy Storage
PV Solar Photovoltaic
DRL Deep Reinforcement Learning
SAC Soft Actor Critic
VV Volt-VAR
MPC Model Predictive Control
VW Volt-Watt

I. INTRODUCTION

SMart control of distributed energy resources (DER) in distribution systems is bringing a fundamental shift in how these networks are maintained within the security limits. Historically, control methods were designed based on conventional approaches, where cyber threats have not been paid attention [1]. The idea of introducing internet protocols in the electrical network to use more advanced protection and control components has created the need to defend the cyber-attacks. Furthermore, a study on [2], [3] has showed that only 62% of cyber-attacks can be recognized after they cause massive damage to the system, which makes it a critical issue for system designers. Nowadays, the digital transformation of the electrical distribution systems has forced lots of restrictions and regulations that must be applied for achieving a secure and resilient system [4]. In the context of cyber-physical security [5], [6], smart attackers can initiate false data injection attacks (FDI) [7], where a slight change in any of the controllable devices (i.e., smart inverters, smart ring main units and digital relays), can result in disturbing the networking security without being detected by existing defense approaches. In this paper, we propose a learning-based approach for the mitigation of cyber-attacks on connected loads and DERs. Deep reinforcement learning (DRL) was opted for its superior capability of learning the power system constraints and achieving optimal control strategy. In [8], the multi-agent RL detection algorithm using deep Q-network is developed, which focuses on detecting FDI not the mitigation

A. Selim, J. Zhao and S. Park are with the Department of Electrical and Computer Engineering and M. Fei is with the Computer Science and Engineering Department, University of Connecticut, Storrs, CT 06269 USA. (e-mail: alaa.selim@uconn.edu, junbo@uconn.edu, sung_yeul.park@uconn.edu).

F. Ding is with National Renewable Energy Laboratory, Golden, Colorado (e-mail: fei.ding@nrel.gov).

This material is based upon work supported by Eversource Energy and the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number 37770.

Authorized licensed use limited to: Carleton University. Downloaded on May 10,2023 at 07:01:44 UTC from IEEE Xplore. Restrictions apply.

strategy. In [9], deep RL based recovery strategy is proposed to minimize the cyber-attack impacts under different scenarios for a transmission system. This is different from distribution system with DERs considered in this work. In [10], [11], the DRL-based approach for generating Volt-VAR and Volt-Watt curves is proposed. The action space aims to control Volt-VAR curves in mitigating cyber-attacks. This paper formulates the mitigation differently considering both DER setpoints and tie-switch actions. To illustrate more about the role of DRL in mitigating cyber-attack scenarios, [8] has demonstrated the conceptual model for maintaining cyber security. It is shown that for an offensive type of attack, the use of artificial intelligence to develop the attack strategy can be smarter and thus cannot be easily detected.

This paper proposes a DRL-based defense approach to mitigate the cyber attacks induced voltage violations and power losses on a distribution network. The main contributions of this paper work can be summarized as follows:

- The cyber attack defense problem is reformulated as a Markov decision-making process with the aim of controlling DERs and tie-switches to address the voltage violations and minimize power losses caused by cyber attacks. The original soft actor-critic (SAC) method for continuous actions has been extended to handle discrete and continuous actions for controlling DERs' setpoints and tie-switches scenarios.
- The proposed method can control DERs while avoiding infeasible switching combinations. Comparison results with the Volt-VAR and Volt-Watt methods show that our method has less power losses and achieves much better performances in regulating the voltage violation issues for both load-altering attacks and DER attacks.

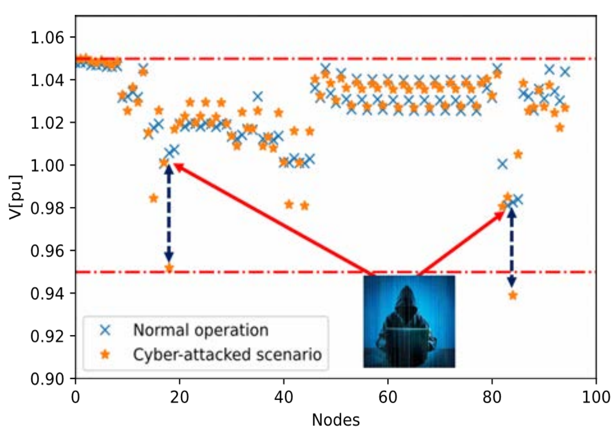

Fig. 1. Voltage violations due to cyber-attacks on the distribution system.

II. Problem Formulation

Cyber attack scenarios for loads and DERs are investigated in this paper, where these attacked nodes are the most vulnerable ones to cause system violations. Assuming the system has N nodes, our target is to maintain the system security by controlling DERs, tie-switches and load-switching actions. Battery energy storage is assumed to be co-located with intermittent DERs, a reasonable assumption as the grid and battery energy storage can be used to mitigate DER variability and uncertainties. The overloading and under-generation scenarios resulted from the cyber-attacks are specifically studied in this paper. The key challenge is to balance the system generation and load and mitigate voltage violations resulted from cyber-attacks on the loads and DERs. Fig. 1 shows one example, where attacks on some distribution system nodes can lead to voltage violations.

The problem can be formulated as an optimization model with DERs setting points and load shedding being the decision variables. At each control step t of the time horizon T, controlling the system to return back to its normal operation aims to minimize the multi-objective function as shown below:

$$\begin{aligned} minF = & \sum_{t\in T}\left[\gamma_{loss}(p_t^{losses})\right] \\ & + \sum_{t\in T}\left[\gamma_{bat}\left(p_t^{ES}-p_{t+1}^{ES}\right)\right] \\ & + \sum_{t\in T}\sum_{i\in N}\left[\gamma_v\left(v_t^i-V_{nom}\right)^2\right] \end{aligned} \tag{1}$$

$s.t.:$

$$p_t^{ch}+p_t^{dis}=p_t^{ES} \tag{2}$$

$$0\leq p_t^{ch}\leq p_{ES,max} \tag{3}$$

$$-\ p_{ES,max}\leq p_t^{dis}\leq 0 \tag{4}$$

$$S_{t+1}^{ES}=S_t^{ES}+\Delta T\left(\alpha_{ch}p_{ch}^t+\frac{p_{dis}^t}{\alpha_{dis}}\right) \tag{5}$$

$$S_{min}\leq S_t^{ES}\leq S_{max} \tag{6}$$

$$P_t^{Total}=\delta \mathrm{P}_t^{grid}+\sigma \mathrm{P}_t^{DER} \tag{7}$$

$$P_t^{grid}+P_{i,t}^{ES}+P_{i,t}^{pv}=P_{i,t}^{d}+P_{i,t}^{uc.} \tag{8}$$

$$P_i^{pv,min}<P_t^{pv}<P_i^{pv,max} \tag{9}$$

$$q_i^{pv,min}<q_t^{pv}<q_i^{pv,max} \tag{10}$$

$$q_i^{pv,min}<q_t^{pv}<q_i^{pv,max} \tag{11}$$

$$p_i^{grid,min}<p_t^{grid}<p_i^{grid,max} \tag{12}$$

$$q_i^{grid,min}<q_t^{grid}<q_i^{grid,max} \tag{13}$$

$$-P_i^{ES,ch}<P_{i,t}^{ES}<P_i^{ES,\text{ disch}} \tag{14}$$

$$0.95\leq|V_{i,t}|\leq 1.05,\forall i\in\mathcal{N} \tag{15}$$

where the coefficient in the objective function are determined based on the priorities of each term and if no specific priority is preferred, equal weights are applied. One assumption behind this optimization-based defense model is that the operator needs to timely detect the attacks, which can be a challenge to achieve in practice. Furthermore, the DER and load uncertainties make the optimization problem even more challenging. This paper develops a DRL-based approach to address them.

978-1-6654-5355-4/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Carleton University. Downloaded on May 10,2023 at 07:01:44 UTC from IEEE Xplore. Restrictions apply.

III. Proposed Extended SAC-based DRL Control for Cyber Attack Mitigation

The cyber-attack problem is first cast into the Markov decision process (MDP). The environment is configured so that the agent learns how to suppress the cyber-attacks in the network. This impact can be viewed through voltage violations and power congestion. The system experiences different conditions at each time step represented by the state space vector. Actions are taken at each time step based on the updated state from the last time step and with respect to the boundaries of the system input actions. Consequently, a reward function is formulated to reflect the effectiveness of control actions at each time step. The key elements for the MDP are shown below.

State Space S_t: the state S_t is used to represent the system status at each time step and is defined as follows:

$$S_t = \left[E_t^{ES}, P_t^{d}, P_t^{loss}\right] \tag{16}$$

Actions Space a_t: the set of available actions at each time step and it determines the continuous operational setpoints of the DERs based on the available power that can be generated and energy storage limits. Actions also include discrete operation of load shedding and tie switching, where load shedding are controlled in discrete steps based on the μ factor and tie switches are switched on/off to form the optimal combination at each time step. Formally, we have

$$a_t = \left[P_t^{DER,k}, q_t^{DER,K}, SW_t, P_t^{d,sh}\right] \tag{17}$$

Reward Function: the reward function represents the multi-objective function we would like to maximize. The reward in this problem is related to minimizing the number of shedded loads and power losses in the network while maintaining energy reserve in energy storage, and penalizing the voltage violation at each node for all time steps. These four objectives are weighted in the reward formulation with respect to their impacts on the system performance. Our target is to maximize the reward function as defined below:

$$\text{Maximize} \sum_{t=0}^{T} r\left(s_t, a_t\right) \tag{18}$$

$$r\left(s_t, a_t\right) = \sum_{t=0}^{T} \left(\alpha\left(P_{loss,t}\right) - \beta\left(E^{ES,Max} - E_{i,t}^{ES}\right)\right. \tag{19}$$

$$\left. - \rho\left(V^{Nor} - V_{i,t}\right) - \mu\left(P_{i,t-1}^{d} - P_{i,t}^{d}\right)\right)$$

Handling Continuous and Discrete Actions: The MDP can be solved by SAC algorithm [12], where the function approximators are used for both soft Q-function and policy. A parameterized soft Q-function and a tractable policy will be considered, where the parameters of these networks are θ and ϕ. For example, the soft Q-function can be modeled as expressive neural networks, and the policy as a Gaussian distribution with mean and covariance given by neural network as shown below:

$$\pi = \arg\max E\left[\sum_{t=0}^{\infty} \gamma^t \left(R\left(s_t, a_t\right) + \alpha H\left(\pi\left(\cdot \mid s_t\right)\right)\right)\right] \tag{20}$$

$$J_\pi(\phi) = E_{s_t \sim D}\left[E_{a_t \sim \pi_\phi}\left[\alpha \log\left(\pi_\phi\left(a_t \mid s_t\right)\right) - Q_\theta\left(s_t, a_t\right)\right]\right] \tag{21}$$

where γ is a future discount coefficient; $R\left(s_t, a_t\right)$ represents the expected discount reward with state s_t and taking actions following actor policy π; $H\left(\pi\left(\cdot \mid s_t\right)\right)$ is the entropy term; $\alpha > 0$ is the trade-off coefficient. Adding the entropy term is to encourage the agent to explore more possibilities in action space. SAC has been shown to have a greater advantage to handle stochastic models of uncertainties and intermittent nature of the DERs via the entropy term. The latter allows the auto-tuning of parameters to handle the percentages of uncertainties in generation or demand. However, the SAC is not designed to address the mixed action space with both continuous and discrete control actions. The continuous actions come from DERs and energy storage while discrete actions are related to switch changes. Motivated by [13], we propose to modify the Q-function so as to deal with both continuous and discrete control actions via:

$$J_\pi(\phi) = E_{s_t \sim D}\left[\pi_t\left(s_t\right)^T\left[\alpha \log\left(\pi_\phi\left(s_t\right)\right) - Q_\theta\left(s_t\right)\right]\right] \tag{22}$$

Inside SAC, there are two neural networks known as actor and critic networks, where actor network is designed to find the best action corresponding to the current state and critic network is designed to find the Q-value of the executed action in the current state. The critic network computes target for the Q function as:

$$y\left(s_t, r_t, s_{t+1}\right) = r + \gamma\left(Q\left(s_{t+1}, a_{t+1}\right) - \alpha \log \pi_\theta\left(a_{t+1} \mid s_{t+1}\right)\right) \tag{23}$$

The final stage of the proposed DRL algorithm is deducing θ and ϕ to continue in an iterative process to reach the best action values for each trained scenario.

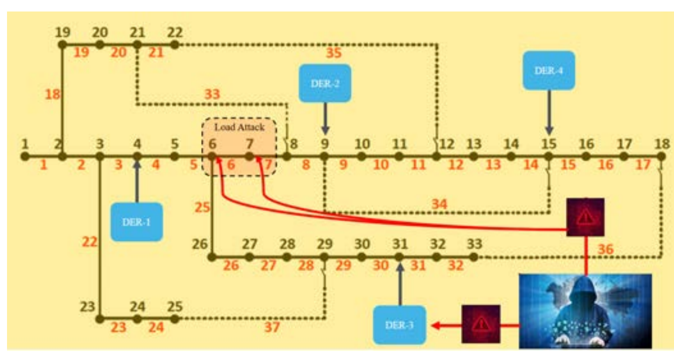

Fig. 2. Modified IEEE 33-node system with tie switches and DERs.

IV. Numerical Results

A modified IEEE 33-node system with three-phase loads and 4 DERs that are utility-owned (i.e., each DER consists of 1 ES unit and 1 PV with installed capacities of 500 kW each unit), is used for testing. The system is modeled using OpenDSS and is configured to be in the grid connected mode. At the first solution evaluated using OpenDSS, no voltage violation has been observed during the normal operation. In addition, 4 tie switches have been added to the network and no

978-1-6654-5355-4/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Carleton University. Downloaded on May 10,2023 at 07:01:44 UTC from IEEE Xplore. Restrictions apply.

sectionalizers are considered for operation. The learning environment is designed according to OpenAI Gym [14], which is a common interfacing library to define DRL environment for the agent. The SAC algorithm is implemented using PyTorch. Specifically, in the SAC, both actor and critic networks are designed as feed-forward neural networks with three hidden layers of 50, 100 and 50 neurons and a ReLU activation function for each layer. Other SAC hyper-parameters are as follows: Adam optimizer is used with a learning rate of 0.0001 and discount factor γ is set to 0.9. The target network is updated by tau $= 0.001$ and random process is applied for better exploration with $\boldsymbol{\alpha} = 0.1$, $\beta = 0.1$ $\rho = 10$ and $\mu = 0.1$; the replay buffer size is 100000 with batch size 256. The offline DRL training spends around 3 hours and 30 minutes on a laptop computer with $3.6\mathrm{GHz}$ Intel i7 processor and 32.0 GB RAM. The proposed DRL defense algorithm is compared with other control algorithms, such as the Volt-VAR and Volt-Watt using the default control curves in OpenDSS [15]). Also, the MPC algorithm is implemented following [16] based on the problem formulation in Section II using the same control and state variables for the proposed DRL algorithm. Attacks are initiated for few timesteps by changing the load and DER setpoints (% power change of loads and/or DERs) in OpenDSS using python interface.

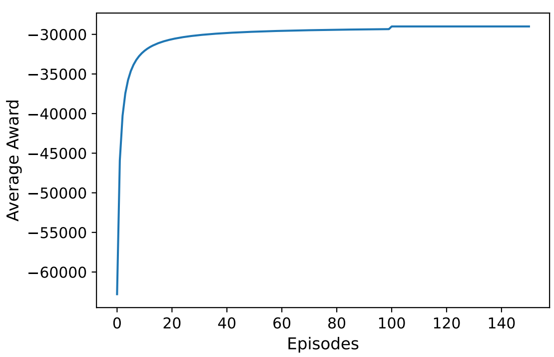

Fig. 3. Learning curve for the proposed SAC agent

The DRL agent is trained for window of 290 time-steps (i.e., each time-step is 5 min, yielding 24 hrs). During this window, DRL explores the best policy to control DER setpoints and tie-switches so as to mitigate voltage violations and minimize the network power losses. Fig. 3 shows the convergence curve. It can be found that the DRL agent can converge to a good reward function just after about 40 episodes. More episodes of training leads to slight improvement on the final reward. This shows that the DRL agent can quickly learn the good control policy in defending the cyber attacks, which will be shown in the next two sections. As for the control execution, DRL and MPC are remotely executed from a centralized controller. However, Volt-VAR and Volt-Watt are locally managed for each corresponding DER. Furthermore, for the MPC and DRL approaches, we assume those system states come from the centralized distribution system state estimator.

A. Load-Altering Attack Scenario

This section evaluates the effectiveness of the proposed approach for defending the load- altering attacks. Following the settings shown in Table I for initiating the load-altering attacks, it is found that attacking nodes 6 and 7 at the same time is most likely to cause voltage violations (based on trial and error). Therefore, these two nodes are used for simulating different types of attack scenarios. After defining the targeted nodes, the performance of each method is tested and the results are shown in Fig. 5 and Table II. It can be observed that the proposed DRL-based approach is able to successfully regulate the voltage within security limit and has the least power losses among all approaches. Volt-Watt and Volt-VAR approaches have low voltage violation issues and the power loss is 10% higher. The MPC-based method has some issues in high voltage violation but its power loss is better than Volt-Watt and Volt-VAR approaches.

TABLE I
LOAD ALTERING ATTACK SETTINGS.

Load Change(%)	Vmin	Vmax	Plosses(%)
0	0.9619	1.0486	7.719
20	0.9394	1.0486	14.62
40	0.9392	1.0486	14.65
60	0.9390	1.0486	15.1
100	0.9389	1.0486	15.12
200	0.9366	1.0485	15.37

TABLE II
STATISTICS OF THE COMPARISON RESULTS UNDER LOAD ALTERING ATTACKS.

Control method	**Vmin**	**Vmax**	**Plosses(kW)**	**Plosses (%)**
Non-control	0.9394	1.0486	585.421	14.62
Volt-Watt	0.9389	1.0482	593.52	15.1
Volt-VAR	0.9394	1.0486	585.421	14.62
Proposed DRL	0.9812	1.0482	62.366	3.185
MPC	0.9645	1.06	407.858	10.95

B. DER Setting Point Attack Scenario

This section evaluates the effectiveness of the proposed DRL approach for defending the attacks on DER setting points. Following the settings shown in Table III in initiating the DER altering attacks. Initially, the system does not have any voltage violations. However, with only 10% of shifting on the setting points of all DERs, the system detects low voltage issues under no control scenarios. The performance of each method is tested and the results are shown in Table IV. It can be observed that the proposed DRL-based approach is able to successfully regulate the voltage within the security limit and has the least power losses among all approaches. By contrast, the other two widely used Volt-VAR and Volt-Watt control approaches have serious over-voltage issues. The proposed DRL approach can regulate the voltage close to the security boundary without violating them. This is because the agent can successfully identify the right combinations of DER setting point changes among non-attacked DERs to mitigate the impacts caused by those with attacks on DERs. It is interesting to notice that the attacks on DERs seem to cause

978-1-6654-5355-4/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Carleton University. Downloaded on May 10,2023 at 07:01:44 UTC from IEEE Xplore. Restrictions apply.

more severe voltage issues than the load-altering attacks. This is expected as the increased penetration of DERs itself is likely to induce voltage security problem.

TABLE III
DER ALTERING ATTACK SETTINGS.

DER Change(%)	Vmin	Vmax	Plosses(%)
0	0.96188	1.2132	7.719
-10	0.94941	1.2132	32.67
-20	0.93921	1.2132	33.79
-40	0.92899	1.2132	45.21

TABLE IV
STATISTICS OF THE COMPARISON RESULTS UNDER DER ALTERING ATTACKS.

Control method	**Vmin**	**Vmax**	**Plosses(kW)**	**Plosses (%)**
Non-control	0.9645	1.2132	619.629	20.64
Volt-Watt	1.0042	1.2132	619.562	16.583
Volt-VAR	0.98292	1.2091	601.545	15.62
Proposed DRL	0.952321	1.050182	432.897	10.892
MPC	0.9734	1.08	538.9	10.95

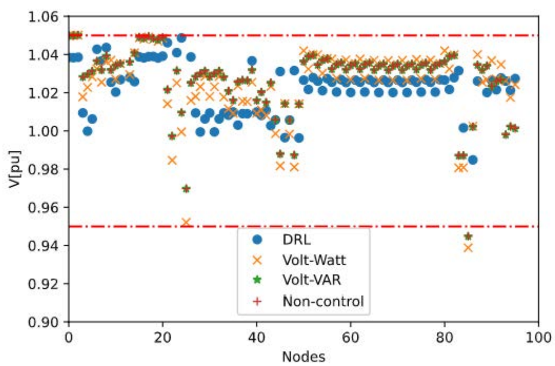

Fig. 4. Voltage regulation performance comparison results under load attacks.

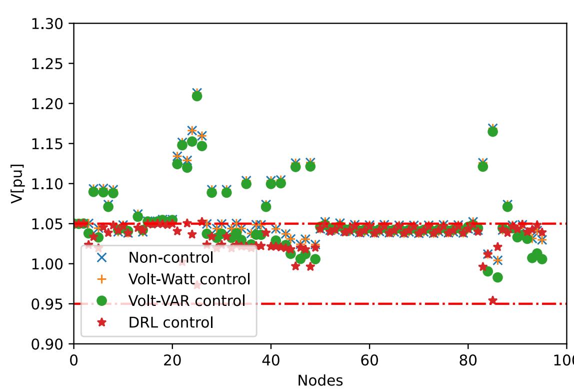

Fig. 5. Voltage regulation performance comparison results under DER attacks.

V. CONCLUSION

This paper proposes a DRL-based defense approach to mitigate cyber-attacks on loads and DERs. The objectives were to regulate the voltage while minimizing the distribution network power losses. Unlike the optimization-based approach that needs to consider specific attack scenarios, our proposed method formulates the cyber-attack defense problem as a MDP problem. We have modified the original SAC approach to deal with discrete and continuous actions for controlling DERs' setpoints and network switches. Comparison results with other control approaches show that the the proposed method controls DERs and network switches effectively to achieve less power losses and voltage violations. Future work will enhance the scalability and robustness of the approach for larger networks and topology changes. We will develop a robust DRL to defense adversarial agents, including corrupted data in both training and online applications.

REFERENCES

[1] H. Zeynal, M. Eidiani, and D. Yazdanpanah, "Intelligent substation automation systems for robust operation of smart grids," *2014 IEEE Innovative Smart Grid Technologies-Asia (ISGT ASIA)*, pp. 786–790, 2014.

[2] J. Jang-Jaccard and S. Nepal, "A survey of emerging threats in cyber-security," *Journal of Computer and System Sciences*, vol. 80, no. 5, pp. 973–993, 2014.

[3] Y. Li and Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments," *Energy Reports*, vol. 7, pp. 8176–8186, 2021.

[4] M. T. A. Rashid, S. Yussof, Y. Yusoff, and R. Ismail, "A review of security attacks on iec61850 substation automation system network," in *Proceedings of the 6th International Conference on Information Technology and Multimedia*, pp. 5–10, IEEE, 2014.

[5] J. Qi, A. Hahn, X. Lu, J. Wang, and C.-C. Liu, "Cybersecurity for distributed energy resources and smart inverters," *IET Cyber-Physical Systems: Theory & Applications*, vol. 1, no. 1, pp. 28–39, 2016.

[6] S. Sahoo, T. Dragičević, and F. Blaabjerg, "Cyber security in control of grid-tied power electronic converters—challenges and vulnerabilities," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 5, pp. 5326–5340, 2019.

[7] D. An, Q. Yang, W. Liu, and Y. Zhang, "Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach," *IEEE Access*, vol. 7, pp. 110835–110845, 2019.

[8] A. J. Abianeh, Y. Wan, F. Ferdowsi, N. Mijatovic, and T. Dragičević, "Vulnerability identification and remediation of fdi attacks in islanded dc microgrids using multiagent reinforcement learning," *IEEE Trans. Power Electron.*, vol. 37, no. 6, pp. 6359–6370, 2021.

[9] F. Wei, Z. Wan, and H. He, "Cyber-attack recovery strategy for smart grid based on deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2476–2486, 2019.

[10] C. Roberts, S.-T. Ngo, A. Milesi, S. Peisert, D. Arnold, S. Saha, A. Scaglione, N. Johnson, A. Kocheturov, and D. Fradkin, "Deep reinforcement learning for der cyber-attack mitigation," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–7, IEEE, 2020.

[11] C. Roberts, S.-T. Ngo, A. Milesi, A. Scaglione, S. Peisert, and D. Arnold, "Deep reinforcement learning for mitigating cyber-physical der voltage unbalance attacks," in *2021 American Control Conference (ACC)*, pp. 2861–2867, IEEE, 2021.

[12] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*, pp. 1861–1870, PMLR, 2018.

[13] P. Christodoulou, "Soft actor-critic for discrete action settings," *arXiv preprint arXiv:1910.07207*, 2019.

[14] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.

[15] W. Sunderman, R. C. Dugan, and J. Smith, "Open source modeling of advanced inverter functions for solar photovoltaic installations," in *2014 IEEE PES T&D Conference and Exposition*, pp. 1–5, IEEE, 2014.

[16] P. Li, J. Ji, H. Ji, J. Jian, F. Ding, J. Wu, and C. Wang, "Mpc-based local voltage control strategy of dgs in active distribution networks," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2911–2921, 2020.

978-1-6654-5355-4/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Carleton University. Downloaded on May 10,2023 at 07:01:44 UTC from IEEE Xplore. Restrictions apply.