NOTE FOR NATIONAL DEFENCE:
Public and Defence Policy Challenges and Innovations on Artificial Intelligence, Autonomous Systems, and Cybersecurity
Part 5: Applications, Challenges, and Ethical Behaviors on Reinforcement and Deep Reinforcement Learning Algorithms

Authors: Neshat Elhami Fard [1], Rastko R. Selmic [2], Khashayar Khorasani [3]

[1] Graduate student, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada
[2] Professor, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada; rastko.selmic@concordia.ca
[3] Professor, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada; kash@ece.concordia.ca

Summary

- Reinforcement learning (RL) algorithms as a subset of Artificial Intelligence (AI) encounters several challenges in Cyber-Physical systems, which are addressed in this report.
- Several deep reinforcement learning (DRL) applications in cyber-security were introduced in the previous report. The rest will be briefly presented and explained in this report.
- Since RL and DRL algorithms can solve a series of problems in complex real-world environments that people are incapable of explaining accurately, this can have significant societal, ethical, and governance implications. A number of these consequences are described in the subsection of "insight into RL and DRL policy and ethical behaviors."
- Some solutions related to the described consequences are explained in the last section.

CONTEXT

Challenges of Reinforcement learning (RL) Algorithms in Cyber-Physical System (CPS):

Using RL Algorithms, CPS, both in cyber-attack and cyber-defence, is hampered by challenges similar to any other method. The following are the most significant challenges to consider:

- Although many RL and deep reinforcement learning (DRL) algorithms present superior performance, hyper-parameters and reward function design highly influence their performance. As a result, there is a great deal of uncertainty in training the controller [1].

- Many RL and DRL algorithms operate based on Markov decision processes (MDP), which are fully observable in terms of the environment. For CPS, in particular, the existence of an observable environment is false due to inherent uncertainties in states. Therefore, modeling this type of system is achieved by utilizing a partially observable Markov decision process (POMDP) [1].

Other Applications of DRL Algorithms in Cyber-security:

- Generating a virtualized smart city network resource allocation that aims to assign virtual resources to a specific user, optimally, using a *double dueling Deep Q-Networks (DDQN)* RL algorithm [2].

- Using an *Asynchronous Advantage Actor-Critic (A3C)* RL algorithm, create a mobile edge caching to maximize offloading traffic [3].

- Robustness-guided falsification of CPS can be done to find false inputs in these systems by *integrating double DQN and A3C* RL algorithms [4].

- Enhancing the robustness of the autonomous system against adversarial attacks to recognize corrupted measurements and decrease the effects of adversarial errors using *trust region policy optimization (TRPO)* RL algorithm [5].

- Producing a secure offloading in mobile edge caching that aims to learn a policy for a mobile device to securely offload data to edge nodes against jamming and intelligent attacks using *DQN RL algorithm with hot booting transfer learning methods* [6].

- Creating an anti-jamming communication scheme for cognitive radio networks (CRN) in order to derive an optimal frequency hopping policy for CRN secondary users (SU) to overcome intelligent jammers based on a frequency-spatial anti-jamming game using *DQN RL algorithm including convolutional neural network (CNN)* [7].

- The development of other anti-jamming communication technologies to expand the previous item. In this scenario, jammers can use various channel-slot-based architectures, including the *recursive CNN algorithm and DQN* RL algorithm [7], [8].

- A spoofing detection method is applied in wireless networks to determine the optimal authentication threshold using a *Q-learning and Dyna-Q* RL algorithm [9], [10].

- *Hot booting Q-learning and DQN* RL algorithms are implemented as part of the development of mobile offloading for cloud-based malware detection, which will enhance detection accuracy and speed [11], [12].

- By developing a protected mobile crowd-sensing (MCS) system using the *DQN* RL algorithm, the payment policy is optimized to enhance the sensing performance facing faked sensing threats by forming a Stackelberg game [13].

- Using the *DQN* RL algorithm, an automated URL-based phishing detection is produced that identifies malicious websites (URLs) [14].

Insights into RL and DRL Policy and Ethical Behaviors:

Since RL and DRL can solve a series of problems in complex real-world environments that people are incapable of explaining accurately, this can have significant societal, ethical, and governance implications [15]. A number of these consequences are:

Human oversight:

Due to the statements of ``European Commission's High-Level Expert Group on AI," human oversight of AI systems is necessary for supporting human autonomy and decision-making. This statement that is published as ``Ethics Guidelines for Trustworthy AI" in 2019, constitutes a significant obstacle for RL and DRL applications, which are intended to enhance machines' autonomy. A designed RL and DRL system must constantly make many small decisions that it is impossible for humans to follow, monitor, and review all of these decisions or change them. Therefore, it is impossible for a human being to detect all errors and intervene in the RL and DRL system in real-time. However, for the off-line system, human oversight may be beneficial. Furthermore, RL and DRL systems may proceed to learn and adjust their policy to have a robust behavior, perhaps at a speed that is too rapid for humans to keep track of and consequently supervise meaningfully. After the above issues, the European Commission in 2020 announced that ``the appropriate type and degree of human oversight may vary from one case to another [15], [16]."

Safety and reliability:

Communication with the environment allows RL and DRL, both subsets of ML algorithms, to learn. However, learning can be a right or a wrong process. For CPS, the critical issue in using RL and DRL algorithms is the security of the system and the proper learning that leads to safe learning [15].

- **Harms from reward function design:**

In designing reward functions for RL and DRL algorithms to utilize in various applications, it is essential to consider the values of all individuals affected by these algorithms because optimizing for the incorrect purpose can have unintended consequences that are difficult to foresee. In other words, definitions of inappropriate reward functions have irreversible effects [15].

- **Automation and the future of work:**

Due to the advances in RL and DRL algorithms, there will probably be an increase in the tendency of jobs to automation, particularly in CPS fields, including robotics and manufacturing. Therefore, this issue has to be considered when analyzing the impact of AI on the labor market [15].

CONSIDERATIONS

To solve the above-discussed problems, the authors of [15] have raised the cases of

- "Finding ways to track progress in DRL and its applications,"
- "Considering the implications of DRL progress for existing AI governance initiatives, including standards and regulation,"
- "Establishing notions of responsible DRL development."

NEXT STEPS

The next report will examine cyber-defence and cyber-attack policy challenges on RL and DRL algorithms. Moreover, different policies in the production and use of AI in the military of various countries are examined, which can be attributed to RL algorithms to a large extent. Therefore, "AI in the Military" for strategic countries, e.g., China, Russia, The United Kingdom, France, Israel, South Korea, will be studied. Afterwards, the regulation of AI in the Americas and the Caribbean countries, e.g., Brazil, Canada, Jamaica, Mexico, United States, will be investigated.

References

[1] A. Balakrishnanand, and J. Deshmukh, "Reinforcement learning for cyber-physical systems," https://viterbi-web.usc.edu/jdeshmuk/teaching/cs599-autocps-spring-2019/rl.pdf, March 2019.
[2] Y. He, F. R. Yu, N. Zhao, V. C. Leung, and H. Yin, "Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach," IEEE Communications Magazine, vol. 55, no. 12, pp. 31–37, 2017.
[3] H. Zhu, Y. Cao, W. Wang, T. Jiang, and S. Jin, "Deep reinforcement learning for mobile edge caching: Review, new features, and open issues," IEEE Network, vol. 32, no. 6, pp. 50–57, 2018.

[4] T. Akazaki, S. Liu, Y. Yamagata, Y. Duan, and J. Hao, “Falsification of cyber-physical systems using deep reinforcement learning,” in International Symposium on Formal Methods. Springer, 2018, pp. 456–465.
[5] A. Gupta and Z. Yang, “Adversarial reinforcement learning for observer design in autonomous systems under cyber attacks,” arXiv preprint arXiv: 1809.06784, 2018.
[6] L. Xiao, X. Wan, C. Dai, X. Du, X. Chen, and M. Guizani, “Security in mobile edge caching with reinforcement learning,” IEEE Wireless Communications, vol. 25, no. 3, pp. 116–122, 2018.
[7] G. Han, L. Xiao, and H. V. Poor, “Two-dimensional anti-jamming communication based on deep reinforcement learning,” in 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2017, pp. 2087–2091.
[8] X. Liu, Y. Xu, L. Jia, Q. Wu, and A. Anpalagan, “Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach,” IEEE Communications Letters, vol. 22, no. 5, pp. 998–1001,2018.
[9] L. Xiao, Y. Li, G. Liu, Q. Li, and W. Zhuang, “Spoofing detection with reinforcement learning in wireless networks,” in2015 IEEE Global Communications Conference (GLOBECOM). IEEE, 2015, pp. 1–5.
[10] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, “Phy-layer spoofing detection with reinforcement learning in wireless networks,” IEEE Transactions on Vehicular Technology, vol. 65, no. 12, pp. 10037–10047, 2016.
[11] X. Wan, G. Sheng, Y. Li, L. Xiao, and X. Du, “Reinforcement learning based mobile offloading for cloud-based malware detection,” in GLOBECOM 2017- 2017 IEEE Global Communications Conference. IEEE, 2017, pp. 1–6.
[12] Y. Li, J. Liu, Q. Li, and L. Xiao, “Mobile cloud offloading for malware detections with learning,” in 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2015, pp.197–201.
[13] L. Xiao, Y. Li, G. Han, H. Dai, and H. V. Poor, “A secure mobile crowd-sensing game with deep reinforcement learning,” IEEE Transactions on Information Forensics and Security, vol. 13, no. 1, pp. 35–47, 2017.
[14] M. Chatterjee, and A. -S. Namin, “Detecting phishing websites through deep reinforcement learning,” in 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), vol. 2. IEEE, 2019, pp. 227–232.
[15] J. Whittlestone, K. Arulkumaran, and M. Crosby, “The societal implications of deep reinforcement learning,” Journal of Artificial Intelligence Research, vol. 70, pp. 1003–1030, 2021.
[16] M. MacCarthy, and K. Propp, “The EU’s white paper on AI: A thoughtful and balanced way forward,” https://www.lawfareblog.com/eus-white-paper-ai-thoughtful-and-balanced-way-forward, March 2020.