Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 15 No. 4, 2024 - www.cisdijournal.net

Development of an Improved Model for Zero-Day Attack Detection Using Reinforcement Learning

Oguntunde, T. & Adeosun, O.A.
Department of Computer science
University of Ibadan
Ibadan. Nigeria.
E-mails: tantos557@yahoo.com [2]midebbie10@gmail.com

ABSTRACT

Zero-day attacks are one of the most critical problems in the contemporary scope of cybersecurity as they can use the unidentified vulnerabilities before the software is modified with a patch or signature. Commodity signature-based intrusion detection systems (IDS) are effective against known attacks but do not recognize new and emerging attack patterns. Such a restriction has led to heightened insecurity violations, loss of finances, and disruption of operations in organizations. Reinforcement Learning (RL) and its adaptive learning feature and dynamic decision-making plans have become one of the promising solutions to identify intelligent and hitherto unknown cyber threats. This study aims at creating a more advanced version of Reinforcement Learning to achieve successful detection of zero-day attack, which is an improvement of the current machine learning and IDS methods. The data for this study was collected from University of New Brunswick website in collaboration with Canadian institute of Cybersecurity (CSE-CIC IDS2018). An improved Deep Adaptive Q-Network(DAQ-Network) Model was ensemble from Actor-Critic and Deep Q-Network Models. The collected data was preprocessed by removing the redundant features, missing values and normalization of the data was done. The preprocessed dataset was partitioned into training and testing in 70:30 proportion respectively. The ensemble DAQ-Network Model was trained with the training dataset to detect Zero-day attack. The performance evaluation was done by comparing Actor-Critic, Deep Q-Network and Deep Adaptive Q-Network Models using accuracy, precision, F1-score and recall as metrics. A total of 98,455 datasets was used, 68,918 for training and 29,537 for testing. Actor-Critic(A2C), DQ-Network, DAQ-Network Models achieved an accuracy of 0.6104, 0.0.7403 and 0.9410, respectively; Precision of 0.5977,0.6144 and 0.8942, recall,0.5206,0.5883,0.9029 while their F1-Score were 0.4913,0.6338 and 0.8985, respectively. This indicated that the ensemble Deep Adaptive Q-Network Model outperformed A2C and DQ-Network models individually. The ensemble Deep Adaptive Q-Network model was better in detection of Zero-day attacks than constituent models. It could be deployed as an add-on to enterprise servers in organization or corporate place to detect zero-day attacks.

Keywords: Deep Adaptive Q-Network, Zero-day attacks detection, Actor-Critic(A2C), Datasets, Data, Reinforcement learning-based model.

CISDI Journal Reference Format
Oguntunde, T. & Adeosun, O.A. (2024): Development of an Improved Model for Zero-Day Attack Detection Using Reinforcement Learning. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 15 No 4, Pp 55-69. Available online at www.isteams.net/cisdijournal. dx.doi.org/10.22624/AIMS/CISDI/V15N4P7

1. INTRODUCTION

The extreme rise in technological progress over the last 15 years and beyond has resulted in a geometric rise of cybersecurity threat, where one of the most difficult to detect is a zero-day attack, and many of them are capable of a substantial level of damage (Leyla, 2019). Conventional intrusion detention system (IDS) tend to be signature-based, and they cannot detect emerging threats, which causes major security violations (Gowthami & Priscila, 2024). The exploits of the zero-day attacks are unknown vulnerabilities either in software or hardware and are highly dangerous because they lack pre-existing signature or patches to counter the weakness. Reinforcement learning (RL) has gained immense popularity and few methods to make the task of identifying zero-day attack a bit easy. This encompass Intrusion Detection System (IDS) and antivirus programs which are mainly based on signature-base and heuristic detection method that are not effective with zero-day threat because they are based on the known attack patterns. This has led to the increased demand of a powerful intrusion detection system that has the capability of detecting a zero-day threat. But such systems must be prepared against unknown attack and unknown signature attack. (Hanan, 2020).

Mechanisms like Intrusion Detection System, firewall, scanner and antivirus software, and others are applied to avoid different kinds of cybersecurity assaults. There are numerous attacks which are often launched on network-connected devices. Nevertheless, current machine learning base zero-day attack detection model do have a number of weaknesses, such as high frequency of false positives, low flexibility to new attack approaches, and inefficiency in terms of computation. Consequently, it is necessary that a better model can be developed that improves the detection accuracy, lowers false positive as well as makes the detection of zero-day attack real-time owing to the shortcoming of the existing IDS, which are based on pre-established pattern and signature of different previous attacks. Furthermore, the existing IDS has very high false positive rates, thereby restricting its performance and practical application in the real world application. Consequently, high volume of zero-day attack has remained unknown and contribute to the escalation of the impact (denial of service, identity theft and stolen customer details). Zero-day attack is never slow and zero-day attackers do not quit their struggle easily. The assault is covert.

Zero-day attack exploit refers to a malicious software known as malware that is used to gain access to the target system. A cybercriminal could exploit a vulnerability in a non-exploitable malicious software that is a zero-day vulnerability or software bug that is not known yet. The zero-day attack takes it further, not only is the system penetrated, but also hackers use the exploit to create damage or take advantage of the system and steal valuable resources. The development of Internet in the early 1990s bonanza exemplifies excel opportunities in the domain of communication, commercial and information exchange. Moreover, new security issues also emerge with the creation of digital evolution. Nevertheless, the first cybersecurity solutions were primarily signature-based ones by which all the known malware codes were detected and stored in a database. Traditional IDS and antivirus software, which is based on this approach was effective against known threat. This approach could not resist new attacks that had never been witnessed before, which were later known as zero-day attacks. With the increased complexity of cyber threat in the late 1990s and early 2000, the use of heuristic approach in addition to rule based detection became the norm in detecting anomalies. These systems, however, tended to produce high positive rate, and were unable to perform with real-time detection.

The 2010s witnessed the transition to machine language and reinforcement learning as a sub-branch of Artificial Intelligent(AI) in cyberspace where both statistical and AI-based models were used. It is a cyber-crime golden age. Over time, threat actors developed. They started ransomware attack instead of botnet that bombarded target with huge amount of spam that was usually randomly selected. Huge technological change was also experienced in the 2010s. The promotion of cloud computing, extensive usage of mobile gadgets and installation of IoT equipment implied that business could not specify the sensitive information that was always stored within their data canters.

2. RELATED WORKS.

Emmah et al., (2021) has used dimensionality reduction methods to compare the Principal Component Analysis (PCA), Truncated Singular Value Decomposition (TruncatedSVD), and Pareto-Based Monte Carlo Filtering (PB-MCFR) in detection of zero-day malware. They utilized the KDD data to demonstrate that PB-MCFR was more accurate than the earlier one with 100 percent accuracy and demonstrated that feature selection can enhance the performance of models.

Muchiri Ndungu (2021) created a model of locating zero-day anomalies. The suggested solution was used to analyze time series data of the UNSW-NB15 dataset to detect abnormal behavior and reach an accuracy of 93 percent in assessments. In this paper, it was shown how RNNs can be trained on a small amount of training data and state why time simulation is an important aspect of cybersecurity.

Sarhan et al., (2023) came up with an algorithm known as zero-shot learning (ZSL) algorithm to test machine learning-based Network Intrusion Detection Systems (NIDS) that allow detection of zero-day threats. The study revealed that the identification of some attacks was very precise though some were not by exclusion of attack classes in the training process and semantic features obtained through known attacks. The outcomes indicate that we must have comprehensive feature engineering and validation metrics like Zero-day Detection Rate (Z-DR), to achieve adequate evaluation of new risks.

Teymourlouei et al. (2023) compared machine learning data reduction method with deep learning in the detection of zero-day vulnerabilities. Their approach was effective to reduce the attributes of datasets in search of threats and a Deep Neural Network had a high accuracy rate (97) and a low rate of a false positive. The findings suggest the use of a multi-layered model that integrates data reduction and deep learning to enhance the accuracy and reliability of threat detection. Diffusion models and generative AI are a new method of creating fake data and discovering issues. This is besides conventional and monitored procedures. They are better training sets because these models work as networks do. They are also able to simplify the process of locating abhorrent risks and enhancing the ability to halt them instantly.

The proposed dual-layered ML framework that applies to real-time zero-day detection is a good example of an hybrid architecture of an unsupervised anomaly environment and a supervised classification environment to boost performance and decrease latency. These works indicate that unsupervised, supervised, and hybrid detection techniques should be combined to ensure full detection of zero-day attacks. Research indicates that Intrusion Detection Systems, which rely on deep learning, cannot handle new threats, and idea drift.

A stronger system is created by new ideas, like normalizing flows to quantify prediction uncertainty, which aim to investigate inputs that are not within the normal range further. Numerous studies have been performed on machine learning to detect zero-day attacks, such as creating feature engineering, advanced neural networks, hybrid and generative models and novel testing techniques. Some of the current problems are that there is no adequate real-world zero-day data, variability in detection, and hardware constraints and deployments that are both interpretable and adjustable.

Schulman et al. (2015) introduced Generalized Advantage Estimation (GAE) which is much more stable in variance in advantage computation and learning especially in continuous-control environments. Later, Schulman et al. (2017) created Proximal Policy Optimization (PPO), a type of Actor–Critic algorithm that implements large policy updates with clipping and has since become the current state-of-the-art baseline on on-policy Reinforcement Learning because of its robustness and easy tuning. On the whole, the A3C, A2C, GAE, and PPO development indicate the significance of the joint optimization of policy and critic-based value estimation, which results in the high-performance, scalable, and stable models of reinforcement learning that found numerous applications in cybersecurity, robotics, and autonomous systems.

Wu et al. (2024) proposed a novel approach to addressing the limitations of traditional Network Intrusion Detection Systems (NIDS) in detecting zero-day attacks. Zero-day attacks pose significant challenges because they involve previously unseen attack patterns, while most existing intrusion detection models rely heavily on large volumes of labeled data and predefined attack signatures. This dependence results in reduced detection accuracy when encountering new or evolving threats. To overcome these challenges, the authors introduce an **Active Deep Q-Network (ADQN)** framework that integrates **active learning, deep reinforcement learning**, and **deep learning** techniques. The framework consists of three main components: a NIDS classifier, a sample selection strategy, and an annotator. A Deep Q-Network (DQN) is employed as an intelligent controller to dynamically select the most informative samples for labeling based on probability distributions, thereby minimizing the need for extensive manual labeling. Additionally, a **Bi-directional Long Short-Term Memory (BiLSTM)** network is incorporated into the DQN to capture temporal correlations in network traffic data, improving decision-making in sample selection. Selected samples are labeled using a Euclidean distance-based approach to enhance labeling accuracy.

The proposed framework is evaluated using two widely recognized intrusion detection datasets, **NSL_KDD** and **UNSW_NB15.** Experimental results demonstrate that the ADQN framework outperforms traditional supervised learning models and existing intrusion detection approaches, particularly in zero-day attack scenarios. The model exhibits improved generalization, stability, and detection accuracy while significantly reducing labeling cost and retraining overhead. Overall, the study shows that combining active learning with deep reinforcement learning provides an effective and scalable solution for zero-day attack detection in dynamic network environments.

Ren et al.(2022) addressed the growing threat of network attacks, particularly unknown and zero-day attacks, which exploit previously unseen vulnerabilities and evade traditional intrusion detection systems (IDS). Such attacks include DoS, DDoS, botnets, brute-force attacks, web attacks, and infiltration attempts, which are increasingly stealthy, diverse, and difficult to detect using signature-based or conventional machine learning techniques.

The authors highlight that the existing IDS models suffer from high false-alarm rates and poor detection performance due to redundant features, data imbalance, and limited capability to generalize to unknown attack patterns. To mitigate these challenges, the study proposed the ID-RDRL model, which combines recursive feature elimination (RFE) and deep reinforcement learning (DRL) to improve attack detection. The attack detection process focused on identifying malicious network traffic by selecting the most relevant features from raw traffic data and learning optimal classification policies through interaction with the environment. By removing approximately 80% of redundant features, the model reduces computational overhead while preserving critical attack characteristics. Reinforcement learning enables the system to adapt dynamically and improve its ability to distinguish attack traffic from benign traffic, even in complex and evolving network environments.

Oguntunde and Momoh (2023) developed an Hybrid Machine Learning Model, Randomized Logistic XGBoost (RL-XGBoost), to Detect Email-based Social Engineering Attacks. Seven thousand, two hundred and seventeen emails were harvested from the university of Ibadan email server for the study. The collected emails were preprocessed and partitioned into 80% (training) and 20% (testing). Randomized Logistic XGBoost (RL-XGBoost) was developed from Logistic regression(LR), Random Forest(RF) and Extreme Gradient boost called Randomized Logistic XGBoost (RL-XGBoost). The performance evaluation of RL-XGBoost was done by comparing it with Logistic Regression, Random Forest and Extreme Gradient boost individually using accuracy,precision, recall, F1-score, ROC-AUC as metrics. Logistic Regression, Random Forest, Extreme Gradient boost and RL-XGBoost had accuracy of 0.86 ,0.89 ,0.92 and 0.94, respectively; precision of 0.71, 0.84 ,0.88 and 0.90, respectively. Moreover, Logistic Regression, Random Forest, Extreme Gradient boost and RL-XGBoost had recall 0.74, 0.86, 0.90, and 0.96, respectively; F1-score of 0.73, 0.85, 0.89, and 0.93, while they have ROC-AUC of 0.78, 0.82, 0.90 and 0.92, respectively. The Randomized Logistic Extreme Gradient Boost outperformed the existing, individual models and can be deployed as and add-on into an existing email filtering systems within organizational mail gateways to automatically flag suspicious messages and assist security teams in early threat detection.

Sindhiramutty, Prabagaran & Jhanjhi (2024) proposed an AI-driven framework for proactive detection and simulation of zero-day vulnerabilities, addressing the limitations of traditional reactive security methods. As modern software systems grow in complexity, conventional techniques such as signature-based scanning and manual audits are increasingly ineffective against unknown threats. The proposed system integrates neural vulnerability prediction, semantic code analysis, and adversarial learning to anticipate exploitable weaknesses before public disclosure. The framework leverages code embedding and graph neural networks to model control flow, data dependencies, and compositional interactions within software, enabling accurate classification of vulnerability types such as memory safety flaws and injection attacks.

Reinforcement learning is employed to simulate realistic exploit strategies, allowing autonomous agents to learn multi-stage attack paths that reflect real attacker behavior. Generative adversarial networks further enhance exploit synthesis by producing diverse and context-aware attack scenarios. An automated red-teaming and validation pipeline verifies predicted vulnerabilities and generated exploits in isolated containerized environments, creating a continuous feedback loop that refines model accuracy and exploit realism. Experimental results on enterprise-scale repositories demonstrate strong predictive performance, particularly for memory-related vulnerabilities, and provide significant temporal lead time over traditional discovery methods. Overall, the study

established predictive offensive security as a viable approach for transforming cybersecurity from reactive defense to intelligence-driven anticipation.

Seraj, Pimenidis and Trovati (2025) addressed the growing threat of Android botnets, particularly zero-day variants that evade traditional detection techniques. Due to Android's open ecosystem and rapid app proliferation, botnets can easily infiltrate devices and enable attacks such as data theft and distributed denial-of-service (DDoS). To overcome limitations of existing methods, the authors proposed a permission-based detection framework using an optimized multilayer perceptron (MLP) neural network. A novel dataset of 3,458 Android applications is introduced, each represented by 455 permission-based features extracted via static analysis. By leveraging the full set of Android permissions, the model captured subtle behavioural patterns indicative of botnet activity, including previously unseen threats. Experimental results show that the proposed MLP achieves 98.5% accuracy, outperforming traditional machine learning classifiers and deep learning baselines. The study demonstrates that permission-based neural models are effective, scalable, and well-suited for proactive zero-day Android botnet detection, contributing significantly to mobile security research.

Mathew & Ans Maria (2025) outlined the setup, configuration, and deployment of a Machine Learning (ML)–based Intrusion Detection and Prevention System (IDPS) designed to detect anomalies and mitigate zero-day attacks within a controlled IoT environment. It specified the required hardware, including a minimum of 8 GB RAM, a multi-core processor, and SSD storage, alongside software requirements such as Python 3.11.9, VS Code, Jupiter Notebook, and essential ML libraries. The system relies on network traffic datasets containing features like packet size, response time, ports, and IP byte statistics.

Furthermore, Oguntunde and Ola(2020) developed a Machine Learning Model to Enhance Incremental Approaches for Anomaly Detection in a Network.Their paper asserted that various Anomaly Detection Systems are capable of detecting both known and unknown attacks and able to work in online mode but with challenges of a high rate of false alarm. A model for real-time detection and classification of network anomaly was developed using a One-class Support vector machine (OCSVM) and Incremental approach to reasonably reduce the false alarm rate.The KDD'99 datasets comprising of 494,021 observations which contain 24 training attack types, with an additional 14 types in the test data was used for the study. The One-class Support Vector Machine (OCSVM) technique predicted accurately at a 95 percent accuracy rate.

3. METHODOLOGY

3.1 Data Collection

A public available and representative network intrusion dataset was used. The dataset was collected from University of New Brunswick website in collaboration with Canadian institute of Cybersecurity. All efforts to collate data locally yielded no result because organizations and financial institutions are not willing to share their enterprise network data for reasons of confidentiality and trust.

3.2 Data preprocessing and Feature Engineering.

The collected data underwent removal of duplicate and invalid field, non-useful columns, redundant feature removed, normalization and partitioning into training and testing dataset in 70:30 ratios respectively.

3.3 Model Selection

Reinforcement learning models was deployed for the study. Two of Reinforcement learning models were used. Deep Q-Network (DQN) model which is value-based and Actor-critic that combines stability and adaptiveness. Each of the selected model compensated for the weakness of one another.

3.3.1. Deep Q-Network Model

Deep Q-Network (DQN) is a type of Reinforcement Learning (RL) algorithm, which integrates both Q-Learning and Deep Neural Networks to allow making intelligent decisions in a complex environment. DQN has been applied in cybersecurity to identify zero-day attacks by learning malicious behavior patterns with the network traffic, without any signatures or prior attack labels. DQN has the capability of identifying zero-day attacks through learning the behavioral patterns, adapting as time progresses, and generalizing to the unknown cyber-attacks. DQN allows an AI agent to perform the most optimal actions through interaction with an environment (e.g., network traffic).

The policy is a rule known to the agent that specifies the optimal course of action in a particular state with maximum time rewards. DQN is taught by deep learning the value of each action in each case. A Q-value that indicates what the future reward is likely to be is used to represent this value. The element of DQN are states(s) that store the network traffic attributes, action: (a) that makes a choice of whether to allow, block or Flag the traffic, Reward(r) will provide +1 of the correct identification and -1 of the wrong identification. The neural network that is used to estimate the Q-values is called Q- Network. The rule of action will be provided by policy. Experience replay is a store of previous experiences to be stored in the form of buffer to be learned stably, also there is target network which is a network dedicated to stable learning updates.

3.3.2. Actor-Critic Model

Actor- Critic model is a complex reinforcement learning architecture that has two parts including the Actor that learns the best policy to take an action, and the Critic that determines the quality of the action taken by the Actor through value functions. The critic gives a feedback signal (TD error) that helps in updating the actor's policy to give better training stability and convergence rate than in pure value-based or policy-based training. Actor-Critic model is effective in zero-day attack detection, where the adaptive and real-time learning of dynamic attacks is possible as it also enables detecting new and elusive cyber threats without pre-existing signature. It is meant to overcome drawbacks of classical RL approaches like Q-Learning and reinforce by enhancing the stability, efficiency, and convergence speed. Actor-Critic systems prove to be particularly useful in complex settings like network security, robotics, self-driving systems and adaptive cybersecurity systems.

3.4 Analysis of the selected hybridized models (Deep Q-Network and Actor-Critic Model)

The hybridized models include Deep Q- Network model and Actor-Critic model. These models were chosen so that the weakness of one model is compensated by the strength of one other. Deep Q-Network (DQN) and Actor-Critic (A2C) are largely dissimilar with regards to the mechanism of learning, the decision-making process, and the process of adaptation to new conditions like the zero-day cyberattack. Nonetheless, DQN is an algorithm that works on values and tries to predict Q-values of all possible states-actions combinations with the help of a deep neural network. Its strength is in the fact that it gives accurate value estimates and strictly optimal courses of action on discrete environments. It implies that DQN is useful in Intrusion Detection activities when allow, block, or

flag/alert actions are explicitly defined. Other techniques like experience replay or target networks have also been applied to DQN and they assist in stabilizing the learning process and making credible decisions using historical attack patterns. Nevertheless, DQN also fails in settings where environments are highly dynamic or continuous or where the threats change rapidly, such as it is with zero-day attacks. The model is very dependent on past observations thus changing gradually to new or unnoticed behaviors.

On the contrary, Actor-Critic model constitutes the above-discussed weakness of DQN by integrating policy-based and value-based learning. The Actor acquires a clear-cut policy to converge the states to the actions, whereas the Critic estimates the decisions of the Actor through a value function and calculates the temporal-difference (TD) error to direct the reinforcement. This two-frame facilitates converge more quickly, facilitates learning and allows exploration of invisible behaviors. Actor-Critic models perform well in non-stationary settings and they adjust rapidly to new attack patterns since the Actor regularly changes the policy regardless of predetermined Q-values. Although Actor-Critic is both more computationally expensive and susceptible to training imbalance between the Actor and Critic networks, it offers greater adaptability, generalization, and resilience and therefore is more applicable to complex cybersecurity tasks like zero-day attack detection. Overall, it is possible to say that DQN is robust in structured problems with fixed patterns, whereas Actor-Critic is more suitable in dynamic environments where threats keep changing, and the system will have to learn constantly.

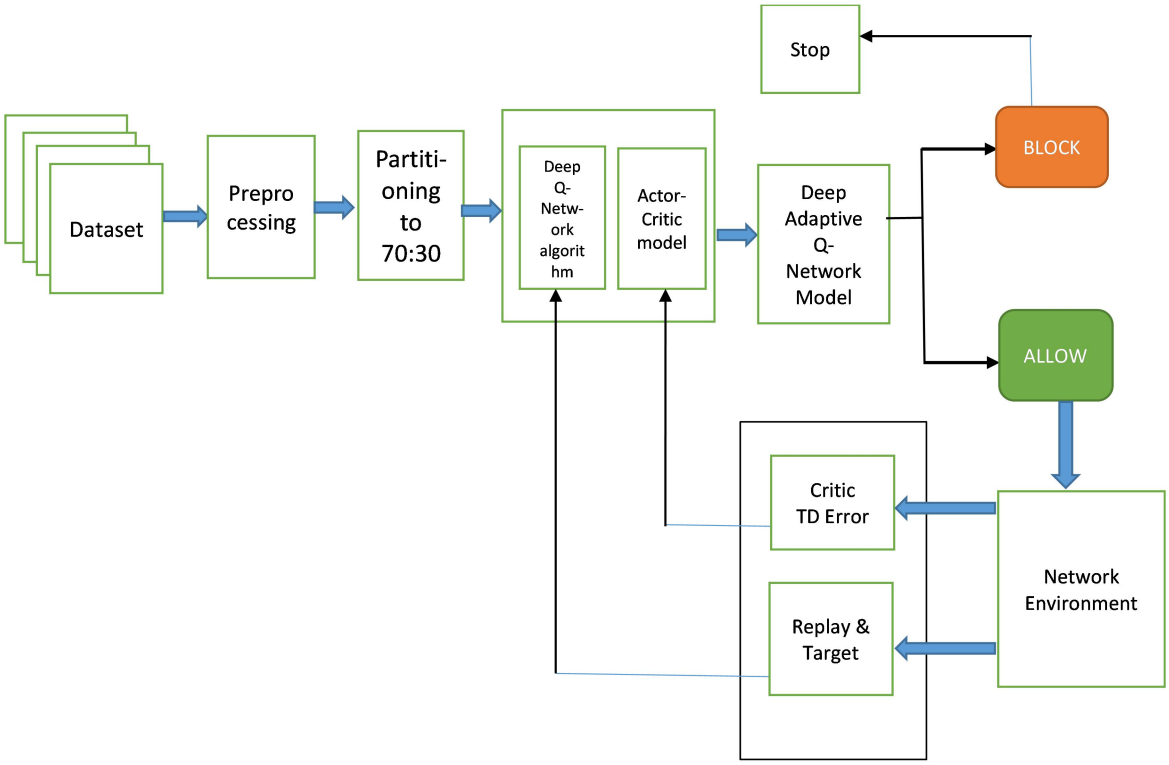

Figure 1: Deep Adaptive Q-Network Model

In Figure 1, Deep Adaptive Q-Network Model is an ensemble of Deep Q-network and Actor-Critic model, one compensating for the weakness of the other. The Deep Q-network takes intelligent decisions based on generalization from malicious behavioral pattern. It does not need any digital approval to authenticate network traffic. The Actor-Critic model fits with the Deep Q-network in that whatever decision made by the Deep Q- network is being judged for quality by the Critic component of the Actor-critic model, which report either error or valid discovery of vulnerability from which the two models learn and make further decision. The Actor-Critic algorithm propose the best action base on policy of network while Deep Q-network algorithm refine and evaluate the proposed decision from Actor algorithm based on Q-value. This means, Actor chooses action, the DQN scores them, both then merge decision into final intrusion decision system (IDS) decision.

4. RESULTS AND DISCUSSION

The evaluation was conducted on the extracted network dataset, harvested from University of New Brunswick website in collaboration with Canadian Institute of Cybersecurity (CSE-CIC IDS2018) using two reinforcement learning-based methods: The Actor-Critic (A2C) model and Deep Q-Network Model, then with Deep Adaptive Q-Network (DAQ-N) model.

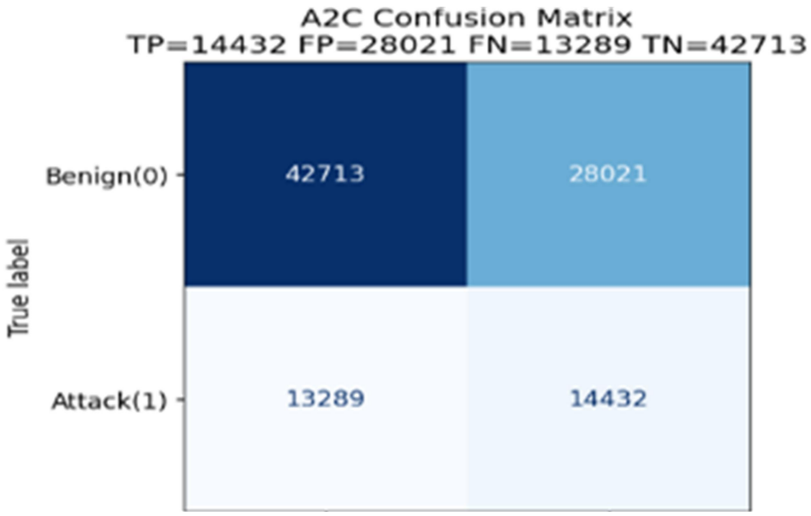

Figure 2: The Confusion Matrix for Actor-Critic Model

The performance of each model was assessed using the following performance metrics:

- Accuracy
- Precision
- Recall / True Positive Rate
- F1 Score
- False Negative Rate (FPR)
- ROC AUC
- Confusion Matrix

The performance of the models was critically compared. The strengths and weaknesses of each model are discussed. The goal of the analysis is to determine the effectiveness of the DAQ-N model in detecting zero-day attacks when compared to a conventional Actor-Critic and deep Q-Network reinforcement learning approach.

4.2.1 Actor–Critic (A2C) Model

The Actor-Critic model combines policy-based and value-based reinforcement learning. Its performance from the implementation is as in table 1

Table 1 Performance Evaluation of Actor-Critic Model

Accuracy	0.6104
Precision	0.5799
Recall	0.5206
F1-Score	0.4913
ROC-AUC	0.5241
False Negative(FN)	13,289

- True Negative (TN) = 42,713 → Benign samples correctly predicted as Benign.
- False Positive (FP) = 28,021 → Benign samples incorrectly predicted as Attack.
- False Negative (FN) = 13,289 → Attack samples incorrectly predicted as Benign.
- True Positive (TP) = 14,432 → Attack samples correctly predicted as Attack.

The A2C model attained an accuracy of 0.6104 and this implies that it rightly classified approximately 61 percent of network traffic samples. Although its precision is moderate, 0.5799, it is not everything that is detected by the model in terms of attacks indicating that, only approximately 58 percent of traffic cases represented as attacks were malicious. This indicates that the model produced significant numbers of false alarms, which can be a burden to the security analysts. The recall value of 0.5206 shows that the model was able to find about 52 percent of real attacks. Though this is an improvement over random detection, it nevertheless implies that almost half of all attacks were not detected, which is a significant security threat. F1-score of 0.4913 also illustrates the low performance of the model since it shows that the model has a poor balance of precision and recall. In general, A2C has an average sensitivity, lacking reliability for use in strong intrusion detection.

4.2.3 Deep Q-Network (DQN)

DQN is a value-based reinforcement learning approach that approximates Q-values using a deep neural network. It is good at stable Q-value estimation but can be slow in adapting to unseen attack patterns, table 2.

Table 2: Deep Q-Network Performance Evaluation.

Accuracy	0.7403
Precision	0.6144
Recall	0.5383
F1-Score	0.6338
ROC-AUC	0.6938
False Negative(FN)	16,469

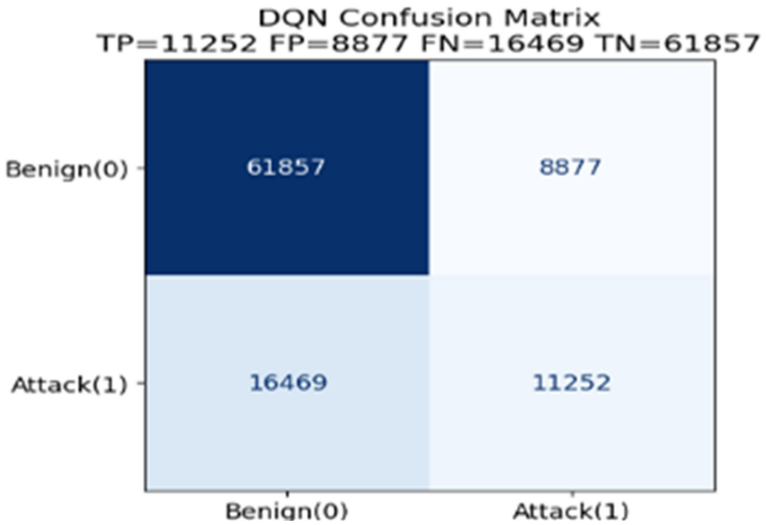

Figure 3: Deep Q-Network confusion matrix

True Negative (TN) = 61,857 ⟶ Benign samples correctly predicted as Benign.
False Positive (FP) = 8,877 ⟶ Benign samples incorrectly predicted as Attack.
False Negative (FN) = 16,469 ⟶ Attack samples incorrectly predicted as Benign.
True Positive (TP) = 11,252 ⟶ Attack samples correctly predicted as Attack.

The Deep Q-Network (DQN) model demonstrated stronger performance compared to A2C, achieving an accuracy of 0.7403, meaning it correctly classified about 74% of all network traffic instances. This indicates that DQN learned more stable patterns from the dataset and generalized better during evaluation. The precision for the attack class was 0.6144, showing that approximately 61% of all traffic predicted as malicious were indeed attacks. This is an improvement over A2C, indicating fewer false alarms and more reliable attack predictions. The recall, 0.5383 shows that the DQN model successfully detected about 54% of actual attacks. While this still leaves room for improvement, it demonstrated better sensitivity than many baseline machine-learning models commonly used in intrusion detection. The F1-score of **0.6338** reflects a reasonable balance between precision and recall, confirming that DQN provides more consistent detection performance. Overall, DQN shows moderate effectiveness and significantly better discrimination of attack traffic than A2C.

4.2.5 Deep Adaptive Q-Network (DAQ-N), the ensemble Model

The DAQ-N is a hybrid model combining both the strengths of A2C and DQN. DAQ-N has best performance compare to the two model.

Table 3: DAQN performance metric

Accuracy	0.9410
Precision	0.8942
Recall	0.9029
F1-Score	0.8985
ROC AUC	0.9143
False Negative(FN)	2,691

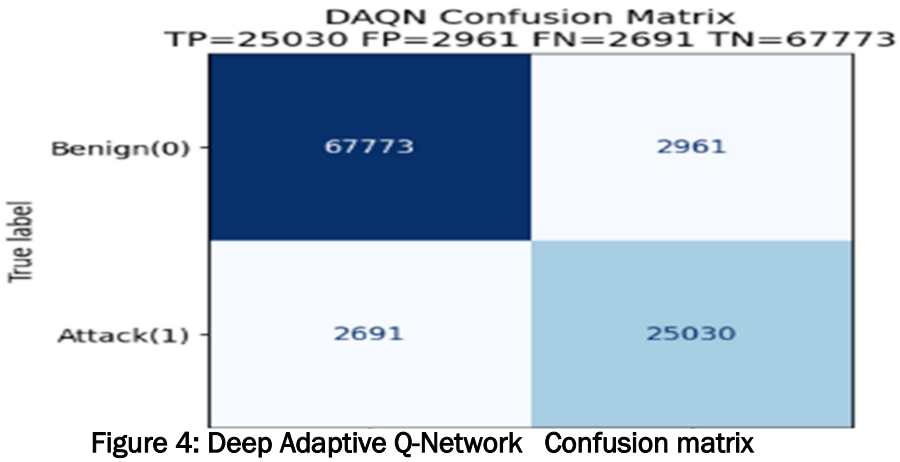

Figure 4: Deep Adaptive Q-Network Confusion matrix

True Negative (TN) = 2,299 ⟶ Benign samples correctly predicted as Benign.
False Positive (FP) = 25,422 ⟶ Benign samples incorrectly predicted as Attack.
False Negative (FN) = 4,013 ⟶ Attack samples incorrectly predicted as Benign.
True Positive (TP) = 66,721 ⟶ Attack samples correctly predicted as Attack.

The overall performance of DAQ-N model is very high in terms of all assessment measures, which signifies the high efficiency and good generalization of intrusion detection system. The model achieved a correct classification of more than 94% of all the instances of traffic with an accuracy of 0.9410, which is a significant improvement relative to both A2C and DQN. The accuracy of the classification of attack, 0.8942, implies that nearly 90% of the traffic that is classified as an attack was actually an attack sample, in other words the model generates very few false alarms, which is a fundamental prerequisite to practical use. The recall figure of 0.9029 also points out a high level of reliability of the model because it identified over 90 percent of real attacks. This will go a long way to minimize the chances of the malicious activity going undetected, which will be one of the key weaknesses in the other models. The F1-score of 0.8985 shows that there was a great equilibrium between the precision and recall, which proves that the model is highly consistent. All in all, the DAQ-N model offers very strong, high-precise and discriminant detection that can be used in zero-day environments of intrusions.

4.3 Comparison of all Models

Table 4: Model performance comparison for A2C, Deep Q-Network and Deep Adaptive Q-Network.

Metric	A2C	DQN	DAQN
Accuracy	0.6104	0.7403	0.9410
Precision	0.5799	0.6144	0.8942
Recall	0.5206	0.5383	0.9029
F1-Score	0.4913	0.6338	0.8985
ROC AUC	0.5241	0.6938	0.9143
FALSE NEGATIVE	13,289	16,469	2,691

The best accuracy belongs to DAQ-N; it is better than both Models. False alarms are also minimized by the Precision. DAQ-N identifies the largest number of attacks and F1 Score has the best total balance. A2C model showed the worst performance in all measures with an accuracy of 0.6104. Its Precision (0.5799) and recall (0.5206) were rather balanced, but the values were not good enough to detect intrusion. The moderate recall suggests that A2C identified some malicious activity, however the low precision and F1-score (0.4913) means that the model made lots of false alarms and gave weak average ability to distinguish between benign and malicious traffic. This implies that A2C was not able to acquire a steady policy in the ever-changing network setting.

The DQN model was significantly better than A2C and had a better F1-score and accuracy of 0.6338 and 0.7403 respectively. Its precision (0.6144) and recall (0.5383) were much larger than those of A2C, indicating that DQN was better used to detect attacks and minimize misclassifications. This has been achieved due to the action selection mechanism based on Q-value and experience replay which stabilized learning and enabled the model to be more generalized based on historical training samples.

The best accuracy (0.9410), precision (0.8942), recall (0.9029), and F1-score (0.7351) were obtained by DAQ-N model in comparison with A2C and DQN. Such findings suggest that the ensembled model was able to exploit the advantages of the two value-based and policy-based learning systems resulting in a more discriminative and powerful intrusion detection system. The high sensitivity towards the malicious activity is indicated by the high recall, whereas the reduced false alarms are indicated by the high precision. All in all, the DAQ-N has better performance and it is the most appropriate model to use in detecting zero-day attacks in this study.

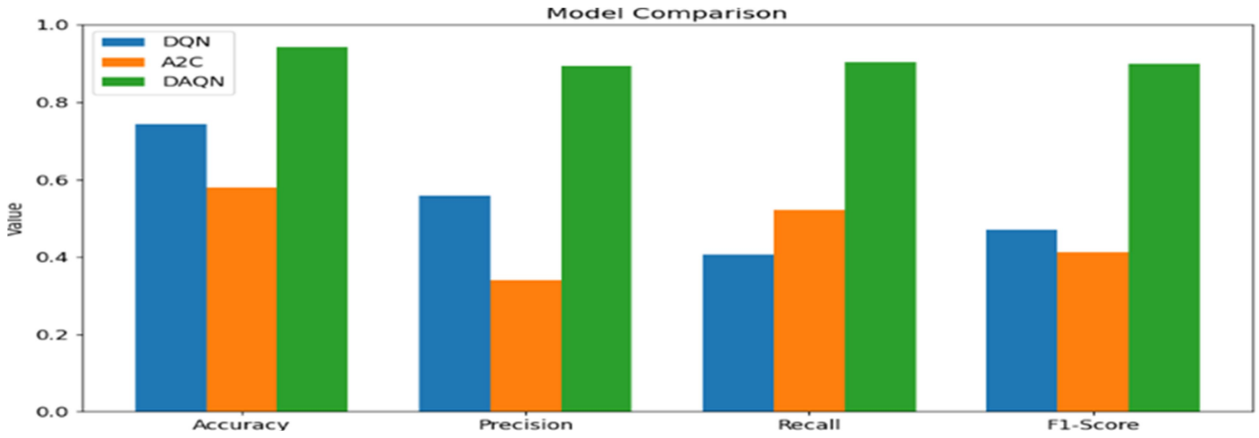

Figure 5: Performance Evaluation of Deep Q-Network model, Actor-Critic model & Deep Adaptive Q-Network models using Accuracy, precision, recall and F1-score as metrics

5. CONCLUSION

The study aimed at creating a superior reinforcing learning-based model that would have a high level of accuracy, adjustability, and stability in detecting zero-day attacks. The experiment conducted and the analysis of three reinforcement learning models, such as A2C, DQN, and the hybrid model, the DAQ-N allow the study to have a holistic outlook on how the reinforcement learning model can be used to improve the current intrusion detection systems. The findings indicate that though reinforcement learning has a promising future in cybersecurity, the model and architecture used have a considerable impact on the performance. Although Actor-Critic (A2C) model had theoretical benefits, its learning behavior was unstable and produced high levels of false alarm, which is why it cannot be applied in practice as a zero-day detector. DQN model was significantly better and it provided more consistent estimation of Q-values and enhanced detection. But it was limited in its ability to adapt to invisible patterns of attack, making it less useful in complicated attacks.

The ensemble hybrid Deep Adaptive Q-Network (DAQ-N) became one of the important contributions of this study. The DAQ-N model was able to tackle the shortcomings of both DQN and Actor-Critic architectures because it was created by applying the strengths of each single model to its detriments. The findings confirm that DAQ-N has better generalization, stability, and adaptability, which are necessary features of identifying hidden and changing threats in real-world networks. The excellent results of the DAQ-N model, where the accuracy is over 94 per cent, and the values of precision and recall are high, prove the feasibility of the hybrid systems of reinforcement learning in advanced cybersecurity application.

The model was also able to minimize false positives and false negatives hence rendering it more suitable to be deployed in operation. These advancements underscore the fact that hybrid RL systems have the capacity to outperform the performance of the conventional signature and machine learning-based systems, in terms of zero-day detection. Finally, the study confirms that combination of reinforcement learning strategies, especially hybridized models can enhance precision, strength, and versatility of intrusion detection systems to a considerable degree. The effectual deployment of DAQ-N is an effective means of curbing zero-day attacks, which has helped in building secure and smarter cybersecurity systems.

6. Bibliography.

1. Alam (2025). *Deep Q-learning intrusion detection system (DQ-IDS): A novel reinforcement learning approach for self-learning cybersecurity ICT Express.* https://doi.org/10.1016/j.icte.2025.05.007
2. Alam et al. (2025). *Adaptive Defense: Zero-Day Attack Detection in NIDS with Deep Reinforcement Learning.* Journal: *IEEE Access.* https://doi.org/10.1109/ACCESS.2025.3585445
3. Alazab, M., et al. (2025). *Advanced Machine Learning Approaches for Zero-Day Attack Detection*. https://www.researchgate.net/publication/388051131
4. Almofarreh et al. (2025). **A Zero-Day Attack Detection Approach Using Deep Learning.** https://www.techscience.com/cmes/online/24887/pdf
5. Bilge, L., & Dumitras, T. (2012). *Before We Knew It: An Empirical Study of Zero-Day Attacks In-The-Wild*. ACM Conference on Computer and Communications Security (CCS). https://doi.org/10.1145/2382196.2382284
6. DeepAI. (2020). *Anomaly Detection Using Autoencoders*.
7. **Emmah, V. T.**, Ugwu, C., & Onyejegbu, L. N. (2021). An Enhanced Classification Model for Zero-Day Attack Detection *and Estimation. European Journal of Electrical Engineering and Computer Science.* https://www.ejece.org/index.php/ejece/article/download/350/215/1337
8. **Evans, K.,** Abuadbba, A., Wu, T., Moore, K., et al. (2021). Reinforcement-aided Spear Phishing Detector. https://arxiv.org/abs/2105.07582
9. Feng et al. (2023). *A Federated Reinforcement Learning Framework for Malware Mitigation arXiv preprint.* https://arxiv.org/abs/2308.05978
10. **Feng, C.,** Huertas Celdran, A., Sanchez Sanchez, P. M., Kreischer, J., von der Assen, J., Bovet, G., & Martinez Perez, G. (2023). A Federated Reinforcement Learning Framework for Malware Mitigation. https://arxiv.org/abs/2308.05978
11. **Gonçalves de Sousa, U.** (2025). Cognitive-Enhanced RL Framework for Cybersecurity Anomaly Detection https://arxiv.org/abs/2509.10511
12. Guo (2023). *A Review of Machine Learning–based Zero-day Attack Detection: Challenges and Future Directions. Computer Communications.*https://doi.org/10.1016/j.comcom.2022.11.001
13. **Gu, Y.** (2023). A Review of Machine Learning-based Zero-day Attack Detection: Challenges & Future. https://doi.sorg/10.1016/j.comcom.2022.11.001
14. **Hossain, M. A.** (2025). Deep Q-learning Intrusion Detection System (DQ-IDS) https://doi.org/10.1016/j.icte.2025.05.007
15. **Huang, Y.,** Huang, L., & Zhu, Q. (2021). Reinforcement Learning for Feedback-Enabled Cyber Resilience. https://arxiv.org/abs/2107.00783
16. Jamshidi et al. (2025). *Carbon-Aware Intrusion Detection: A Comparative Study of Supervised and Unsupervised DRL for IoT Edge Gateways* https://arxiv.org/abs/2511.18240

17. Jamshidia et al. (2025). *Application of Deep Reinforcement Learning for Intrusion Detection in Internet of Things: A Systematic Review. arXiv preprint* https://arxiv.org/abs/2308.05978
18. Johnson (2024). *Leveraging AI for Zero-Day Attack Detection: Challenges and Future Directions. Journal of Artificial Intelligence Research.* Reviews AI-driven zero-day detection methods including future RL integration potential.
19. **Jonnalagadda, N.**, Yan, Z., & Khan, I. (2025). Comparative Study of ML and DL for Zero-Day Detection. DOI 10.21203/rs.3.rs-6211970/v1
20. Lopez-Martin et al. (2020). *Application of Deep Reinforcement Learning to Intrusion Detection for Supervised Problems. Expert Systems with Applications*
21. Mathew & Ans Maria (2025). ML-Based Zero-Day Attack Detection. ML methods for IDS https://norma.ncirl.ie/id/eprint/8225
22. Oguntunde, T. & Momoh, M.S. (2023). Development of an Hybrid Machine Learning Model to Detect Email-based Social Engineering Attacks. Social Informatics, Business, Politics, L:aw, Environmental Sciences & Technology Journal. Vol. 9, No. 3. Pp 21-336.
23. Oguntunde T., Ola M. O., (2020). Development of a Machine Learning Model to Enhance Incremental Approaches for Anomaly Detection in a Network. University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR), Vol. 6 No. 2, pp. 68-78
24. Ren et al. (2022). *A deep reinforcement learning-based feature selection intrusion detection Model. Scientific Reports.* https://www.nature.com/articles/s41598-022-19366-3
25. **Seraj, S.**, Pimenidis, E., & Trovati, M. (2025). Zero-day Android Botnet Detection Using Neural Networks. https://doi.org/10.1007/s00521-024-10818-7
26. Sewak & Sahay (2025). Deep Reinforcement Learning for Cybersecurity Threat Detection & Protection broad cybersecurity RL review.
27. **Sindhiramutty, S. R.**, Prabagaran, K. R., & Jhanjhi, N. Z. (2024). AI-Driven Zero-Day Simulation: Predictive Offensive Modeling. https://jisem-journal.com/index.php/journal/article/download/13722/6478/23282
28. Wu et al. (2024). *An active learning framework using deep Q-network for zero-day attack detection. Computers & Security.* https://doi.org/10.1016/j.cose.2024.103713