New Jersey Institute of Technology

ENGL 102 Introduction to Research Writing

Deep Reinforcement Learning: The Future of Cybersecurity

Dennis Perdomo

As the world constantly evolves into a digital world, technology has become essential to the daily lives of billions around the globe. Records are kept in digital storage, phone calls are made daily, and the internet is used as a crossroads to share ideas, culture, and community with others. While technology has transformed the world, keeping private data safe is a considerable risk. Over the past few years, the damaging trend shows that cyber breaches keep increasing in volume and potency, resulting in devastating amounts of stolen information, impacting billions of records worldwide. One prominent example of the danger of cyber attacks is the 2017 Equifax data breach that saw the personal information of 146.6 million Americans leaked, leading to Equifax paying up to $425 million to help people affected by the data breach (FTC, 2024). Hackers and cyber criminals have adapted modern technology to break cybersecurity systems, and an innovative technique is needed to combat the growing threat of cyber attacks. One such innovative technique is deep reinforcement learning (DRL). DRL is a branch of artificial intelligence (AI) that combines the human-like neural networks of deep learning with the trial-and-error-based reinforcement learning, providing a software that can accurately adapt to different environments and become more efficient over time. Think of DRL as teaching a computer to learn from experience. DRL is the future of cybersecurity, as the revolutionary technology can provide machine-like precision in response to cyber crimes while adapting to emerging trends.

Using DRL methods, intrusion detection systems (IDSs) can outperform traditional methods in detecting intrusion and classifying malware. One key development within DRLs is Deep-Q-Networks (DQN), which applies deep learning to Q-learning. Q-learning is a reinforcement learning algorithm that “directly learns the optimal action-value function (or Q function) from [observing trends]” (Shah & Xie, 2018). Essentially, by trial and error, Q-learning

teaches the system which steps earn the biggest payoff, like training a dog with treats. DQNs and subsequently, Double Deep-Q-Networks (DDQN), use neural networks to process the results from Q-learning. DDQN sets itself apart from DQN by using two networks, one to choose the best course of action and the other to evaluate how good the action is. In simple terms, DQNs act similarly to human brains, allowing for computational programs to use reasoning and logic to choose the action that would yield the best reward. Sometimes DQNs develop a bias and become overly optimistic about a decision, even if it may not be optimal.

DDQNs allow the programs to get a second opinion and see the problem from different angles. DDQNs improve stability over DQNs and reduce overstimulation. Experimental results show that in a realistic environment, DDQNs are competitive with or outperform traditional machine learning (ML) models (e.g. Random Forest, Decision Trees…) in detection performance, measured by accuracy, precision, recall, and F1(the balanced measure of precision and recall) (Lopez-Martin et al., 2019, p.13). When detecting cyber attacks, the DDQN model scored 89.8% in accuracy, 91.2% in F1, 89.4% in precision, and 93% in recall. Compared to other cybersecurity methods already in practice, the DDQN received the highest score of accuracy and F1, and came second only to the statistics-based Gaussian Naive Bayes model in recall. DDQNs can efficiently learn how to adapt to a constantly changing environment, which is crucial when dealing with cyber breaches. In 2020, researchers concluded that DDQN's algorithm showed higher convergence, requiring 20,000 iterations to reach the optimal policy, compared to the Q-learning algorithms' more than 65,000 iterations, proving that, "the DDQN algorithm can efficiently learn how to optimally control the traffic flow matching in a highly dynamic environment, which is required in practice" (Phan et al., 2020, p.1357). DDQN's higher convergence means the algorithm can learn stably and smoothly, adding stability in high-stakes

situations. One such situation could be a busy hospital's intensive care unit, where dozens of life-support devices rely on network stability. If an attacker infiltrates that network, they could manipulate a ventilator's settings or shut down critical alarms. The DDQN can promptly converge to a strategy that maximizes patient safety while minimizing downtime by avoiding false positives. In this case, the DDQN could be a life-saving upgrade over simpler cybersecurity systems. Through DDQNs, DRL has proven a viable option for preventing cyber breaches. It provides more accurate, precise results without sacrificing control.

DRL algorithms offer security and peace of mind, demonstrating remarkable robustness in dynamic environments and leading to higher performance. One key issue of integrating AI methods into cybersecurity is lower performance due to less optimal algorithm setups. Such setups lead to network failures and the loss of privacy for valuable data. DRL systems can change that narrative; the results from DRL experiments offer a promising glimpse into the future of cybersecurity. Proximal policy optimization (PPO) is a DRL algorithm that learns by building upon its policy steadily, ensuring natural progression without taking risks that could leave the IDS vulnerable. In an article published in the National Library of Medicine, PPO competed against other algorithms implemented in the Stable-Baselines Python library, a "set of improved implementations of Reinforcement Learning algorithms based on OpenAI Baselines" (Hill, 2021). When placed into a randomized environment, PPO was the "only agent that [circumvented] the obstacles and [reached] the goal… [testifying to] PPO's strong exploration strategy and robust trust-region-based policy update strategy" (Larsen et al., 2021, p. 12). While other algorithms tried to process their surroundings and perform adequately, PPO was the only successful model. PPO is more efficient and better-equipped to handle complex situations than other standard IDS practices. The same experiment also yielded positive results regarding time

efficiency, stating that the behavioral analysis revealed “PPO’s superior path adherence capability without sacrificing collision avoidance or time efficiency” (Larsen et al., 2021, p. 15). In a world where time is of the essence, it is imperative that IDSs provide secure defence while moving efficiently. In a similar paper, a PPO agent was matched against a random action defender and a rules-based defender developed by human cyber experts within a simple denial of service attack scenario (like when an attacker floods a web server with thousands of bogus HTTP requests per second, saturating the server’s CPU and network links). Results demonstrated clear evidence of the strengths of DRL cyber defenders, showing that the PPO agent's average episode reward “outperformed the rules-based agent in ~50% of cases, and the random agent in 100% of cases” (Miles et al., 2023, p. 4). The PPO agent quickly achieved near-perfect performance after upgrading the environment’s defensive action space. The study also showed that the PPO agent “adapted to an unintentional network misconfiguration, which is a plausible risk that was missed by the rules-based agent” (Miles et al., 2023, p. 4). Aside from merely providing stability, PPO agents can adapt to attacks and unique scenarios in real time, providing extra security compared to the rules- and random-based agents. Flexibility and adaptability are crucial as hackers become more creative as technology advances. DRL is the next step in cybersecurity because it can provide secure protection while ensuring stability in complex situations.

Critics of DRL systems claim that the models are too reliant on the necessity of already existing data to learn from, which can lead to less effective solutions to rare cyber attacks. However, by combining DRL with modern technologies, IDSs can offer a much stronger and resilient defence against attacks. By integrating DRL techniques into adaptive cybersecurity defense, network systems will be more efficient in recognizing and responding to threats. DRL techniques are dynamic, allowing compatibility with other cutting-edge technologies to combat

cyber breaches better. One such technology is generative adversarial networks (GANs). GANs are part of the general deep learning branch of AI. GANs behave as a two-player game, where the Generator keeps trying to deceive the Discriminator by producing fake results. At the same time, the Discriminator determines whether or not the sample is genuine. DRL networks can work with GANs, yielding better performance in ML defense systems against real-world cyber breaches. In 2024, the Department of Computer Science at the University of Western Ontario noticed that IDSs struggled at recognizing and correctly classifying rare, minority classes, as there was too little information in the available dataset to gain accurate information. In response, GANs were trained to produce synthetic attack flows from the existing data, and then a DRL model was trained to combat rare attacks. The researchers proved that "[DRL-based IDS trained by GANs can] interact with the network and identify attack classes with competitive accuracy. In addition, ... generating synthetic data for underrepresented classes can improve the precision and recall within these classes, thus acting as a potential solution for imbalanced datasets" (Strickland et al., 2024, p. 15). The IDS could easily handle random, rare attacks by integrating revolutionary technologies. It is crucial that the software can adapt to the ever-changing world as hackers keep getting more creative. GANs and DRL techniques are valuable cybersecurity tools; merging the two can lead to more accurate and precise simulations that can help with real-world threats, leading to a safer, more secure cyber landscape.

In conclusion, DRL implementation is needed to combat the complex digital world and keep private information safe. At a time when traditional methods are starting to show rust, innovative techniques must be implemented to respond appropriately to the increasing frequency of cyber attacks. DRL provides multiple techniques that can have a positive real-world impact on intrusion detection systems. DQNs and DDQNs offer precision and accuracy, already competing

with contemporary methods, and have succeeded greatly. PPO brings stability and security to IDSs, which are much-needed attributes in the volatile, unpredictable digital world. In addition, integrating DRL systems and generative adversarial networks brings the best of both worlds to the table and truly exemplifies the expanding possibilities within cybersecurity. DRL and GAN hybrid models can better train IDSs to learn from rare attacks, removing any element of surprise. By combining and continuously improving these methods, DRL offers a diverse arsenal to aid in shutting down the surge in cyber crimes while adapting to new developments. Through deep reinforcement learning, the digital world can be hopeful for a safer, more secure future.

References

Equifax Data Breach Settlement. Federal Trade Commission. (2024, November 4). https://www.ftc.gov/enforcement/refunds/equifax-data-breach-settlement

Hill, A. (2021, April 6). *Stable-baselines*. PyPI. https://pypi.org/project/stable-baselines/

Larsen, T. N., Teigen, H. Ø., Laache, T., Varagnolo, D., & Rasheed, A. (2021). Comparing Deep Reinforcement Learning Algorithms' Ability to Safely Navigate Challenging Waters. Frontiers in robotics and AI, 8, 738113. https://doi.org/10.3389/frobt.2021.738113

Lopez-Martin, M., Carro, B., & Sanchez-Esguevillas, A. (2019, September 18). *Application of deep reinforcement learning to intrusion detection for supervised problems*. Expert Systems with Applications. https://www.sciencedirect.com/science/article/pii/S0957417419306815

Miles, I., Farmer, S., Foster, D., Harrold, D., Palmer, G., Parry, C., Willis, C., Mont, M. C., Gralewski, L., Menzies, R., Morarji, N., Turkbeyler, E., Wilson, A., Beard, A., Marques, P., Roscoe, J. F., Bailey, S., Cheah, M., Dorn, M., … H, J. (2023). *Reinforcement Learning for Autonomous Resilient Cyber Defence*. FNC Tech Report. https://www.fnc.co.uk/media/mwcnckij/us-24-milesfarmer-reinforcementlearningforautonomousresilientcyberdefence-wp.pdf

Phan, T. V., Nguyen, T. G., Dao, N.-N., Huong, T. T., Thanh, N. H., & Bauschert, T. (2020, September). *IEEE Xplore*. DEEPGUARD: Efficient Anomaly Detection in SDN With Fine-Grained Traffic Flow Monitoring. https://ieeexplore.ieee.org/document/9123430

Shah, D., & Xie, Q. (2018). *Q-Learning with nearest neighbors*. MIT Libraries. https://dspace.mit.edu/bitstream/handle/1721.1/137946/7574-q-learning-with-nearest-neighbors.pdf?sequence=2&isAllowed=y

Strickland, C., Zakar, M., Saha, C., Soltani Nejad, S., Tasnim, N., Lizotte, D. J., & Haque, A. (2024, April 25). *DRL-Gan: A hybrid approach for binary and multiclass network intrusion detection*. MDPI. https://www.mdpi.com/1424-8220/24/9/2746