Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 45 (2024), No. 8, pp. 2299–2307
DOI : 10.47974/JIOS-1807

The role of deep reinforcement learning in developing adaptive cybersecurity defenses for smart grid systems

Hatoon S. AlSagri [§]
Information Systems Department
College of Computer and Information Sciences
Imam Mohammad Ibn Saud Islamic University (IMSIU)
Riyadh 11432
Saudi Arabia

Shahab Saquib Sohail [†]
School of Computing Science and Engineering
VIT Bhopal University
Kothrikalan
Sehore 466114
Madhya Pradesh
India

Shiju Sebastian *
School of Business and Management
CHRIST University
Ghaziabad 201003
Uttar Pradesh
India

Abstract

The integration of profound fortification learning procedures inside cybersecurity techniques has risen as a promising approach to invigorate the flexibility of savvy network frameworks against advancing dangers. Savvy lattice frameworks, with their complex organize of interconnected gadgets and basic foundation, show special challenges for conventional cybersecurity measures. In this setting, DRL offers energetic and versatile arrangement by leveraging its capacity to memorize from involvement and optimize decision-making in complex, energetic situations. This paper looks at the part of DRL in

[§] *E-mail:* **hsagri@imamu.edu.sa**
[†] *E-mail:* **shahabsaquibsohail@vitbhopal.ac.in**
* *E-mail:* **shijusebastian@christuniversity.in (Corresponding Author)**

© Taru Publications

creating versatile cybersecurity guards custom-made particularly for shrewd network frameworks. By utilizing DRL calculations, such as profound Q-networks (DQN) and profound deterministic arrangement slope, shrewd lattice protections can independently adjust to changing risk scenes, distinguish peculiarities, and moderate assaults in real-time. Moreover, DRL empowers the creation of proactive resistance instruments that can expect potential dangers and preemptively alter security conventions. Through a comprehensive survey of existing writing and case thinks about, this paper highlights the adequacy of DRL in improving the vigor and flexibility of cybersecurity measures for keen lattice frameworks. Additionally, it investigates the challenges and openings related with the integration of DRL methods in viable cybersecurity systems, clearing the way for future inquire about and usage in securing basic framework against developing cyber dangers.

Subject Classification: *68Q10.*

Keywords: *Deep reinforcement learning (DRL), Smart grid systems, Critical infrastructure protection, Dynamic environments, Deep Q-Networks (DQN).*

1. Introduction

Shrewd network frameworks are exceptionally vital for upgrading and moving forward vitality conveyance systems in a time when computers and the web are getting to be increasingly common [1]. In any case, shrewd networks are simple targets for hacking since so numerous contraptions are presently connected to each other and depend on computerized lines of communication. When programmers get into these imperative frameworks, the comes about can be exceptionally awful, from benefit interferences to conceivable hurt that has impacts on society as entirety [2]. To secure the steadfastness, soundness, and security of our vitality framework, we must secure savvy framework frameworks against unused online threats. Indeed, although a few conventional strategies to cybersecurity work, they frequently can't keep up with how rapidly and successfully modern cyber dangers alter [3].

Deep Reinforcement Learning (DRL) is based on the idea that operators learn how to form choices in a certain arrange by expanding the whole sum of benefits they get over time [4]. Since DRL can naturally learn from botches and make the best decisions in complicated, changing settings, it may be a great choice for managing with the issues that uncertainty in keen networks brings. Keen framework security can proceed to alter and respond to modern dangers in genuine time by utilizing DRL calculations like Profound Q-Networks (DQN) and Profound Deterministic Approach Angle (DDPG). The purpose of this consider is to see into how DRL can be utilized to form adaptable cyber defenses for shrewd network

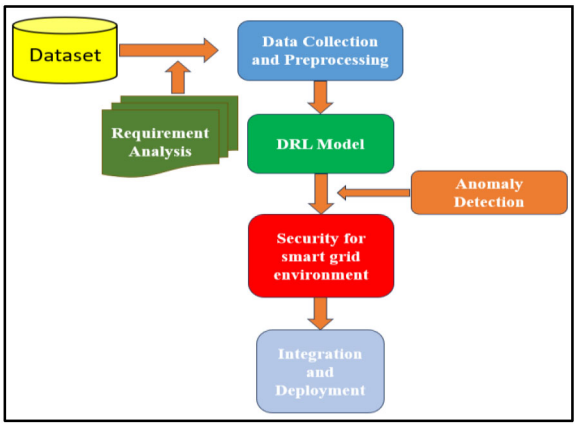


Figure 1

Block Diagram of Proposed Model

frameworks [5]. We'll see at a parcel of current writing, case thinks about, and modern patterns to find out how DRL can be utilized to form assurance strategies more dependable and viable in protecting key vitality frameworks

2. Methodology Used

A few steps are taken with the gotten information amid the information gathering and planning stage. To begin, clamor is taken out to create beyond any doubt the quality of the information. At that point, strategies like estimation or expulsion are utilized to bargain with purge numbers. Normalizing highlights keeps things the same over components. Moreover, strict steps are taken to ensure information security and take after rules like GDPR or HIPAA. To keep private data secure, encryption, anonymization, and get to control are utilized [6]. Taking these steps together makes sure that the information being dealt with remains correct, private, and in line with the rules amid the total prepare of building a hacking defense for savvy framework (figure 1).

3. DRL Model

Within the show creation stage, Profound Q-Networks (DQN) are made to work with savvy lattice systems' adaptable hacking security. Fortification learning operators are planned to remember what they've learned within the past. They do this by utilizing strategies like encounter

rehash to memorize from the distinctive sorts of circumstances that come up in cybersecurity operations [7]. The application includes methods for finding strange things, which use DQN's ability to find trends in smart grid data streams that point to bad behavior [8]. By interacting with the world all the time, the bots learn more about what is normal and what isn't, which lets them spot threats before they happen. Also, the DQN-based models make it easier to find threats and coordinate responses in real time (equation 1 and 2), so protection tactics can change as cyber threats change.

Algorithm:

Step 1: Initialization:
- Initialize Q-function Q(s, a; θ) with parameters θ.
- Set hyperparameters: α, γ, ε, N.

Step 2: Data Collection:
- Collect data: (s, a, r, s').
- Preprocess data.

Step 3: Action Selection:
- Choose action a:

a = { random with probability ε
argmax_a Q (s, a; θ) otherwise }

Step 4: Execute Action and Observe Reward:
- Execute a in smart grid.
- Observe s' and r.

Step 5: Replay Buffer Management:
- Store (s, a, r, s') in replay buffer D.

Step 6: Experience Replay:
- Sample batch B from D.
- Compute target:

$$target = r + \gamma \cdot maxa'Q(s', a'; \theta-) \quad (1)$$

Step 7: Update Q-Network:
- Update θ by minimizing loss:

$$L(\theta) = (1/|B|)\sum(s, a, r, s') \in B(Q(s, a; \theta) - target)2 \quad (2)$$

Step 8: Exploration Rate Decay:
- Decay ε.

4. Security for Smart Grid Environment

In smart grid settings, the Isolation Forest method can be used in several different ways to find anomalies, threats, and coordinate responses [9]. At first, it builds a forest of isolation trees by randomly splitting the data into groups, with fewer groups for anomalies [10]. The average length of the road needed to separate data points is then used to calculate anomaly scores. A cutoff is used to tell the difference between anomalies and normal cases. When strange things are found, they are analyzed to find possible cyber risks by matching trends with known attack signs. After that, the right reactions are set up to stop risks, like separating the damaged parts or sending out alerts. Isolation Forest for Security is as follows as per algorithm and equation 3 and 5.

Algorithm:

1. Initialization:
   - Set hyperparameters: n_trees, sub_sample, max_depth.
2. Data Preprocessing:
   - Normalize features:

$$x' = \frac{x - \mu}{\sigma} \tag{3}$$

3. Training:
   - Construct Isolation Forest:
   - Randomly select subsets of data points and features.
   - For each tree:
   - Randomly select feature f and split value s.
4. Anomaly Score Calculation:
   - Compute anomaly score:

$$s(x) = 2^{-\frac{E(h(x))}{c(n)}} \tag{4}$$

5. Threshold Determination:
   - Define threshold: $Threshold = q_{th(Anomaly\ Scores)}$.
6. Anomaly Detection:
   - Identify anomalies:

$$Anomaly(x) = True\ if\ s(x) > Threshold \quad (5)$$

7. Threat Identification:
   - Analyze characteristics of detected anomalies.
8. Response Orchestration:
   - Orchestrate appropriate responses to mitigate threats.

5. Performance Analysis

The graph shows, in figure 2, how well the DQN algorithm did on different rating criteria. Accuracy, Precision, Recall, F1-score, and AUC are some of the metrics shown on the x-axis. On the y-axis are the numbers for each measure, shown as a percentage (%).

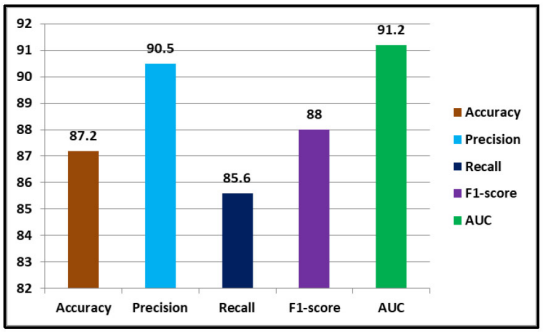


Figure 2

Performance evaluation of DQN

The figure 3 shows the algorithm can make correct positive guesses and classify things correctly overall. In the same way, both the recall and F1-score numbers are high, which suggests that true positive cases are being found effectively. The AUC score, which shows how well the algorithm can tell the difference between classes, shows impressive results, proving that it works well for classification tasks.

The figure 4 shows what happened when the Isolation Forest algorithm was used to find strange things in a smart grid setting. With a level of accuracy of 92.5%, the program does a great job of identifying cases. With a precision of 89.3%, the program is able to make as few fake positive guesses as possible while finding as many true positives as possible. According to the program, a high recall rate of 94.8% means that it correctly identified a lot of true positive cases.

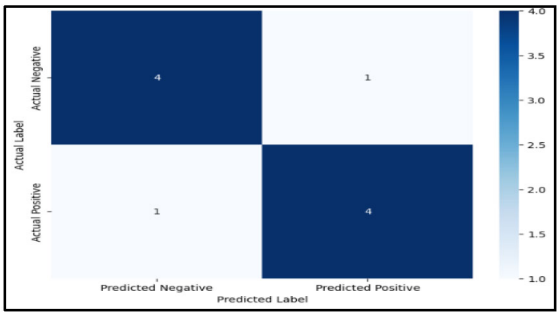


Figure 3

Representation of Confusion Matrix of DQN

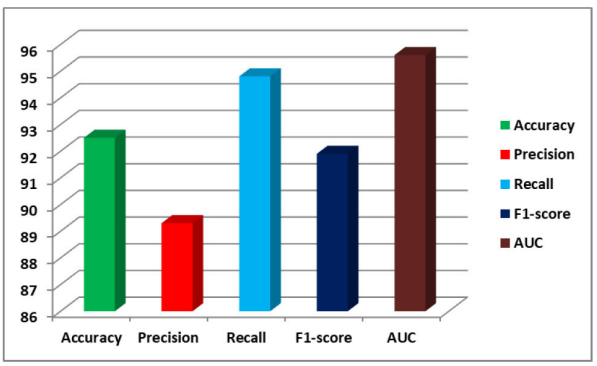


Figure 4

Representation of Performance metrics of Isolation Forest

The F1-score, which measures accuracy and memory, is 91.9%, which means that the person did well in classification tasks, shown in figure 4. The AUC number of 0.956 also shows that the algorithm is very good at telling the difference between regular and abnormal situations, which shows how well it works at finding problems in the smart grid system.

A confusion matrix, shown in figure 3, shows how well a classification method works by showing the difference between the real and projected class names. The rating measures for the DQN and Isolation Forest algorithms are shown in the bar graph. Each bar shows a different measure, like Accuracy, Precision, Recall, F1-score, or AUC, and its height shows the score in percentage (%). Isolation Forest usually does better than DQN in most tests, showing better results in Accuracy, Recall, and AUC, comparison shown in figure 5. DQN, on the other hand, has a slightly higher Precision

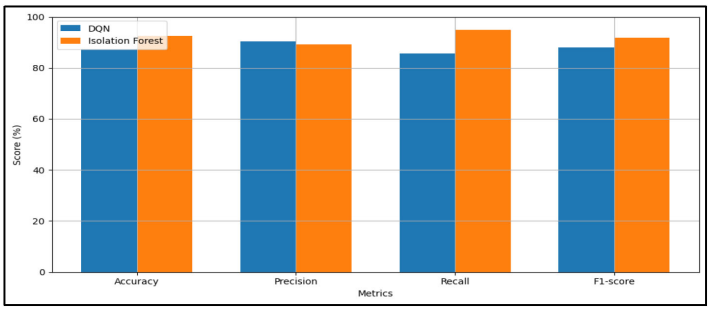


Figure 5

Representation of performance metric on DQN and Isolation Forest using bar graph

and F1-score. You can see how well each program finds anomalies by comparing them on the graph.

4. Conclusion

In conclusion, Deep Reinforcement Learning (DRL) looks like a good way to make smart grid systems safer from cyberattacks. With the help of DRL algorithms like Deep Q-Networks (DQN), smart grids can react instantly and change to new online dangers. DRL models can learn from past mistakes, which lets them find and stop threats before they happen. This makes smart grid systems more resistant to cyberattacks. As long as smart grids are an important part of modern energy delivery networks, adding DRL-based adaptable hacking protection could be a huge step toward protecting important assets and making sure the energy supply is safe and reliable.

Acknowledgment

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-RP23115).

Funding Statement

The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) for funding this work through (grant number IMSIU-RP23115).

References

[1] H. Dong and I. Kotenko, "Hybrid multi-task deep learning for improved IoT network intrusion detection: Exploring different CNN structures," in *Proc. 2024 16th Int. Conf. COMmunication Systems & NETworkS (COMSNETS)*, Bengaluru, India, pp. 7–12 (2024).

[2] I. Kotenko and I. Saenko, "Applying machine learning methods to detect abnormal user behavior in a university data center," in *Intelligent Distributed Computing XV*, vol. 1089, Cham, Switzerland: Springer (2023).

[3] D. Gaifulina and I. Kotenko, "Selection of deep neural network models for IoT anomaly detection experiments," in *Proc. 29th Euromicro Int. Conf. Parallel Distrib. Network-Based Process. (PDP)*, pp. 260–265 (2021).

[4] E. Mushtaq, A. Zameer, M. Umer, and A. A. Abbasi, "A two-stage intrusion detection system with auto-encoder and LSTMs," *Appl. Soft Comput.*, vol. 121, p. 108768 (2022).

[5] J. Mills, J. Hu, and G. Min, "Multi-task federated learning for personalised deep neural networks in edge computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 3, pp. 630–641, Mar. (2022).

[6] S. Kaushik, R. C. Poonia, and S. K. Khatri, "Cryptographic key distribution using artificial intelligence for data security and location privacy in VANET," *J. Discrete Math. Sci. Cryptogr.*, vol. 25, no. 7, pp. 2195–2203 (2022).

[7] X. Lu, P. Shen, Y. Tsao, and H. Kawai, "A pseudo-task design in multi-task learning deep neural network for speaker recognition," in *Proc. 2016 10th Int. Symp. Chinese Spoken Lang. Process. (ISCSLP)*, Tianjin, China, pp. 1–5 (2016).

[8] N. V. Varghese and Q. H. Mahmoud, "Optimization of deep reinforcement learning with hybrid multi-task learning," in *Proc. 2021 IEEE Int. Syst. Conf. (SysCon)*, Vancouver, BC, Canada, pp. 1–8 (2021).

[9] B. Pragathi, D. K. Nayak, and R. C. Poonia, "Solar PV system-frequency adaptive disturbance observer control algorithm for compensating harmonics and frequency disturbances," *J. Interdiscip. Math.*, vol. 23, no. 2, pp. 573–581 (2020).

[10] M. Hessel, H. Soyer, L. Espeholt, W. Czarnecki, S. Schmitt, and H. van Hasselt, "Multi-task deep reinforcement learning with popart," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 3796–3803 (2019).

Received July, 2024