An AI-based Approach for Scalable Cyber-physical Optimal Response in Power Systems

Shining Sun, *Student Member, IEEE,* Shamina Hossain-McKenzie, *Member, IEEE,*
Leen Al Homoud, *Student Member, IEEE,* Khandaker Akramul Haque, *Student Member, IEEE,*
Ana Goulart, Katherine Davis, *Senior Member, IEEE*
Email: sshh2@tamu.edu, shossai@sandia.gov, leen.alhomoud@tamu.edu,
akramwired@tamu.edu , goulart@tamu.edu , katedavis@tamu.edu

***Abstract*—Numerous research studies are being conducted to enhance the resilience of the power grid by detecting potential cyber or physical disturbances on the system. However, the development of effective mitigation techniques and remediation actions for cyber-physical systems (CPS) facing disturbance scenarios is in an early stage. Therefore, this paper focuses on building a framework of scalable cyber-physical optimal response. A review of artificial intelligence methods relevant to the design of the response framework is conducted. Then, an artificial intelligence method based on controller sensitivities is presented and initial results are discussed for a 9-bus system to motivate its use in improving AI-based intrusion response.**

***Index Terms*—artificial intelligence, cyber-physical system, optimal response engine, power system resilience, mitigation and remediation actions**

I. Introduction

The resilience of power system networks and infrastructures are crucial to a nation's economy and public safety. However, despite advancements in technology, power grids remain vulnerable to an increasing number of disturbances, whether cyber or physical. Awareness of the importance of cybersecurity in achieving and maintaining power grid systems' reliability and resilience is increasing. This has resulted in widespread enhancements to the defense of power systems against cyber-originated threats over the past decade, particularly in the research and development of cyber-aware grid planning and monitoring [1].

The challenge is that digitization in power systems continues to grow, while new advanced attack techniques continue to appear, leading to increases in novel cyber disturbances. Disturbances can propagate between systems, highlighting the need for observation of the system as a whole [2], [3]. For example, a Denial of Service (DoS) attack to the Automatic Generation Control (AGC) could cause improper resource allocation or load imbalance [3].

S. Sun, L. Al Homoud, K. Haque, K. Davis, A. Goulart are with Texas A&M University, College Station, TX. S. Hossain-McKenzie is with Sandia National Laboratories, Albuquerque, NM. This work was supported by the US Department of Energy under award DE-CR0000018.

979-8-3503-3120-2/24/$31.00 ©2024 IEEE

Recent research has emphasized false data detection and cyber mitigation. While identifying attacks is crucial, it is equally important to respond to disturbances, take remedial actions to prevent further damage, and restore normalcy. As a result, optimal responses and remediation strategies in cyber-physical systems have gained significant attention in the field of power system resilience and security. In 2023, cyber-informed transmission planning has become North American Electric Reliability Corporation (NERC)'s top work priority [1].

Artificial intelligence (AI) methods are playing increasingly important roles. However, mitigation and restoration are complicated decision-making processes for balancing authorities, individual regulated utility companies, and Independent System Operators (ISO) [4]. The question remains: how can AI enable system operators to improve situational awareness and response capabilities during an incident?

In this paper, we propose a framework for a Scalable Cyber-physical Optimal Response Engine (SCORE), an automatic optimal response engine to ensure resilience of the grid when facing cyber and physical disturbances. The requirements of SCORE are as follows:

- Provide real-time operational guidance for the large-scale power system under different network topologies and multiple disturbance scenarios.
- Provide a physics-informed AI-based approach to facilitate a fast and efficient decision-making process.

Section II presents related work. Recent techniques in deep learning relevant to the design of SCORE are surveyed in Section III. Section IV elaborates the framework of SCORE. Section V provides preliminary results, and Section VI concludes the paper.

II. Related Work

A. Theoretical Approaches

To investigate how to solve the problem of optimal cyber-physical intrusion response in power systems, some initial previous studies [5] were conducted using Markov models, but these studies lacked the use of AI and did not consider the entire end-to-end control loop in the cyber-physical system, which are major innovations addressed in this work. The initial

2024 IEEE Texas Power and Energy Conference (TPEC) | 979-8-3503-3120-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/TPEC60005.2024.10472265

studies in [5] were also small and noted but did not attempt to address the scalability issues. Related work includes proposing a machine learning-based attack classification framework, developing an adaptive special protective scheme (SPS) to mitigate the cyber-physical disturbances, and a case study on Markov Decision Process (MDP) approach [6].

B. Testbed Emulation Approaches

The Resilient Energy Systems Lab (RESLab) testbed replicates the power system cyber-physical environment in research studies. RESLab is mainly composed of a power system simulator, a network emulator, a DNP3 master, a SNORT intrusion detection system, and a data fusion engine [7]. PowerWorld Dynamic Studio (PWDS) serves as a real-time power system simulator that also sends DNP3 outstation packets [8]. For emulating the power system communication and control network, Common Open Research Emulator (CORE) is employed that can connect to other virtual machines (VM) as well as external hardware, to replicate a cyber-physical power system [7]. SNORT serves as an intrusion detection system that sends alerts to the DNP3 master [7]. The data fusion engine is [7] and collects data from Wireshark, using Elasticsearch, as well as SNORT logs. The fusion engine aids in visualization and enables machine learning and AI techniques to efficiently detect cyber-physical attacks [7].

C. Cyber-Physical Energy Management System

The Cyber-Physical Resilient Energy Systems (CYPRES) EMS is presented in [9]. The solution is an end-to-end system that manages the models, visualization, system monitoring and control, physical and cyber alerts, and mitigation. In RESLab, it interacts with CORE and PWDS [9]. The CYPRES EMS is tested on a 2000-bus synthetic cyber-physical power system, in the RESLab architecture described above. This application has an interactive map to visualize and interact with the system in real-time. The EMS allows for monitoring traffic in real-time, critical for cyber-physical threat mitigation and response. A Bayesian inference framework is set up in the EMS to help detect cyber-physical threats and for risk assessment. The EMS employs a tool named CyPSA-Live, which performs an online risk analysis to improve the system's situational awareness.

D. Improved Threat Identification

Cyber attacks pose threats to power systems [10]. A multi-step or multi-stage intrusion involves the intruder's privilege escalation within a network. Details of the taxonomy of attacks can mapped using the MITRE ATT&CK Framework [11], a knowledge base that contains details on different intrusion techniques and attack vectors, with suggested detection and mitigation tools for each. These techniques are available for Industrial Control Systems (ICS), with application to the power grid. Privilege escalation is one of the steps. This type of attack is studied in [12], where the authors develop a numerical cyber-physical security index CPIndex that aids in assessment of cyber-physical vulnerabilities in power systems. Some steps may also include Denial-of-Service (DoS), Man-in-the-Middle (MiTM), and False Data Injection (FDI) [10]. In DoS, the intruder would flood traffic into its target with the goal of shutting down a critical service or process. In MiTM, an intruder compromises the communication between two assets in a system. Lastly, in FDI, an intruder falsifies measurements to mislead an operator or an algorithm.

III. Deep Learning Techniques and Integration into SCORE's Design

In this section, we will investigate state-of-the-art technologies and emerging trends that have the potential to lead to innovative solutions and breakthroughs in response engines.

A. Reinforcement Learning Techniques

With improvements in both computing algorithms and hardware, machine learning techniques such as reinforcement learning and deep learning methods are poised for breakthroughs in policy optimization and cybersecurity [13]. Due to the performance of reinforcement learning (RL) in the control and optimization process [14], it is an optimistic approach for enhancing power system resilience [15]. The RL approach is a category of machine learning algorithms in which an agent learns to make optimal decisions in an environment by interacting with it and receiving feedback in the form of rewards [14]. The agent learns from its experiences and uses this knowledge to make decisions in subsequent interactions [14]. RL algorithms typically involve estimating the action-value or the state-value function, aiming to maximize the rewards that guides the agent's actions.

An MDP is a discrete decision making process in a stochastic environment [16]. Within RL, a variety of algorithms exist, such as State-Action-Reward-State-Action (SARSA), Deep Reinforcement Learning (DRL), Q-learning, etc [17]. DRL combines reinforcement learning with deep neural networks, which serve as function approximators in DRL. By taking advantage of neural networks, DRL can handle complex and high-dimensional state and action spaces [18]. Proximal Policy Optimization(PPO), Advantage Actor-Critic (A2C) and Q-learning are prominent algorithms within the DRL family. PPO aims to optimize policies in RL with improved training stability and sample efficiency, while A2C algorithm balances the dual roles of actor and critic to enhance learning efficiency and policy performance. Moreover, Q-learning algorithms are able to approximate the value function by iteratively updating Bellan equation [17]. The characteristic of Q-learning is a perfect solution for off-policy decision optimization problems [17]. Another RL approach takes advantage of estimating and updating Q-values [17]. The SARSA algorithm acquires the optimal policy by converging to maximum reward values [17]. In Table I, we compare papers applying RL techniques to support the power system optimal solution process during disturbances.

In Table I, the DRL approach is used to learn optimal parameters, vulnerability, and recovery stategy in [19]- [25]. [19] - [22] utilized PPO and A2C algorithms to analyze optimal power flow solution, volt-var control and adversarial

TABLE I
RL ALGORITHMS PRESENTED IN PAPERS

Learning Algorithm		Use Cases	References No.
DRL	PPO	Defense policy-making by learning an adversary to improve security	[19]
		Generating AC optimal power flow solutions in DER	[20]
		Obtaining effective and timely power dispatch policies	[21]
	PPO\ A2C	Volt-var control in distribution system	[22]
	Q-learning	Optimal operation and maintenance management	[23]
		Mitigation strategy caused by cyber attack	[24]
		Generation restoration process	[25]
SARSA		Cascading failure mitigation strategy	[26]

training during cyber events. Q-learning further adds benefits to the optimal recovery solution and mitigation strategies [23]-[25]. [26] proposed the cascading failure mitigation approach applying SARSA. Each of these applications in the four approach is new and has value to offer in response engine design.

B. Meeting Data Fusion and Data Trust Needs for RL

In developing RL or any AI-based defense, a robust design of its data flows and security is essential. Data fusion plays a significant role in enhancing cybersecurity by combining information from different sources, such as physical sensors, control systems, and network logs. It enables (1) a comprehensive view of the system's security posture with more accurate understanding of the system's behaviors, for more effective threat detection and response, (2) anomaly detection and early indicators to minimize impact on the system, (3) correlation of cyber and physical events with system behavior for improved threat intelligence. By leveraging fused data and advanced analytic techniques, organizations can more effectively protect their systems.

C. Including Physics in the AI with Controller Sensitivities

The response engine will use multiple avenues to ensure that physics are included to ensure scalability and accuracy. To achieve this, the distributed controller role and interaction discovery (RID) algorithm, detailed in [27] is considered. The RID algorithm is based on real and reactive power flow sensitivities with respect to control changes. It identifies essential, critical, and redundant controllers for the controllability of a system, as well as control support groups.

- *Essential controllers* are the minimal set of devices required to maintain system controllability.
- *Critical controllers* are essential controllers that are irreplaceable and mandatory for system controllability.
- *Redundant controllers* are the devices that reinforce the control capability of essential controllers.
- *Control support groups* contain devices that are highly coupled in terms of impact on the control objective and with each other.

In [27], the effectiveness of the RID algorithm for reducing the corrective control search space is shown. Additionally, in [28] the RID algorithm reduces power system violations by leveraging different controllers. The RID capability for tackling the dimensionality curse for power system-side corrective actions can be directly applied to SCORE to inform response decisions, as shown in Fig. 1.

However, for cyber-physical systems, where both cyber-physical corrective actions may be needed, no such characterization is available for cyber controls. Furthermore, although efforts are looking into incorporating cyber corrective actions, the strategy for selecting one cyber corrective action over another is not available.

Therefore, to investigate how that characterization and strategy can be achieved, we investigate leveraging the RID algorithm with characterizing cyber corrective actions. Section V-A details this investigation and preliminary results. Specifically, the RID algorithm will help SCORE address the challenge of scalability and directly inform the RL decisions.

IV. GENERAL FRAMEWORK

The proposed framework is shown in Fig. 1, with modules as follows:

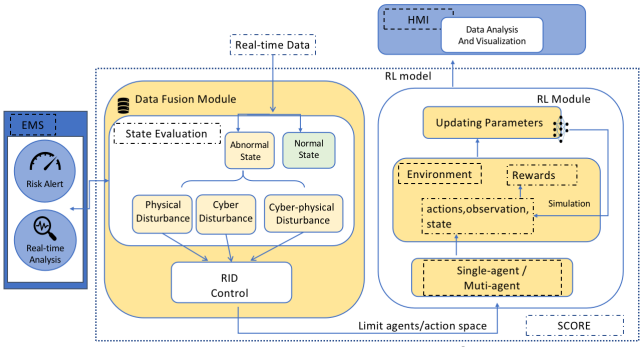

Fig. 1. General Framework of SCORE.

A. Data Collection Module

Real-time data and risk information will send to the *Data Fusion* module from DNP3 master, PowerWorld Dynamic Studio (PWDS) and CORE network. Data fusion techniques will be deployed and used to examine data from multiple sources, such as sensors, historical records, and extract data features.

B. State Evaluation Submodule

The main purpose of the state evaluation module is assessing the overall status of the current power system and determine whether or not the system is facing a disturbance. A comprehensive assessment with the impacts caused by the disturbances will provide system operators with decision support. This classification allows for a more accurate and effective

response to system disruptions, enabling the deployment of targeted mitigation strategies.

C. RID Submodule

A significant challenge that MDP-based models face is the dimension curse [29]. It is important to assess methods to reduce the number of agents and the action spaces by figuring out the most relevant corrective actions. In [6], [30], the roles and groups are determined using the 3-step process in the RID algorithm, which is defined as followed:

1) Obtaining Sensitivity Matrix: The sensitivity matrix $\mathbf{\Psi}$ provides a linearized relationship between control actions and response of the system to the actions.
2) Finding Controllability-Equivalence Sets: The sensitivity matrix rows that show the mutual influence of controls within controllers are clustered to generate the control support groups. Row vectors $\mathbf{v_i}$ and $\mathbf{v_j}$ are compared using the Coupling Index (CI), which is the cosine similarity as shown below.

$$CI = \cos(\theta_{v_i v_j}) = \frac{\mathbf{v_i} \cdot \mathbf{v_j}}{\|\mathbf{v_i}\|\|\mathbf{v_j}\|} \tag{1}$$

3) Finding Critical, Essential and Redundant Sets: The columns of $\mathbf{\Psi}$ are used to identify the critical, essential and redundant controllers in the system, where they are placed, and to create applications for mitigation and response. [31] describes the LU factorization method. The decomposition of $[\mathbf{\Psi}]^{\mathbf{T}}$ is obtained as:

$$[\mathbf{\Psi}]^{\mathbf{T}} = \mathbf{P}^{-1}\mathbf{L_F}\mathbf{U_F} \tag{2}$$

$$\mathbf{L_F} = \begin{bmatrix} \mathbf{L_b} \\ \mathbf{M} \end{bmatrix} \tag{3}$$

Using the Peters-Wilkinson [31] method, $[\mathbf{\Psi}]^{\mathbf{T}}$ is factored, where $\mathbf{P}$ is the permutation matrix and $\mathbf{L_F}$ and $\mathbf{U_F}$ are the lower and upper triangular factors of dimension n, respectively. $\mathbf{M}$ is a sparse, rectangular matrix. The new basis has the structure:

$$\mathbf{L_{CER}} = \mathbf{L_F}\mathbf{L_b^{-1}} = \begin{bmatrix} \mathbf{C_E} \\ \mathbf{C_R} \end{bmatrix} \tag{4}$$

Each accessible controller is represented by a row in the modified matrix . [32]. The identification matrix, or CE, has rows that correlate to key controllers. The redundant controllers are represented by the rows of CR.

D. RL-based Module

A RL-based module is initiated, consisting of three submodules: *Agent*, and *Simulation Environment*.

1) Agent Submodule: While the RID algorithm is applied to select the most relevant actions and agents, limited agent and action spaces are decided by the RID module. By receiving feedback from the environment in form of rewards or penalties by the chosen actions, the agent will finalize and provide the optimal solution [17]. The possible agents would be physical components like generators or lines, or cyber components such as breakers or routers. A collection of actions to be applied will be described in Section IV-D2.

2) Action Spaces for CPS Remediation and Mitigation: From NERC PRC-012-2 [33], remedial action schemes should maintain reliability limits and prevent disturbance propagation. A number of deployed solutions that help grid reliability are detailed in [34]. A related work is HARMONIE-SPS, an adaptive, response solution for cyber-physical disturbances [6]; Disturbances are first classified to three categories: cyber, physical and cyber-physical. Hence, we adopt separate action spaces for different agents, shown in Table II.

TABLE II
POSSIBLE ACTION SPACES FOR CPS DURING DISTURBANCES

Impacts		**Agent**	**Actions**
Impacts on Cyber-physical System	Impacts on Physical System	Generator	Change Real Power
			Change Reactive Power
			Isolate/Shut down
			Change to different buses
			Change Voltage Maganitude
		Tranmission Line	Trip/Close
		Load	Increase/Decrease
			Change to different buses
			Isolate/Shut down
	Impacts on Cyber Network	Router	Connect/Disconnect
		HMI	
		Ethernet Switches	
		Breaker/Relay	Open/Close
		Firewall	Connect/Disconnect
			Reconfigure

3) Simulation Enviornment Submodule: Similar to Grid2op [35], the simulation environment in SCORE must support with power system analysis based on its cyber-physical properties. Power flow and contingency analysis must be incorporated into the environment to enable the performance of RL agents to be tested and evaluated.

E. Human Machine Interface (HMI)

An HMI is applied for visualization and interaction. The responsibilities of HMI include displaying the suggest corrective actions, interacting with the users, and receiving feedback throughout the process. If a solution appears incorrect or unreasonable, the users can intervene and send a negative feedback to the RL model. The rewards or feedback from users further refine the RL model.

The SCORE engine is not intended to be a fully automated machine. Rather, the goal is a decision support system that incorporates model information and human expertise, while safely harnessing the power of artificial intelligence.

V. PRELIMINARY RESULTS: HOW SENSITIVITIES CAN HELP AI-BASED CYBER-PHYSICAL OPTIMAL RESPONSE

In this section, we verify our preliminary results on how RID and RL algorithms can facilitate building SCORE and

the RL-based model. We considered the modified WSCC 9-bus system. [36] with 9 buses, 2 generators and 1 battery storage. Two 4000 kVAR capacitors are added to buses 7 and 9, respectively, as shown in Fig.2. A battery supply takes the place of one of the generator. In this instance, the transformers are regarded as voltage regulators or online tap changing. We assumed Battery 1 facing Denial of Service attacks and causing voltage drop and load imbalanced.

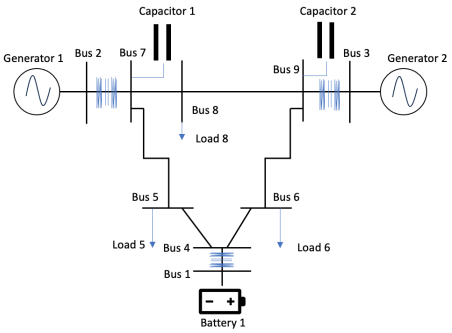

Fig. 2. WSCC 9-bus one-line diagram.

A. RID Algorithm

A challenge that MDP-based models face is the dimensionality curse [29]. Thus, it is important to assess methods to reduce that problem for action spaces.

In this circumstance, we focus on investigating how RID algorithm could apply to determine the capacitors and provide the necessary reactive power support. The first step is to set up the sensitivity matrix. The sensitivities offer an understanding of the relationship between the capacity of available capacitors and buses under overload conditions. We create a sensitivity matrix based on the relationship between the injected reactive power and the resulting change in voltage

$$\Delta V_{\text{bus,overloaded}} = [\mathbf{\Psi}] \cdot \mathbf{\Delta Q}_{\text{MVar}} \tag{5}$$

Subsequently, the RID algorithm is applied to the sensitivity matrix, leveraging LU decomposition. For this case study, both the capacitors are the critical ones. This provides insight into next steps for RL technique.

TABLE III
WSCC-9 BUS RID RESULT

Violations	Critical Controllers
Bus voltage violations	Capacitor 1, 2

B. RL Technique

To keep the voltage within 0.95 - 1.05 p.u., capacitors and generators located at strategic positions are controlled for the optimization of the Volt-Var under voltage profile, power losses and device constraints.

A loss function could be generated as Eq.6:

$$\min_{s.t.} {}_x F_{volt}(x) + F_{ctrl}(x) + F_{power}(x) \tag{6}$$

Two popular policy based RL algorithms namely Proximal Policy Optimization (PPO) and Advantage Actor Critic (A2C) are used for benchmarking [37], [38]. The results of normal state rewards and rewards facing cyber-disturbance can be illustrated by Fig.3 and Fig.4 respectively.

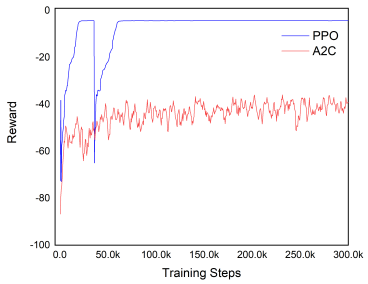

Fig. 3. Reward of RL algorithms in 300k steps in Normal Condition

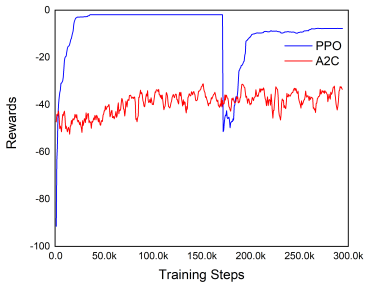

Fig. 4. Reward of RL algorithms in 300k steps in Battery1 Denial of Service Condition

It is evident from comparing Figs 3 and 4 that the scenario with DoS interference has a larger reward at the beginning of the PPO algorithm's training phase. The load profile's random initialization was the cause of this. The load values in the load profile range from 0 to 1. When the condition is substantially loaded, or 1, the voltage value decreases significantly. When the condition is weakly loaded, or 0, the opposite thing occurs. Following an adequate length of training, the load profile's random initialization is thoroughly investigated, and as a result, the reward eventually settles at a lower value than it was during the first training phase. Based on the two figures, we can infer that the PPO algorithm maximizes the reward because it performs better than the A2C method in less steps. After verifying the outcome in the surroundings, Fig. 4 shows that all bus voltages are within the ±5% tolerance. It is demonstrated through validation that an RL-based model could restore both scenarios' typical conditions for the system.

VI. CONCLUSION

A framework of SCORE with RL and RID and a review on supportive techniques are presented. This work introduces an

automatic optimal response framework against cyber-physical disturbances. This continued investigation will directly influence cyber-physical data analysis and metric types as well as provide a cyber corrective action characterization approach. The cyber-physical extension of RID can help SCORE tackle the scalability challenge and directly inform its RL-based decisions with *a-priori* corrective action analysis. In our subsequent research, we aim to extend the application of this methodology across a larger case study.

REFERENCES

[1] NERC and the Six Regional Entities, "ERO enterprise publishes cyber-informed transmission planning white paper," North American Electric Reliability Corporation, Tech. Rep., May 2023.

[2] S. Hossain-McKenzie, N. Jacobs, A. Summers, R. Adams, A. Chatterjee, A. Layton, K. Davis, and H. Huang, "Towards the characterization of cyber-physical system interdependencies in the electric grid," *IEEE Power and Energy Conference at Illinois (PECI)*, 2023.

[3] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2012.

[4] W. Jang, H. Huang, K. R. Davis, and T. J. Overbye, "Considerations in the automatic development of electric grid restoration plans," in *2020 52nd North American Power Symposium (NAPS)*, 2021, pp. 1–6.

[5] A. Sahu, H. Huang, K. Davis, and S. Zonouz, "Score: A security-oriented cyber-physical optimal response engine," in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2019, pp. 1–6.

[6] S. Hossain-McKenzie, N. Jacobs, A. Summers, B. Kolaczkowski, C. Goes, R. Fasano, Z. Mao, L. Al Homoud, K. Davis, and T. Overbye, "Harmonized automatic relay mitigation of nefarious intentional events (HARMONIE)-special protection scheme (SPS)." Sandia National Lab (SNL-NM), Albuquerque, NM (United States), Tech. Rep., 2022.

[7] A. Sahu, Z. Mao, P. Wlazlo, H. Huang, K. Davis, A. Goulart, and S. Zonouz, "Multi-source multi-domain data fusion for cyberattack detection in power systems," *IEEE Access*, vol. 9, pp. 119 118–119 138, 2021.

[8] H. Huang, C. M. Davis, and K. R. Davis, "Real-time power system simulation with hardware devices through DNP3 in cyber-physical testbed," in *2021 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2021, pp. 1–6.

[9] A. Sahu, K. Davis, H. Huang, A. Umunnakwe, S. Zonouz, and A. Goulart, "Design of next-generation cyber-physical energy management systems: Monitoring to mitigation," *IEEE Open Access Journal of Power and Energy*, vol. 10, pp. 151–163, 2023.

[10] F. Li, X. Yan, Y. Xie, Z. Sang, and X. Yuan, "A review of cyber-attack methods in cyber-physical power system," in *2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP)*, 2019, pp. 1335–1339.

[11] MITRE. MITRE ATT&CK framework. [Online]. Available: https://attack.mitre.org

[12] C. Vellaithurai, A. Srivastava, S. Zonouz, and R. Berthier, "CPIndex: Cyber-physical vulnerability assessment for power-grid infrastructures," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 566–575, 2015.

[13] A. Uprety and D. B. Rawat, "Reinforcement learning for IoT security: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8693–8706, 2021.

[14] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[15] Z. Zhao, P.-Y. Chen, and Y. Jin, "Reinforcement learning for resilient power grids," *arXiv preprint arXiv:2212.04069*, 2022.

[16] C. C. White III and D. J. White, "Markov decision processes," *European Journal of Operational Research*, vol. 39, no. 1, pp. 1–16, 1989.

[17] R. S. Sutton and A. G. Barto, "Reinforcement learning: an introduction mit press," *Cambridge, MA*, vol. 22447, 1998.

[18] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, and F. Blaabjerg, "Reinforcement learning and its applications in modern power and energy systems: A review," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1029–1042, 2020.

[19] A. Pan, Y. Lee, H. Zhang, Y. Chen, and Y. Shi, "Improving robustness of reinforcement learning for power system control with adversarial training," 2021.

[20] Y. Zhou, W. Lee, R. Diao, and D. Shi, "Deep reinforcement learning based real-time ac optimal power flow considering uncertainties," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 5, pp. 1098–1109, 2022.

[21] Y. Zhao, J. Liu, X. Liu, K. Yuan, K. Ren, and M. Yang, "A graph-based deep reinforcement learning framework for autonomous power dispatch on power systems with changing topologies," in *2022 IEEE Sustainable Power and Energy Conference (iSPEC)*, 2022, pp. 1–5.

[22] T.-H. Fan, X. Y. Lee, and Y. Wang, "Powergym: A reinforcement learning environment for volt-var control in power distribution systems," in *Learning for Dynamics and Control Conference*. PMLR, 2022, pp. 21–33.

[23] R. Rocchetta, L. Bellani, M. Compare, E. Zio, and E. Patelli, "A reinforcement learning framework for optimal operation and maintenance of power grids," *Applied Energy*, vol. 241, pp. 291–301, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261919304222

[24] J. Appiah-Kubi and C.-C. Liu, "Cyberattack correlation and mitigation for distribution systems via machine learning," *IEEE Open Access Journal of Power and Energy*, vol. 10, pp. 128–140, 2023.

[25] H. Li, X. Yang, and C. Zhai, "A reinforcement learning approach for robust restoration of generators in power system," in *2022 China Automation Congress (CAC)*, 2022, pp. 1651–1656.

[26] Y. Zhu and C. Liu, "Mitigating multi-stage cascading failure by reinforcement learning," in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2019, pp. 3724–3728.

[27] S. Hossain-McKenzie, E. Vugrin, and K. Davis, "Enabling online, dynamic remedial action schemes by reducing the corrective control search space," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2020, pp. 1–6.

[28] S. Hossain-McKenzie, K. Raghunath, K. Davis, S. Etigowni, and S. Zonouz, "Strategy for distributed controller defence: Leveraging controller roles and control support groups to maintain or regain control in cyber-adversarial power systems," *IET Cyber-Physical Systems: Theory & Applications*, vol. 6, no. 2, pp. 80–92, 2021.

[29] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: reinforcement learning framework," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 427–435, 2004.

[30] S. Hossain-McKenzie, E. Vugrin, and K. Davis, "Enabling online, dynamic remedial action schemes by reducing the corrective control search space," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2020, pp. 1–6.

[31] G. Peters and J. H. Wilkinson, "The least squares problem and pseudo-inverses," *The Computer Journal*, vol. 13, no. 3, pp. 309–316, 1970.

[32] J. Chen and A. Abur, "Placement of PMUs to enable bad data detection in state estimation," *Power Systems, IEEE Transactions on*, vol. 21, no. 4, pp. 1608–1615, Nov. 2006.

[33] NERC, "PRC-012-2 – Remedial Action Schemes (RAS)(PRC-012-2)," Tech. Rep., 2016.

[34] J. Giri, "Real-time grid management: Keeping the lights on!" *IEEE Power and Energy Magazine*, vol. 21, no. 3, pp. 51–60, 2023.

[35] B. Donnot, "Grid2op- A testbed platform to model sequential decision making in power systems. ," https://GitHub.com/rte-france/grid2op, 2020.

[36] A. S. Al-Hinai, "Voltage collapse prediction for interconnected power systems," *Graduate Theses, Dissertations*, vol. 1065, 2000. [Online]. Available: https://researchrepository.wvu.edu/etd/1065/

[37] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[38] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*. PMLR, 2016, pp. 1928–1937.