2024 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) | 979-8-3503-1360-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISGT59692.2024.10454163

Deep RL-based Volt-VAR Control and Attack Resiliency for DER-integrated Distribution Grids

Kundan Kumar, *Member, IEEE,* Gelli Ravikumar, *Member, IEEE,*
Department of Computer Science and Electrical and Computer Engineering, Iowa State University, United States, IA, 50010
Email:kkumar@iastate.edu, gelli@iastate.edu

***Abstract*—Integrating distributed energy resources (DERs) into a power system requires more advanced control mechanisms. One of the control strategies used for Volt-VAR control (VVC) is to manage voltage and reactive power. With the increase in the complexity of the power system, there is a need to develop an autonomous and robust control mechanism using deep reinforcement learning (DRL) to enhance grid performance and adjust voltage and reactive power settings. These adjustments minimize losses and enhance voltage stability in the grid. In this paper, we proposed a novel approach to develop a DRL-based VVC framework and mitigation techniques to protect against stealthy white-box attacks targeting the trained control policies of the DRL model. The mitigation technique on the trained DRL is proposed to control the voltage violations on the smart grid to enhance the stability of the grid and minimize voltage irregularities. Our proposed mitigation technique provided better control policies for DRL-based VVC, successfully mitigating 100 percent of voltage violations in the smart grid environment. The results show that the mitigation technique enhances the security and robustness of trained DRL VVC agents.**

***Index Terms*—adversarial attacks, distributed energy resources (DERs), deep reinforcement learning (DRL), grid security, smart grid, Volt-VAR Control (VVC), reinforcement learning (RL)**

I. Introduction

Smart grids are enhanced by integrating an advanced distributed management system (ADMS) and control mechanisms to optimize electricity generation, distribution, and consumption. The Volt-VAR control (VVC) mechanism technique is used to regulate voltage levels typically between 0.95 and 1.05 p.u. of voltage according to the ANSI C84.1-2020 standard [1], ensures efficient power delivery and grid stability. Adversarial attacks disrupt the normal operation of ADMS. These adversarial attacks, such as injecting false data [2], involve injecting false voltage or reactive power measurements into the power system. Denial of service attacks [3] target the communication infrastructure of VVC systems. Data validation and filtering defense mechanisms detect and mitigate the false data injection [4], [5]. These defense mechanisms are crucial for ensuring the reliable and secure operation of smart grid VVC applications. The utilization of artificial intelligence (AI) with DRL models for the enhancement of control logic and its capacity to adjust grid conditions is becoming common in smart grids. The adversarial attacks on DRL models [6] manipulate control policies and cause vulnerabilities in the model. The defense mechanisms such as Adversarial training [7] to mitigate the adversarial attack impact. The risk of adversarial attacks [8] is increasing because attackers exploit vulnerabilities in the AI model, leading to unexpected results.

In the last five years, research studies have proposed methods for centralized resilient secondary control of energy storage systems versus DoS attacks [9] with multi-agent reinforcement learning. The researchers proposed a few defense mechanisms [10], [11] that detect and prevent adversarial attacks in smart grids. These attacks are modeled on datasets, and the authors propose mitigating strategies using DRL for modeled attack datasets.

Based on the literature conducted by the author in the field of power systems, especially in VVC, it has been observed that the proposed strategies and techniques are developed on input or historical data sets. However, there seems to be a lack of mitigation techniques and strategies for the trained DRL-based model. In light of this, we proposed a clipping technique that could potentially mitigate attacks on trained DRL-based VVC models in the smart grid. The key contributions of our research are:

- Developed a DRL framework for VVC policies for the distribution grid for regulating bus voltages and optimizing power distribution.
- Proposed a stealthy cyberattack technique on the trained advantage actor-critic (A2C) DRL agent, which compromises the control policies of VVC.
- Proposed mitigation techniques against stealthy cyberattacks to enhance grid stability and minimize voltage irregularities in the smart grid.
- Performed impact analysis to determine the effectiveness of the mitigation technique on the IEEE-13 bus system.

The paper is organized as follows. Section II proposed a DRL-based VVC framework for smart grids. Section III proposes stealthy cyberattacks and mitigation strategies for the trained DRL-based VVC model. Section IV demonstrates and evaluates the mitigation technique on the trained DRL model for the IEEE-13 bus system. Finally, Section V presents the conclusion.

II. Proposed DRL-based Volt-VAR Control Framework for Smart Grid

In this section, we propose a DRL-based VVC using the Markov Decision Process (MDP) modeled with the IEEE-13 circuit OpenDSS datasets and train with the A2C DRL model.

A. MDP for DRL

In a smart grid environment, the VVC manages voltage levels and reactive power, and it is designed as an MDP. The state space S_t in MDP is defined by voltage and reactive power levels, while the action space A_t represents the possible adjustments to reactive power. The reward function encourages the agent to maintain voltage levels within the desired range

while minimizing the cost of reactive power generation. At each discrete time step t, the agent takes an action during the periods. The goal of the agent is to maximize the cumulative reward over these actions while taking into account the control error f_{ctrl}, power loss f_{power}, and voltage violation f_{volt} as shown in (1) of the reward function over a time period. The detailed reward function is described here [12].

$$R(s, s_0, i) = -f_{\text{volt}}(s_0) - f_{\text{ctrl}}(s, s_0, i) - f_{\text{power}}(s_0) \quad (1)$$

In the given scenario, s refers to all observations, whereas s_0 refers to the next step. The variable i indicates the episode step, which ranges from 0 to $H-1$, where H is the total number of timesteps in the MDP. The control error f_{ctrl} measures the deviation between the agent's intended reactive power adjustments of regulators, capacitors, and batteries and the actual state of the smart grid system. f_{power} is the ratio of power loss to total power, and f_{volt} represents the sum of voltage violations in all phases and nodes of the system. The reward function stated in (1) tends to indicate better performance and lower susceptibility to anomalies when the reward value is closer to zero.

B. Data Preparation and Feature Extraction for DRL Model

In the data preparation and feature extraction, we simulate the circuit components (buses, lines, transformers, etc.) with their respective properties. Voltage magnitudes $V = v_1, v_2, \ldots, v_n$ are collected at various bus locations using the OpenDSS API query functions for each time step. After loading the circuit topology, a graph representation is constructed using circuit information. Each node corresponds to a bus, and the edges represent the connections (lines, transformers) between the buses. The state includes voltage magnitudes and power flow information about the current grid conditions and records voltage values over time to analyze voltage trends.

C. A2C DRL algorithm for VVC

The Advantage Actor-Critic algorithm (A2C) is a model for decision-making in an environment that combines actor and critic networks. The A2C algorithm uses value- and policy-based approaches to improve exploration and exploitation while minimizing voltage perturbations at each bus and maximizing overall reward. The actor-network, also known as the policy network, selects actions based on the current state of the environment and outputs a probability distribution over possible actions. The action a is sampled using a Gaussian distribution. $\pi(a|s;\theta_{\text{actor}})$ is the stochastic policy, where $\pi(a|s)$ is the probability of taking action a given state s, and θ_{actor} are the parameters of the policy of the actor neural network as shown in Fig. 1. The critic network, also known as the value network, estimates the expected cumulative reward of being in a particular state and following the actor's policy. It inputs the current state and outputs a value function as $V(s)$. The estimated value of state s under the value function of the critic is represented by $V(s;\theta_{\text{critic}})$, where θ_{critic} represents the parameters of the critic neural network as shown in Fig. 1.

In the A2C algorithm, The actor-network uses the critic's evaluation with temporal difference (TD) error to adjust the policy and favor actions expected to yield higher rewards. It creates a continuous feedback loop, where the actor learns to improve its policy based on the critic's assessments. Improve voltage adjustment based on current levels and estimate expected rewards while mitigating voltage violations.

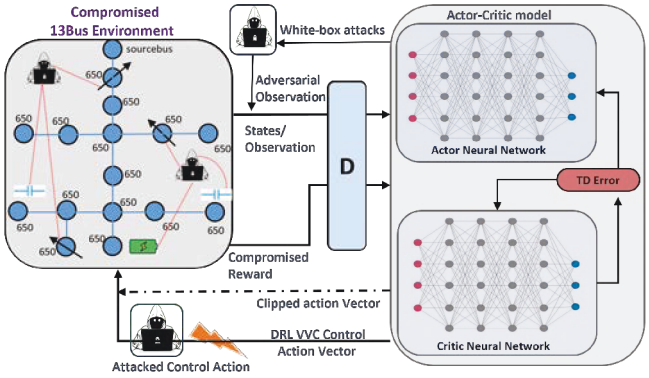


Fig. 1: Proposed DRL-based VVC framework for stealthy attack and mitigation.

III. Proposed Stealthy Cyberattack and Mitigation Strategies on the DRL Model

In this section, we propose a technique to address potential stealthy cyberattacks on the DRL model deployed in the smart grid. The focus is on mitigating voltage violations in the smart grid system and improving overall reward performance. It is needed as we know that IT security ensures data integrity and provides secure communication between data and the environment, but no system is fully secured, so each part has to be secure individually. We assume that attackers pass IT security and have access to the trained DRL model.

A. Stealthy Cyberattacks on Action Vectors in DRL Model

Stealthy cyberattacks target the actions available to the agent in a given state of the environment. These attacks, known as action vector attacks in DRL models, aim to manipulate or perturb the action choices made by the DRL model without being detectable. The goal of these attacks is to degrade the decision-making process of the DRL agent, causing unusual behavior on the smart grid VVC application component.

1) Stealthy white-box attack on A2C model for action vector: A stealthy white box attack is a kind of adversarial attack in which the attacker gains access to the internal parameters of the actor θ_{actor} and critic θ_{critic} networks, including neural network weights and architecture, responsible for determining the model action vector and estimating the value function for a given state. It is different from a black-box attack, as the black-box attacker has limited or no knowledge about the internal workings of the target system. In this attack vector, the attacker has only access to the control action vector of the DRL VVC agent.

The paper discusses how the control policies of the DRL VVC model are targeted by attackers through a series of stealthy cyberattacks. The attackers intentionally modify the control policies, which undermines the decisions made by the

979-8-3503-1360-4/24/$31.00 ©2024 IEEE

model and disrupts the functionality of the actuators in the smart grid VVC. The attackers focus on the action vectors of the trained DRL model rather than modifying its internal parameters, as modifying the internal parameters affects the behavior and performance of the overall DRL model, but we cannot analyze the individual components (batteries, capacitors, and regulators) of the smart grid environment, which are key factors in decision making of voltage levels and reactive power. The attacker's primary objective is to maximize the expected reward by perturbing the actions generated by the A2C model to improve voltage control. Fig. 1 shows how the IEEE-13 distribution test feeders are compromised due to control actions by DRL VVC, which hampers the operational functionalities of capacitors, batteries, and regulators.

$$\max_{\delta} \mathbb{E}[R(\pi(a+\delta, s))] \tag{2}$$

The action vector in VVC often includes adjustments to the reactive power generation in the voltage regulators, capacitors, and regulators of the smart grid, which aim to control voltage levels within a desired range. The perturbations (δ) to the action vector cause changes to the control actions and alter the expected reward, ultimately degrading the voltage control as defined in (2). The expected reward R in (2) quantifies how good or bad a state s action a pair achieves the agent's objectives by maximizing the expected cumulative reward over time in the policy π.

B. Proposed detection and mitigation techniques against stealthy white-box attacks on trained DRL model

In this section, we propose a detection and mitigation strategy to reduce VVC bus voltage violations in the trained A2C DRL model. We first detected the attacks and incorporated a few techniques to improve the effectiveness of voltage violation mitigation on the trained model.

1) Detection of stealthy attack for DRL action vector: We detect attacks by creating the detector module D as shown in Fig. 1, which continuously monitors the control actions of the DRL VVC agent and checks actions that violate operational constraints or safety limits according to the ANSI standard [1].

2) Enforcing Constraint on DRL Attack Action Vector: In this method, the L2-norm constraint, as shown in (3), is used to measure the quantification of the perturbation vector δ, which changes in control actions that impact voltage and reactive power levels, and its magnitude signifies the degree of deviation from optimal control actions. By constraining the L2-norm($||\delta||_2$) of δ to be less than or equal to a specified threshold represented by ϵ as shown in (3), the constraint ensures that control actions remain within certain predefined limits to maintain the stability and effectiveness of VVC operations.

$$||\delta||_2 \leq \epsilon \tag{3}$$

3) Mitigation with clipping action vector during inference: To further protect VVC operations during inference, the action vector is clipped. This technique modifies control actions by limiting the action vector to prevent potential stealthy adversarial white-box attacks, as shown in (4).

$$a' = \text{clip}(a+\delta, a_{\text{min}}, a_{\text{max}}) \tag{4}$$

In (4), the clipped action vector modifies the control actions used for VVC and is denoted by a'. These clipped actions are adjusted to ensure that they fall within predetermined safe bounds, where a_{min} and a_{max} represent the lower and upper bounds of the action vector, respectively. a_{min} is the minimum allowable setting for control actions related to voltage regulation and reactive power components of the smart grid to prevent undervoltage. Conversely, a_{max} is the upper bound of the action vector, representing the maximum allowable setting or adjustment for control actions to avoid overvoltage. By employing the action clipping technique using (4), the modified control actions (a') remain within these predefined bounds (a_{min} and a_{max}), ensuring grid stability and safety and preventing significant deviations in control actions.

4) Updated Objective with Clipping: The primary aim of the attacker is to maximize the expected reward by manipulating the actions of the DRL agent, as shown in (5). The reward function evaluates the expected reward for an agent's specific action. The agent policy $\pi(\text{clip}(a+\delta, a_{\text{min}}, a_{\text{max}}), s)$ determines how it selects actions given the current state s and the action is clipped to ensure that it stays within the range of a_{min} and a_{max} even as attacker attacks the action vector.

$$\max_{\delta} \mathbb{E}[R(\pi(\text{clip}(a+\delta, a_{\text{min}}, a_{\text{max}}), s)) \tag{5}$$

The objective function in (5) is critical to prevent attackers from forcing the DRL agent to take actions that result in voltage or reactive power violations. Even if an attacker tries to manipulate the agent actions (δ), capping the actions ensures the resulting actions remain within safe operational limits. However, this clipping differs from DRL in that action clipping ensures that the sampled actions generated by the policy network remain within the valid action space. It is a way to enforce constraints on the actions taken by the agent. The clipping adds a layer of security to the DRL system, making it more resilient to adversarial attacks and protecting it against actions that compromise its safety and reliability.

C. Performance of success

Performance P is calculated as the percentage change in voltage violations after the attack and mitigation techniques described in (6).

$$\text{P}(\%) = \left(\frac{V_{vio_{attack}} - V_{vio_{mitigation}}}{V_{vio_{attack}}} \right) \times 100 \tag{6}$$

IV. TESTBED-BASED IMPLEMENTATION ON IEEE-13 BUS

In this section, we pre-processed the data and trained the DRL A2C algorithms on the IEEE-13 Bus distribution grid, measured the total voltage violation across the nodes in the trained DRL model, and also performed the stealthy attack on the trained DRL, and performed a mitigation technique to reduce the impact of the attack.

979-8-3503-1360-4/24/$31.00 ©2024 IEEE

A. Data pre-preprocessing for DRL-based VVC

To pre-process the data for DRL, we load the IEEE-13 circuit model into the simulation and iterate through it for each time step. We use the OpenDSS with Python API query functions at each time step to gather voltage magnitudes ($V = v_1, v_2, \ldots, v_n$) at different bus locations. The agent takes one action per hour over 24 hours, with the number of actions equal to 24 x T, where T is the number of time steps within each hourly period. The value of T varies depending on how time is discretized within each hour. In our scenario, the agent takes only one action for each time step.

B. Training DRL Algorithm for VVC

After pre-processing the data, the data are passed to the A2C architecture. θ_{actor} chooses actions that offer the highest expected rewards, while θ_{critic} evaluates the value of states to determine the quality of the actions. During the training phase, we fine-tuned various hyperparameters, such as the learning rate and discount factor, to ensure that the model could effectively regulate the voltage and control the reactive power. This involved carefully balancing exploration and exploitation by configuring these hyperparameters, which enabled the model to learn optimal control policies. Consequently, the model maintains grid stability and reduces voltage violations across all 13 Bus test feeder nodes. To develop the actor and critic, we implemented a neural network architecture consisting of two hidden layers, each with 32 neurons ([32, 32]) with a discount factor of 0.9 and a smaller learning rate of 0.001, which allowed for smaller steps during training.

After training the DRL A2C algorithms, we discovered that there were only 11 V_{vio} instead of 472 V_{vio} without trained DRL algorithms, as demonstrated in Fig. 2a and Fig. 2b.

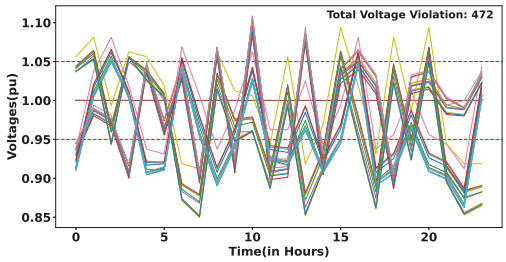
TABLE I: Voltage violations comparison after white-box attack on trained DRL VVC Components

Attack-vector	***Voltage Violations across IEEE-13 Node***
No Attack	11
Attack on Capacitors	49
Attack on Regulators	81
Attacks on Battery	345
Attacks on All of them	840

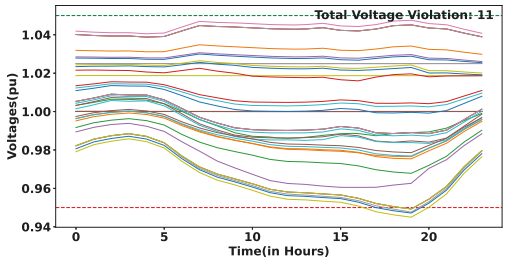
C. Perform a stealthy attack on trained DRL VVC algorithm

The white-box attack is performed over time (hourly) to target the action vector of the trained DRL model, which leads to 840 V_{vio} from 11 V_{vio} on the IEEE-13 Bus environment as shown in Fig. 3a which has one battery and two capacitors and three regulators as shown in Fig. 1. As shown in Fig. 3a, all bus voltage levels dropped below the critical threshold of 0.95 pu, posing a significant risk of voltage collapse and potential power failures and reducing system efficiency.

Furthermore, we quantified the attacked action vector of dedicated components of the IEEE-13 Bus. Table I shows that attacks on the action vector of the capacitors caused 49 V_{vio}, while the action vector of the regulators caused 81 V_{vio} and the battery had 345 V_{vio}, with attacks on all components leading to 840 V_{vio}. It gives an insight into the individual components of the 13Bus grid that get affected and is also useful for taking appropriate measures to prevent attacks and improve the overall security of the grid. The reward function is evaluated, as demonstrated in Fig. 4a. In this study, attacks are executed on individual capacitors (attack1 and attack2), regulators (attack3, attack4, and attack5), and batteries (attack6), as well as on the three components (attack7), using the action vector of a trained DRL VVC agent. Fig. 4a illustrates the impact of multiple stealth attacks on the components of the IEEE-13 Bus, indicating the number of attacks that caused the model to perform poorly with new data points. A reward close to zero suggests that the model is less susceptible to perturbations, while more negative values indicate suboptimal system performance characterized by an increased incidence of control errors, greater power losses, and more frequent voltage violations.



(a)



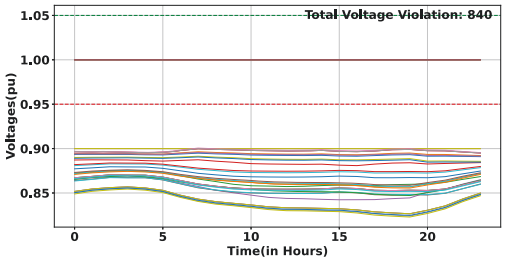
(b)

Fig. 2: (a) Voltage violations without DRL over the 24-hour time horizon. (b) Voltage violations with DRL algorithms over the 24-hour time horizon.

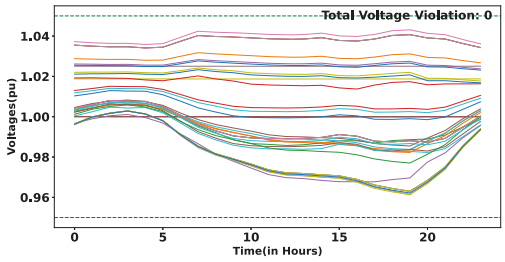
D. Perform mitigation techniques for voltage violations

The mitigation technique is applied to restrict the ability of the attacker to modify the action vector within a certain range. To mitigate the voltage violations, the L2 norm is used, as described in (3), whereby the value of ϵ is determined by identifying the maximum magnitude of the action vector. The clipping parameter is also used, as illustrated in (4). The process of selecting the best maximum and minimum clipping boundaries, a_{min} and a_{max}, respectively, is iterative. In this case, a_{min} is set to 10 while a_{max} is set to 20 of the action vector of the trained DRL model. Applying clipping mitigation strategies reduced voltage violations from 840 to 0, as shown in Fig. 3b, where all bus voltages are now within acceptable limits [1]. We measured the performance P as described

979-8-3503-1360-4/24/$31.00 ©2024 IEEE

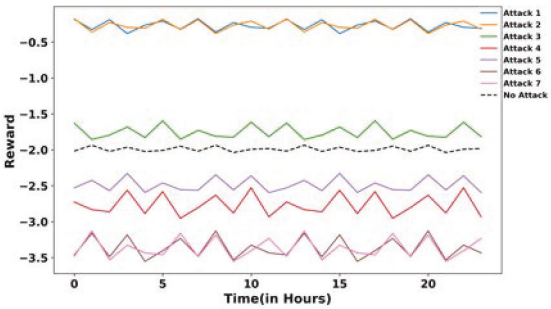


(a)

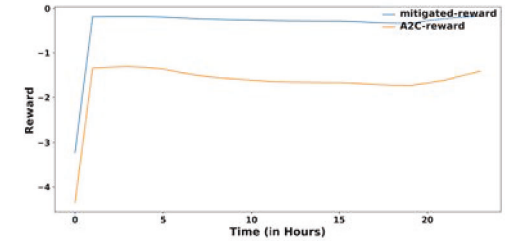


(b)

Fig. 3: (a) Voltage violations after stealthy attacks over a 24-hour time horizon. (b) Voltage violations after mitigation technique on the stealthy attack on the trained DRL algorithms over a 24-hour time horizon.



(a)



(b)

Fig. 4: (a) Comparison of reward on stealthy attack on the different components of IEEE-13 Bus test feeder. (b) Comparison of reward after mitigated trained DRL with trained A2C-DRL.

in (6), showing that the proposed mitigation strategies have successfully mitigated 100% of voltage violations across bus voltages. The cumulative reward, calculated while using the clipping limit as described in (4), is depicted in Fig. 4b, which demonstrates that the mitigated trained DRL model is a better reward than the trained DRL model, indicating that the model is more robust and has better resiliency against attacks.

V. CONCLUSION

This paper presented a mitigation technique to protect against adversarial attacks on the trained DRL model. First, we develop the DRL framework to train the A2C DRL model and perform a stealthy cyberattack on the trained A2C model. Second, clipping mitigation techniques are used to reduce voltage violations of the attacked model. With this approach, we successfully mitigated voltage violations of 100%, and the reward also performed better than the trained DRL model, making that model more robust against stealthy adversarial attacks and improving its accuracy. The future focus will be developing control policies to improve computational efficiency and make a robust and interpretable DRL model that ensures better transparency and trustworthiness.

ACKNOWLEDGMENT

This research is funded partly by US NSF Grant # CNS 2105269 and US DOE CESER Grant DE-CR000016.

REFERENCES

[1] American National Standards Institute, "American national standard for electric power systems and equipment – voltage ratings (60 hz)," American National Standards Institute, Tech. Rep., 2020, aNSI C84.1-2020.

[2] S. E. Huang, W. Wong, Y. Feng, Q. A. Chen, Z. M. Mao, and H. X. Liu, "Impact evaluation of falsified data attacks on connected vehicle based traffic signal control," *ArXiv*, vol. abs/2010.04753, 2020.

[3] A. Huseinović, S. Mrdović, K. Bicakci, and S. Uludag, "A survey of denial-of-service attacks and solutions in the smart grid," *IEEE Access*, vol. 8, pp. 177 447–177 470, 2020.

[4] A. Sargolzaei, K. Yazdani, A. Abbaspour, C. D. Crane III, and W. E. Dixon, "Detection and mitigation of false data injection attacks in networked control systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4281–4292, 2020.

[5] G. Ravikumar and M. Govindarasu, "Anomaly detection and mitigation for wide-area damping control using machine learning," *IEEE Transactions on Smart Grid*, pp. 1–1, 2020.

[6] H. Liang, E. He, Y. Zhao, Z. Jia, and H. Li, "Adversarial attack and defense: A survey," *Electronics 2022, Vol. 11, Page 1283*, vol. 11, p. 1283, 4 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/8/1283/htm https://www.mdpi.com/2079-9292/11/8/1283

[7] D. M. Ziegler, S. Nix, L. Chan, T. Bauman, P. Schmidt-Nielsen, T. Lin, A. Scherlis, N. Nabeshima, B. Weinstein-Raun, D. Haas, B. Shlegeris, and N. Thomas, "Adversarial training for high-stakes reliability," *ArXiv*, vol. abs/2205.01663, 2022.

[8] K. Ren, T. Zheng, Z. Qin, and X. Liu, "Adversarial attacks and defenses in deep learning," *Engineering*, vol. 6, pp. 346–360, 3 2020.

[9] P. Chen, S. Liu, B. Chen, and L. Yu, "Multi-agent reinforcement learning for decentralized resilient secondary control of energy storage systems against dos attacks," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 1739–1750, 2022.

[10] F. Wang, M. C. Gursoy, and S. Velipasalar, "Adversarial reinforcement learning in dynamic channel access and power control," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021, pp. 1–6.

[11] A. T. El-Toukhy, M. M. Badr, M. M. E. A. Mahmoud, G. Srivastava, M. M. Fouda, and M. Alsabaan, "Electricity theft detection using deep reinforcement learning in smart power grids," *IEEE Access*, vol. 11, pp. 59 558–59 574, 2023.

[12] T. Fan, X. Y. Lee, and Y. Wang, "Powergym: A reinforcement learning environment for volt-var control in power distribution systems," *CoRR*, vol. abs/2109.03970, 2021. [Online]. Available: https://arxiv.org/abs/2109.03970