Adaptive Deep Reinforcement Learning: A Novel Framework for DDoS Detection on Resource-Constrained Edge Devices

Syaifuddin Saif
Department of Electrical and Information Technology, Universitas Gadjah Mada, Indonesia | Department of Information Technology, University of Muhammadiyah Malang, Indonesia
saifuddin@umm.ac.id (corresponding author)

Widyawan Widyawan
Department of Electrical and Information Technology, Universitas Gadjah Mada, Indonesia
widyawan@ugm.ac.id

Ridi Ferdiana
Department of Electrical and Information Technology, Universitas Gadjah Mada, Indonesia
ridi@ugm.ac.id (corresponding author)

Received: 29 November 2025 | Revised: 4 January 2026, 14 January 2026, and 19 January 2026 | Accepted: 21 January 2026

 DOI: https://doi.org/10.48084/etasr.16594

ABSTRACT

The rapid growth of the Internet of Things (IoT) has significantly increased exposure to Distributed Denial of Service (DDoS) attacks, particularly due to the limited resources and heterogeneous traffic characteristics of IoT devices. Conventional intrusion detection approaches, including supervised learning models, often fail to adapt to dynamic traffic patterns and zero-day attacks. This study proposes an adaptive DDoS detection framework based on Deep Reinforcement Learning (DRL) designed for deployment on resource-constrained edge devices. The framework is evaluated using the IoT-DH dataset, a real-world multi-protocol IoT traffic dataset collected via a honeypot. Experimental results show that the proposed DRL-based approach consistently outperforms static models, achieving an F1-score of up to 0.98–0.99 on in-distribution data while maintaining a low False Positive Rate (FPR). Cross-dataset evaluation on public benchmarks further demonstrates stable performance with F1-scores above 0.96 under distribution shifts. Implementation on a Raspberry Pi 3 confirms that the model operates with low inference latency and acceptable resource usage. These results indicate that adaptive DRL provides an effective and practical solution for real-time DDoS detection in edge-based IoT environments.

Keywords-Deep Reinforcement Learning (DRL); DDoS detection; Internet of Things (IoT); edge computing; Dueling Deep Q-Network (DQN); Prioritized Experience Replay (PER)

I. INTRODUCTION

The application of the Internet of Things (IoT) has expanded rapidly across diverse domains, including smart homes, healthcare, industrial automation, and transportation, highlighting its growing importance in modern digital infrastructures [1]. This rapid adoption is accompanied by a continuous increase in the number of interconnected devices, which is projected to exceed 39 billion active connections worldwide by 2030 [2]. While this growth enables significant technological advancements, it simultaneously expands the attack surface and exposes IoT ecosystems to increasingly sophisticated cybersecurity threats, particularly Distributed Denial of Service (DDoS) attacks [3]. The inherent characteristics of IoT devices—such as limited computational resources, memory, and energy capacity—restrict the deployment of complex security mechanisms and make them attractive targets for large-scale attacks [4]. Furthermore, the heterogeneity of IoT communication protocols complicates the implementation of uniform defense strategies, as each protocol exhibits distinct traffic patterns and vulnerabilities [5]. As a result, protecting IoT environments against dynamic and large-scale DDoS attacks remains a critical and unresolved challenge [6].

Existing intrusion detection systems (IDSs) for IoT environments predominantly rely on rule-based and supervised machine learning approaches. Although these methods can achieve high detection accuracy under controlled and balanced datasets, they fundamentally depend on static decision boundaries and historical traffic distributions. Consequently,

their effectiveness degrades when confronted with heterogeneous protocols, evolving attack strategies, and previously unseen traffic patterns. This limitation has been consistently reported across multiple studies, regardless of the specific classifier or feature engineering techniques employed [7-10]. Even recent deep learning-based IDSs, which offer improved classification performance and lightweight architectures suitable for edge deployment, remain constrained by offline training paradigms. Once deployed, these models lack intrinsic adaptability, limiting their ability to respond effectively to zero-day attacks and abrupt distribution shifts in real-world IoT traffic [11, 12]. Collectively, these observations indicate that the core limitation of current IoT DDoS detection approaches lies not in model complexity, but in the absence of adaptive learning capabilities under non-stationary conditions.

To address this limitation, adaptive learning paradigms have gained increasing attention. Deep Reinforcement Learning (DRL), in particular, enables agents to continuously refine their decision policies through interaction with the environment, making it inherently suitable for dynamic and non-stationary scenarios. Prior studies have demonstrated the potential of DRL to improve learning stability and convergence efficiency in complex environments [13, 14]. However, the application of DRL to IoT-based intrusion detection remains largely limited to cloud-centric or simulation-based settings. As a result, critical aspects such as false alarm behavior under real traffic, robustness to zero-day attacks, and feasibility on resource-constrained edge devices are often insufficiently addressed [15, 16]. This gap underscores the need for a lightweight yet adaptive DRL-based intrusion detection framework that is explicitly designed for practical deployment in edge-based IoT environments.

Accordingly, this study addresses the identified research gap by proposing an adaptive DRL-based DDoS detection framework tailored for resource-constrained IoT edge devices. Unlike prior approaches that rely on static supervised models or cloud-dependent DRL architectures, the proposed framework emphasizes real-world adaptability, robust detection under heterogeneous traffic conditions, and controlled false positive behavior. Multiple DRL variants—namely Deep Q-Network (DQN), Double DQN, and Dueling DQN—are systematically evaluated to assess their suitability for dynamic IoT traffic and zero-day attack scenarios. The framework is further designed to support efficient edge deployment, ensuring that adaptive learning does not compromise inference latency or resource utilization.

The main contributions of this study are threefold. First, it introduces an adaptive DRL-based intrusion detection framework capable of responding to evolving IoT traffic patterns and zero-day DDoS attacks without frequent manual retraining. Second, it provides a comprehensive evaluation of multiple DRL architectures using a real-world IoT honeypot dataset that captures heterogeneous, multi-protocol traffic. Third, it demonstrates the feasibility of deploying adaptive DRL models on resource-constrained edge devices by balancing detection accuracy, false alarm rate, and inference efficiency. Through these contributions, this study aims to advance the development of practical and adaptive IDSs for future IoT environments.

II. METHODOLOGY

This research was conducted through a systematic methodology comprising five principal stages, from the initial literature study to the final reporting of results. The overall workflow of this process is depicted in Figure 1.

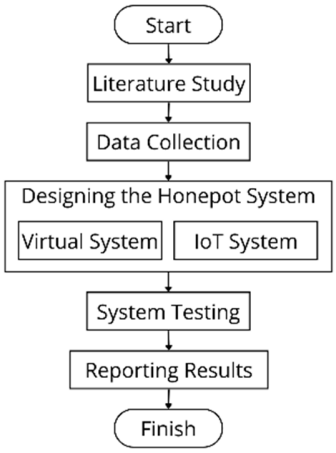

Fig. 1. General research flow.

A. Literature Study

The literature review was conducted through a systematic search of prominent scientific databases, primarily focusing on ACM, IEEE, ScienceDirect, and Springer. The search utilized specific keywords such as "Detections AND Prevention AND Classification," "DDoS-IoT," and "Honeypot AND Architecture." This process yielded 185 articles, which were meticulously selected to identify studies relevant to IoT security, DDoS detection, and honeypot design. The reviewed studies provided a comprehensive understanding of existing methods, their limitations, and emerging trends in adaptive security solutions for IoT environments, particularly regarding DDoS attack detection mechanisms and the integration of honeypots to mitigate such threats.

B. Data Collection

This study utilizes the IoT-DH dataset, generated through the implementation of the RIoTPot honeypot within a real-world IoT network environment, which is publicly available and further described in [17]. The dataset is released under the CC BY 4.0 license, and all uses in this study comply with the license terms. As depicted in Figure 2, the data-recording process began with capturing raw network traffic from six key IoT protocols (MQTT, CoAP, HTTP, Telnet, SSH, and ICMP) into PCAP files using tcpdump. To maintain realism and represent real-world conditions, this dataset contains both benign traffic and observed real attacks—such as TCP Flood (including SYN Flood), UDP Flood, Slowloris, and protocol-specific attacks against MQTT/CoAP services—which were then further processed using TShark and Zeek for feature extraction.

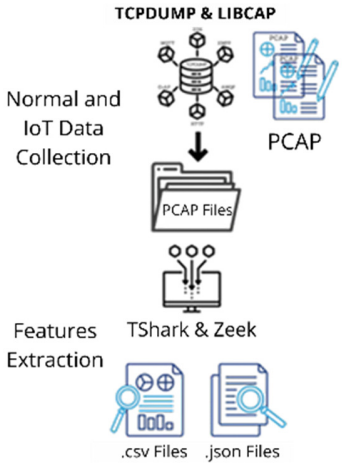

Fig. 2. Data recording flow.

The original dataset comprised a large and highly imbalanced collection of instances across the different protocols. To ensure experimental validity and prevent model bias, a balanced subset was curated for this study. This curated dataset consists of 5,000 samples for each category ('normal' and 'attack') across all six protocols, yielding a total of 60,000 instances for training and evaluation. Finally, to address the high dimensionality of network traffic data and minimize computational overhead on edge devices, a rigorous feature selection process was conducted. We employed an embedded method based on the Random Forest algorithm to evaluate feature relevance. Unlike filter methods that assess features independently, this embedded approach captures feature interactions during the training process, utilizing Gini impurity as the ranking criterion. Crucially, this selection was performed globally across the entire aggregated dataset—rather than individually per protocol—to ensure the identification of generalizable attack patterns across heterogeneous IoT traffic. Based on the importance rankings, the top-10 predictive features were retained to construct the final state space for the DRL agent. The ranking of these selected features is visualized in Figure 3, with attributes such as pktrate, protocol, tx_kbps, and rx_kbps demonstrating the highest contribution to the detection task.

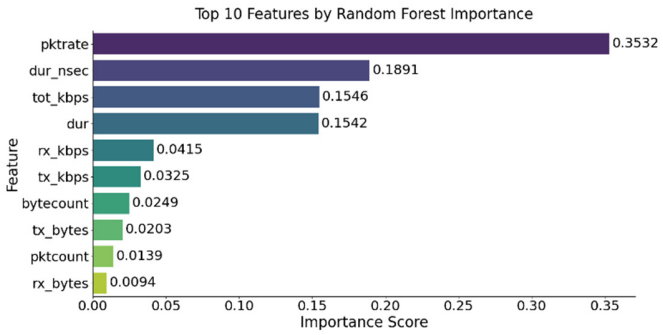

Fig. 3. Top ten traffic features based on Random Forest importance scores.

C. Designing the Honeypot System

The architecture of the proposed security solution, depicted in Figure 4, is designed as a hybrid, two-layer framework. This design strategically separates computationally intensive model training (represented as the 'Build Model' component)—which is conducted on an HPC server equipped with an AMD Ryzen 7 processor (3.4 GHz), 32 GB of RAM, and an NVIDIA GeForce RTX 3080 GPU (10 GB VRAM)—from the real-time inference process handled by resource-constrained edge devices. As illustrated on the 'Production Server,' a Honeypot is deployed to capture network traffic, which populates the Attack Dataset. This dataset serves as the source of the state input for the DRL agent. The agent processes this state information through a Running Model to determine the Best Action—classifying traffic as either benign or malicious. To enhance learning efficiency and stability within the proposed DRL framework, several optimization mechanisms were integrated into the training process. Prioritized Experience Replay (PER) was employed to improve sample efficiency by prioritizing transitions that contribute more significantly to policy updates [13, 18]. Previous studies have shown that PER can accelerate convergence and improve learning robustness in DRL systems [19]. In the context of IoT-based intrusion detection, this mechanism is particularly beneficial for emphasizing critical traffic state transitions under dynamic and non-stationary conditions.

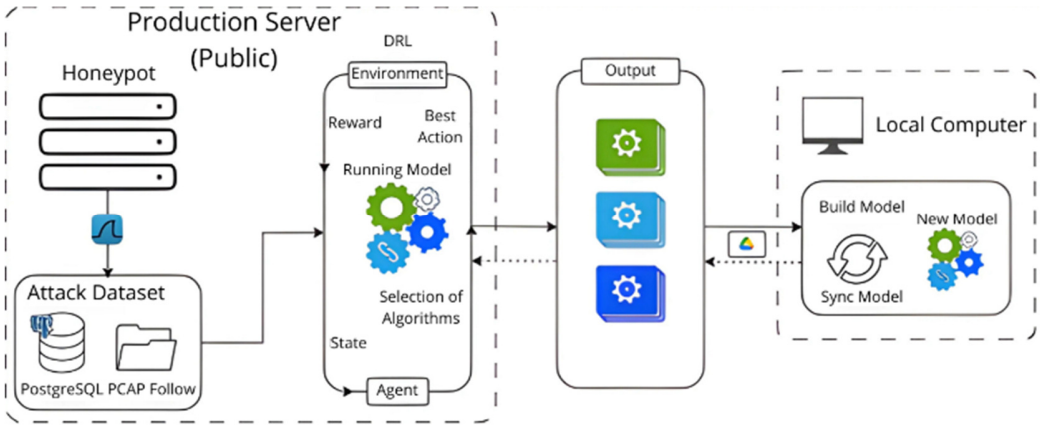

Fig. 4. Proposed system security architecture.

In addition, an Adaptive Reward Shaping (ARS) strategy was incorporated to guide the learning process by dynamically adjusting reward signals based on detection outcomes. Reward shaping has been demonstrated to improve convergence behavior and policy stability in complex reinforcement learning tasks [20]. In this study, ARS is designed to balance detection accuracy and the False Positive Rate (FPR) while maintaining responsiveness to evolving attack patterns.

Furthermore, the Huber loss function was adopted to stabilize neural network weight updates during training. Compared to conventional mean squared error loss, Huber loss reduces sensitivity to outliers and abrupt reward variations, thereby mitigating unstable learning behavior. The combined integration of PER, ARS, and Huber loss aims to improve convergence stability and robustness, which are essential for deploying DRL-based IDSs on resource-constrained edge devices.

The selection and optimization of the DRL models—including Q-Learning, DQN, Double DQN, and Dueling DQN—are performed offline on the HPC server. Each model is enhanced with PER and the Huber loss function to improve learning stability and efficiency. The trained models are then synchronized to edge devices through a Sync Model loop. To ensure feasibility on resource-constrained hardware, the models undergo post-training quantization using the ONNX Runtime engine. This process compresses the model weights from FP32 to INT8 format, reducing the file size from approximately 1.2 MB to 0.3 MB. This optimization explicitly enables the system to operate in a lightweight, portable manner and perform real-time attack detection in resource-constrained IoT environments.

D. Deep Reinforcement Learning Hyperparameter Settings and Reward Formulation

To ensure reproducibility, the specific hyperparameters used for training the Dueling DQN agent with PER are summarized in Table I. The priority exponent α dictates how much prioritization is used, whereas the importance-sampling weight β corrects the bias introduced by this prioritization. The Huber loss parameter δ is set to 1.0 to balance the squared error loss for small errors and absolute error loss for large errors (outliers).

TABLE I. DRL TRAINING HYPERPARAMETERS

Parameter	Symbol	Value	Description
Learning rate	η	0.0001	Adam optimizer step size
Discount factor	γ	0.99	Importance of future rewards
Batch size	B	64	Number of samples per training step
Replay buffer size	N	100,000	Maximum capacity of experience replay
PER priority exponent	α	0.6	Controls the degree of prioritization
PER importance sampling	β	0.4 → 1.0	Annealed linearly over training episodes
Huber loss delta	δ	1.0	Threshold for quadratic/linear transition
Exploration rate	ε	1.0 → 0.01	Epsilon-greedy decay strategy

Furthermore, the ARS mechanism is formulated to guide the agent toward an optimal policy by penalizing false alarms more heavily than missed detections, which is critical for edge-based security. The reward function R_t at time step t is defined as shown in (1):

$$R_t = \begin{cases} +1, & \text{if } A^t = L^t \text{ (Correct Classification)} \\ -1, & \text{if } A^t \neq L^t \text{ Attack (False Negative)} \\ -1(1+\rho), & \text{if } A^t \neq L^t \text{ Normal (False Positive)} \end{cases} \quad (1)$$

where A^t is the action taken, L^t is the true label, and ρ (set to 0.5) is an additional penalty factor to suppress false positives, ensuring that the model prioritizes service availability on the IoT network.

E. System Testing

The system testing was designed to assess three main aspects, namely classification accuracy, learning stability, and system performance on edge devices with limited resources. In terms of classification, the evaluation was conducted using several performance metrics, including accuracy, precision, recall, F1-score, and FPR. These metrics were used to assess the effectiveness of the system in detecting DDoS attacks on dynamic IoT traffic. The learning stability aspect was tested by monitoring changes in rewards in each training episode. This analysis is divided into three main phases: the initial phase (episodes 1–8) to observe the learning process for known attack patterns, the transition phase (episodes 9–10) where the model is exposed to previously unseen zero-day attack scenarios, and the adaptation phase (episodes 11–15) to assess the model's ability to adjust its detection strategy after exposure to new attacks. The reward change patterns in these three phases were used to evaluate the model's adaptability. In addition, aspects of system performance were tested based on CPU usage (%), memory consumption (MB), and inference latency (ms) to ensure the feasibility of implementation on edge devices.

Testing was conducted on several variants of DRL algorithms, namely Q-Learning, DQN, and Dueling DQN, both in their standard forms and after applying performance-improvement mechanisms such as PER, ARS, and Huber loss. For comparison, conventional supervised learning models such as Extra Trees, Random Forest, LightGBM, XGBoost, Gradient Boosting, Decision Tree, AdaBoost, Logistic Regression, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Ridge Classifier, Naive Bayes, and Quadratic Discriminant Analysis (QDA) were also included to identify differences in characteristics between static and adaptive approaches. In the final stage, the trained and optimized models were tested on edge devices (Raspberry Pi 3) to evaluate operational efficiency and inference stability under resource-constrained conditions, simulating real-world deployment in an IoT environment.

To evaluate cross-dataset generalization capability, the trained DRL model was further tested on multiple publicly available intrusion detection datasets, namely UNSW-NB15, Bot-IoT, and IoT-23. In this evaluation protocol, the model was trained exclusively on the IoT-DH dataset, after which the network parameters were frozen and directly applied to each public dataset without retraining or fine-tuning. To ensure consistency, preprocessing transformations learned from the

IoT-DH dataset were applied uniformly to the external datasets. Model performance was evaluated using accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) on each dataset. Performance degradation relative to IoT-DH was recorded as an indicator of distribution shift robustness and analyzed from a methodological perspective, considering differences in traffic characteristics, protocol composition, and class imbalance.

To ensure methodological clarity, the zero-day attack scenario in this study does not refer to synthetic or artificially generated traffic. Instead, zero-day conditions are constructed by deliberately withholding specific attack patterns during the training phase and introducing them only during the testing stage. In particular, the dataset is temporally and behaviorally partitioned such that certain DDoS attack variants (denoted as Attack-B) are completely excluded from the training data and first exposed to the model during the evaluation phase. This protocol simulates realistic zero-day conditions, where detection systems encounter previously unseen attack behaviors rather than randomly perturbed samples.

The zero-day dataset is derived from real-world traffic captured by the RIoTPot honeypot deployment. Variations in attack behavior include differences in packet rate, burst patterns, protocol targeting (e.g., MQTT/CoAP), and temporal characteristics. By introducing these withheld attack patterns during evaluation, the testing process emulates abrupt distribution shifts commonly observed in operational IoT environments. This design allows the assessment of both performance degradation in static supervised models and adaptive learning behavior in DRL-based approaches.

F. Reporting Results

The results of all testing scenarios were thoroughly analyzed and presented to address the research questions and demonstrate the effectiveness of the proposed adaptive DRL approach. This included a comparative analysis of various model variants, the performance of the system in zero-day attack scenarios, and an assessment of the model's operational efficiency on edge devices. Additionally, ablation tests were conducted to evaluate the contribution of each performance-improvement component (PER, ARS, Huber loss) to overall system performance. The findings were discussed in terms of detection accuracy, learning stability, inference efficiency, and adaptability, thus validating the proposed model's capacity to effectively handle dynamic IoT traffic and emerging attack patterns.

III. RESULTS AND DISCUSSION

A. Baseline Performance of Static Models

Zero-day experiments reveal a significant performance gap between training data (in-distribution) and test data containing new attack patterns. The initially high accuracy on training data sharply declines when faced with different traffic distributions, confirming that the supervised approach heavily relies on historical distributions. The heterogeneity of IoT protocols further exacerbates this performance degradation, rendering static approaches unable to keep pace with the dynamics of new attacks. Figure 5 provides empirical evidence that supervised models are less adaptive, highlighting the need for more flexible detection mechanisms against zero-day attacks. This performance degradation under zero-day conditions directly answers the first research question, demonstrating that static supervised models are unable to maintain detection reliability when IoT traffic distributions shift.

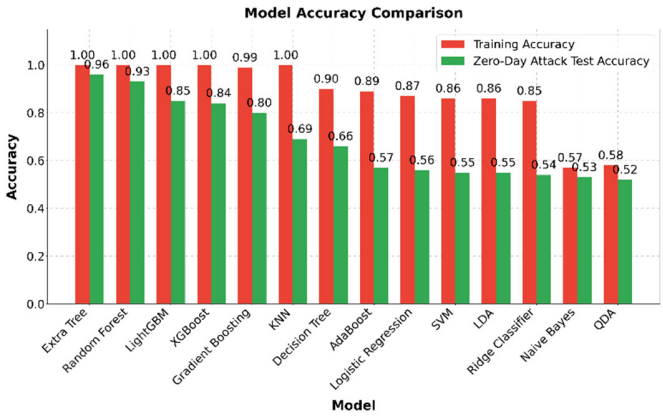

Fig. 5. Comparison of machine learning model accuracy on training data and zero-day attack test.

B. Core Performance of the Adaptive Deep Reinforcement Learning Framework

The core performance evaluation compares several DRL variants under the in-distribution scenario using accuracy, precision, recall, and F1-score, as shown in Figure 6.

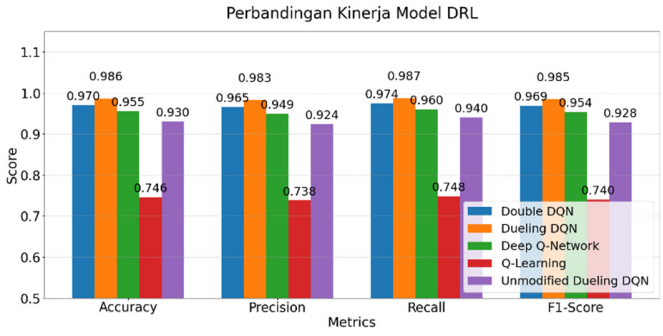

Fig. 6. Performance metrics comparison chart between DRL algorithms.

Q-Learning exhibits the lowest performance, with accuracy around 0.75 and F1-score below 0.75, indicating limited capability in modeling complex IoT traffic behavior. DQN significantly improves performance, achieving accuracy and precision above 0.95, although noticeable variation remains across recall and F1-score. Double DQN further stabilizes performance by mitigating value overestimation, resulting in consistently high scores across all metrics, with accuracy and F1-score reaching approximately 0.97. The unmodified Dueling DQN shows improved recall (≈ 0.94), reflecting better attack sensitivity, but its precision and F1-score remain slightly lower than Double DQN. The best overall performance is achieved by the performance-enhanced Dueling DQN, which integrates PER, ARS, and Huber loss. This model consistently

attains the highest scores across all metrics, with accuracy, precision, recall, and F1-score all approaching or exceeding 0.98. These results demonstrate a well-balanced detection capability, effectively combining high sensitivity with reduced false alarms, making the proposed model the most stable and suitable for deployment on edge-based IoT security systems. From a system design perspective, this balance between high recall and low FPR is essential for edge-based IDSs, as excessive false alarms can overwhelm limited computational and operational resources in IoT deployments.

C. Cross-Dataset Generalization Evaluation

To evaluate the generalization capability of the proposed DRL framework, the trained model was further tested on multiple publicly available intrusion detection datasets, namely UNSW-NB15, Bot-IoT, and IoT-23, without retraining or fine-tuning. As shown in Table II, the model maintains consistently high performance across all datasets, with F1-scores remaining above 0.96 despite substantial differences in traffic characteristics, protocol composition, and attack distributions.

TABLE II. CROSS-DATASET GENERALIZATION PERFORMANCE OF THE PROPOSED DRL MODEL

Dataset	Accuracy	Precision	Recall	F1-score
IoT-DH	0.982	0.978	0.985	0.981
UNSW-NB15	0.982	0.981	0.980	0.980
Bot-IoT	0.975	0.974	0.972	0.973
IoT-23	0.968	0.970	0.965	0.967

These results directly address the research question concerning robustness and adaptability, demonstrating that the proposed DRL framework does not overfit to the primary IoT-DH dataset. Instead, it learns generalized detection policies that remain effective under cross-dataset distribution shifts. The modest performance degradation observed on Bot-IoT and IoT-23 reflects realistic domain differences rather than model instability.

In contrast to conventional supervised learning models, which typically suffer sharp performance drops when evaluated across different datasets, the proposed DRL-based approach preserves stable detection capability by relying on adaptive policy learning rather than fixed decision boundaries. This behavior further supports the suitability of the proposed framework for real-world IoT deployments, where traffic patterns and attack behaviors continuously evolve.

D. Learning Dynamics of Reward Behavior in Adaptive Deep Reinforcement Learning Models

The learning stability analysis focused on the best model, namely performance-enhanced Dueling DQN, because this variant had previously been proven to provide the highest and most stable performance among all DRL algorithms tested. The evaluation was carried out by monitoring changes in reward in each training episode, as shown in Figure 7.

In the initial phase (episodes 1–8), the reward increased steadily, indicating that the agent rapidly converges toward a stable baseline detection policy when exposed to known IoT traffic patterns. This behavior is critical for edge-based IDSs, as it demonstrates that the proposed reward design enables fast policy stabilization, thereby minimizing false alarms during early deployment. This indicates the model's basic ability to associate network conditions with optimal detection actions. Next, in the transition phase (episodes 9–10), the reward decreased sharply when the zero-day attack scenario was introduced. This reward degradation reflects a realistic detection challenge under zero-day conditions, where previously optimal policies become temporarily suboptimal due to abrupt distribution shifts. Importantly, this behavior highlights the limitation of static detection strategies and motivates the need for adaptive policy updates in real-world IoT environments. Entering the adaptation phase (episodes 11–15), the reward increased again and reached a new stability. This recovery and subsequent stabilization of rewards demonstrate that the proposed DRL framework can autonomously adapt its detection policy without explicit retraining, thereby validating its suitability for handling non-stationary IoT traffic and real zero-day attack scenarios. In other words, the performance-enhanced Dueling DQN successfully adjusts the detection policy based on environmental feedback (reward feedback) without requiring explicit retraining.

Fig. 7. Reward dynamics of the optimized Dueling DQN model.

In addition to the observed reward fluctuations, a closer inspection of misclassification behavior indicates that false positives predominantly occur during the early exposure to zero-day attack scenarios. At this stage, certain benign IoT traffic patterns—particularly bursty normal traffic generated by protocols such as MQTT and CoAP—exhibit statistical similarities to early-stage DDoS attacks, including elevated packet rates and temporal density. As a result, the model initially tends to adopt conservative detection decisions, temporarily labeling such traffic as malicious. This behavior reflects the inherent ambiguity of real-world IoT traffic rather than model instability. As learning progresses, adaptive mechanisms gradually refine the decision policy, leading to a reduction in false positives and a more balanced trade-off between sensitivity and false alarm rate.

Overall, the increase in reward after the decline proves that the optimized Dueling DQN has strong self-adaptive learning capabilities. This characteristic makes it the most promising model for detecting and responding to the ever-changing dynamics of cyberattacks in an edge computing-based IoT ecosystem.

E. Ablation Study on Prioritized Experience Replay, Adaptive Reward Shaping, and Huber Loss

The ablation study results in Table III highlight the contribution of each performance-improvement component to the performance of Dueling DQN. PER accelerates convergence time and slightly improves the F1-score; ARS effectively reduces FPR despite an initial reduction in precision; and Huber loss plays a crucial role in maintaining stability during weight updates. When all three are combined, the performance improvement is most significant in terms of both F1-score enhancement and FPR reduction. This positive synergy confirms that the weaknesses of one technique can be mitigated by the strengths of another, resulting in a more adaptive and efficient model. These findings underscore the relevance of implementing DRL-based IDSs at the edge, which demand high accuracy, low false alarm rates, and real-time efficiency.

TABLE III. RESULTS OF THE ABLATION STUDY ON DUELING DQN PERFORMANCE-IMPROVEMENT COMPONENTS

Scheme	Δ F1-score	Δ FPR	Convergence time (episodes)
Dueling DQN (baseline)	0.00	0.00	100%
+ PER	+0.03	–0.02	92%
+ ARS	+0.02	–0.06	95%
+ Huber loss	+0.04	–0.03	90%
+ PER + ARS + Huber (full)	+0.10	–0.09	82%

F. Edge Inference Efficiency and Resource Utilization

The evaluation of resource utilization efficiency on the edge device focuses on memory consumption and CPU utilization patterns during the DRL inference process. As shown in Figure 8, memory usage exhibits a slight increase during the initial execution phase, rising from approximately 863 MB to 866 MB, which corresponds to model initialization, parameter loading, and internal data structure allocation. After reaching a steady-state condition, memory consumption remains stable with no noticeable fluctuations, indicating that the inference process operates without progressive memory growth or memory leakage. The sharp memory drop observed at the final index is associated with process termination rather than inference degradation, confirming that the DRL model can be executed continuously on memory-constrained edge devices.

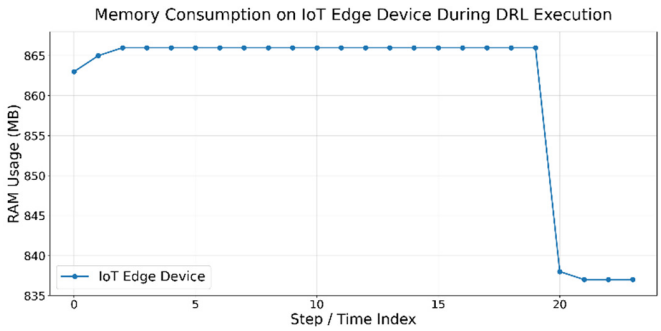

Fig. 8. Memory consumption on IoT edge device during DRL inference execution.

Meanwhile, the CPU utilization pattern during inference, illustrated in Figure 9, shows a fluctuating profile characterized by intermittent spikes reaching up to approximately 2.4 CPU cores, corresponding to active inference computations. These spikes are short-lived and followed by extended low-utilization periods, indicating that the inference workload does not cause sustained CPU saturation. This utilization pattern confirms that the proposed DRL-based detection framework can coexist with other edge services without inducing resource contention, thereby strengthening its feasibility for continuous, real-time deployment in operational IoT environments.

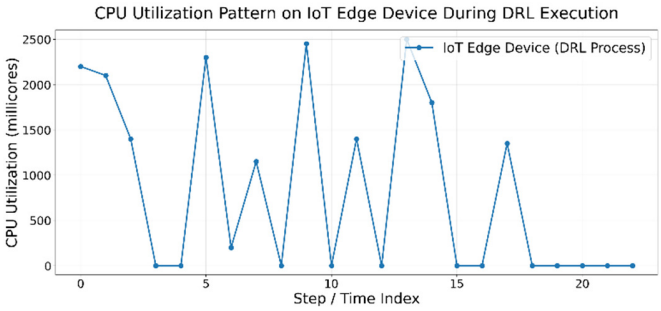

Fig. 9. CPU utilization pattern on IoT edge device during DRL inference execution.

Furthermore, the applied quantization strategy yielded significant performance gains on the Raspberry Pi 3. Specifically, the average inference latency was reduced by approximately 14% (improving from 4.2 ms to 3.6 ms per sample) compared to the non-quantized baseline. Crucially, this efficiency improvement was achieved with a negligible accuracy degradation of less than 0.5%, confirming that the framework maintains robust detection capabilities even under resource-constrained conditions.

G. Comparison with Previous Studies

Most existing studies on IoT DDoS detection predominantly rely on static supervised or rule-based learning paradigms, which demonstrate high accuracy under controlled and in-distribution datasets but experience substantial performance degradation when exposed to evolving or previously unseen attack patterns. As summarized in Table IV, prior works [7, 8] confirm this limitation, as their detection models depend on fixed decision boundaries and lack mechanisms for adaptive policy updates under non-stationary IoT traffic.

Lightweight deep learning approaches [12] improve inference efficiency but remain constrained by offline training paradigms, limiting robustness against zero-day attacks. Similarly, resource-efficient machine learning techniques that emphasize feature optimization and computational feasibility [16]—also summarized in Table IV—achieve practical deployment goals but sacrifice adaptability, resulting in reduced resilience to distribution shifts commonly observed in real-world IoT environments.

In contrast, the proposed framework differs fundamentally in three aspects: (i) it employs DRL to enable continuous

policy adaptation based on environmental feedback, (ii) it conducts explicit zero-day evaluation using real, withheld attack patterns rather than synthetic perturbations, and (iii) it validates practical edge deployment on a Raspberry Pi 3. As shown in Table IV, unlike prior DRL-focused studies that remain largely conceptual or simulation-based [15], this work provides end-to-end empirical validation, positioning the proposed method as a more realistic and deployment-ready benchmark for adaptive IoT DDoS detection. From a technical perspective, supervised classifiers rely on fixed feature-to-label mappings optimized during offline training, making them inherently incapable of updating decision boundaries when traffic statistics, protocol behavior, or attack intensity shift over time.

TABLE IV. COMPARATIVE ANALYSIS WITH PREVIOUS STUDIES ON IOT DDOS DETECTION

Study	Learning paradigm	Dataset type	Zero-day evaluation	Edge deployment	Best reported F1-score
[7]	Rule-based + ML	Public / simulated	No	No	0.92
[8]	Supervised ML	Public IoT	No	Partial	0.93
[12]	Lightweight DL	Public IoT	Limited	No	0.95
[15]	DRL (survey)	Simulation-based	Conceptual	No	–
[16]	Feature-based ML	Public IoT	No	Yes	0.91
This work	Adaptive DRL (Dueling DQN + PER + ARS + Huber)	Real honeypot (IoT-DH)	Yes (withheld real attacks)	Yes (Raspberry Pi 3)	0.98–0.99

ML=Machine Learning, DL=Deep Learning.

H. Implications and Findings

The final discussion of this study confirms several empirical implications that distinguish it from previous research. This study builds a primary dataset (IoT-DH) through the RIoTPot honeypot in a real IoT environment, rather than relying on older public datasets. This approach ensures higher data relevance and validity by reflecting current traffic. Methodologically, the proposed framework demonstrates that DRL-based models can adapt to heterogeneous IoT traffic dynamics and new attacks, unlike conventional machine learning methods, which are static and prone to performance degradation in zero-day scenarios. Architecturally, a hybrid strategy that involves training on high-performance servers and performing inference at the gateway/edge has proven effective in maintaining efficiency without sacrificing accuracy.

Another important finding is that the combination of performance-improvement mechanisms used (such as PER, ARS, and Huber loss) positively impacts convergence speed and learning stability. Technical details regarding the contribution of each mechanism are described in the results section; thus, it suffices to emphasize here that the synergy of these three mechanisms results in a more robust and reliable model for DDoS detection on limited devices. Furthermore, implementation test results demonstrate that the model can be deployed on real-world edge devices with limited resources, such as the Raspberry Pi 3, after undergoing distillation and quantization processes. This validates the feasibility of DRL-based adaptive detection systems for direct application at the edge layer, reducing dependence on cloud computing and accelerating response times.

Future development efforts can focus on expanding coverage of non-IP protocols such as AMQP or MODBUS, which are widely used in industrial environments, implementing online learning to allow DRL agents to adapt directly to actual traffic, and integrating automatic mitigation modules capable of blocking or limiting attacks in real-time. Thus, this research opens new directions for developing adaptive, lightweight IDS solutions in future IoT environments.

IV. CONCLUSION

This study addresses a fundamental limitation of existing Internet of Things (IoT) Distributed Denial of Service (DDoS) detection systems, which rely on static supervised learning and consequently degrade under non-stationary traffic distributions and zero-day attack conditions. To overcome this gap, we propose an adaptive Deep Reinforcement Learning (DRL)-based framework tailored for resource-constrained edge devices, integrating a Dueling Deep Q-Network (DQN) architecture with Prioritized Experience Replay (PER), Adaptive Reward Shaping (ARS), and Huber loss to enhance stability, adaptability, and false alarm control. Trained on a real-world honeypot-derived dataset (IoT-DH) and evaluated using withheld genuine attack patterns, the proposed approach achieves F1-scores of 0.98–0.99 on in-distribution data and maintains robust cross-dataset performance above 0.96 on UNSW-NB15, Bot-IoT, and IoT-23 without retraining, outperforming static supervised models and baseline DRL variants. Learning dynamics analysis further confirms that the agent autonomously adapts its detection policy after exposure to zero-day attacks, recovering stable performance without offline retraining.

Deployment on a Raspberry Pi 3 demonstrates practical feasibility, with quantization reducing inference latency by approximately 14% while preserving accuracy and stable resource utilization. Unlike prior studies that remain cloud-centric, simulation-based, or reliant on fixed decision boundaries, this work provides end-to-end validation of adaptive DRL-based DDoS detection using real honeypot data, explicit zero-day evaluation, and edge-level deployment, positioning the proposed framework as a practical and scalable solution for securing evolving IoT environments.

ACKNOWLEDGMENT

Funding for this work was provided by the Center for Education Financial Services (Puslapdik) and the Indonesia Endowment Funds for Education (LPDP) under Number 00787/J5.2.3./BPI.06/9/2022. The authors are very grateful to our university, Infokom UMM, and to Salman Alfarisi, Aisah Alfiyatus Ramadhanti, Romadhon Wiratama, Larynt Sawfa Kenanga, and Fath' Hana Sarla Bajsair for their assistance in completing this research.

REFERENCES

[1] N. Sharma, M. Shamkuwar, and I. Singh, "The History, Present and Future with IoT," in *Internet of Things and Big Data Analytics for Smart Generation*, V. E. Balas, V. K. Solanki, R. Kumar, and M. Khari, Eds. Cham, Switzerland: Springer International Publishing, 2019, pp. 27–51, https://doi.org/10.1007/978-3-030-04203-5_3.

[2] S. Sinha. "Number of connected IoT devices growing 14% to 21.1 billion." IoT Analytics. https://iot-analytics.com/number-connected-iot-devices/.

[3] M. Gelgi, Y. Guan, S. Arunachala, M. S. S. Rao, and N. Dragoni, "Systematic Literature Review of IoT Botnet DDOS Attacks and Evaluation of Detection Techniques," *Sensors*, vol. 24, no. 11, June 2024, Art. no. 3571, https://doi.org/10.3390/s24113571.

[4] K. Yang, D. Blaauw, and D. Sylvester, "Hardware Designs for Security in Ultra-Low-Power IoT Systems: An Overview and Survey," *IEEE Micro*, vol. 37, no. 6, pp. 72–89, Nov. 2017, https://doi.org/10.1109/MM.2017.4241357.

[5] C. Tagliaro, M. Komsic, A. Continella, K. Borgolte, and M. Lindorfer, "Large-Scale Security Analysis of Real-World Backend Deployments Speaking IoT-Focused Protocols," in *Proceedings of the 27th International Symposium on Research in Attacks, Intrusions and Defenses*, Padua, Italy, 2024, pp. 561–578, https://doi.org/10.1145/3678890.3678899.

[6] M. A. Shyaa, N. F. Ibrahim, Z. Zainol, R. Abdullah, M. Anbar, and L. Alzubaidi, "Evolving cybersecurity frontiers: A comprehensive survey on concept drift and feature dynamics aware machine and deep learning in intrusion detection systems," *Engineering Applications of Artificial Intelligence*, vol. 137, Nov. 2024, Art. no. 109143, https://doi.org/10.1016/j.engappai.2024.109143.

[7] A. Hussain, E. Marín Tordera, X. Masip-Bruin, and H. C. Leligou, "Rule-Based With Machine Learning IDS for DDoS Attack Detection in Cyber-Physical Production Systems (CPPS)," *IEEE Access*, vol. 12, pp. 114894–114911, 2024, https://doi.org/10.1109/ACCESS.2024.3445261.

[8] R. Alsulami, B. Alqarni, R. Alshomrani, F. Mashat, and T. Gazdar, "IoT Protocol-Enabled IDS based on Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12373–12380, Dec. 2023, https://doi.org/10.48084/etasr.6421.

[9] Z. Deng, "Developing machine learning-based intrusion detection systems for IoT environments," in *Joint Proceedings of the BIR 2025 Workshops and Doctoral Consortium co-located with 24th International Conference on Perspectives in Business Informatics Research*, Riga, Latvia, 2025, pp. 246–255.

[10] B. R. Kikissagbe and M. Adda, "Machine Learning-Based Intrusion Detection Methods in IoT Systems: A Comprehensive Review," *Electronics*, vol. 13, no. 18, Sept. 2024, Art. no. 3601, https://doi.org/10.3390/electronics13183601.

[11] X. Lin, "A Survey of AI-Based Zero-Day Attack Detection Methods," *Applied and Computational Engineering*, vol. 164, pp. 171–179, July 2025, https://doi.org/10.54254/2755-2721/2025.PO25664.

[12] R. H. Altaie and H. K. Hoomod, "An Intrusion Detection System using a Hybrid Lightweight Deep Learning Algorithm," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16740–16743, Oct. 2024, https://doi.org/10.48084/etasr.7657.

[13] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized Experience Replay," in *4th International Conference on Learning Representations*, San Juan, Puerto Rico, 2016, https://doi.org/10.48550/arXiv.1511.05952.

[14] M. Hessel *et al.*, "Rainbow: Combining Improvements in Deep Reinforcement Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, pp. 3215–3222, Apr. 2018, https://doi.org/10.1609/aaai.v32i1.11796.

[15] S. Jamshidi, A. Nikanjam, K. W. Nafi, F. Khomh, and R. Rasta, "Application of deep reinforcement learning for intrusion detection in Internet of Things: A systematic review," *Internet of Things*, vol. 31, May 2025, Art. no. 101531, https://doi.org/10.1016/j.iot.2025.101531.

[16] N. Gavric, G. Prasad Bhandari, and A. Shalaginov, "Towards Resource-Efficient DDoS Detection in IoT: Leveraging Feature Engineering of System and Network Usage Metrics," *Journal of Network and Systems Management*, vol. 32, no. 4, Aug. 2024, Art. no. 69, https://doi.org/10.1007/s10922-024-09848-2.

[17] S. Saif, W. Widyawan, and R. Ferdiana, "IoT-DH dataset for classification, identification, and detection DDoS attack in IoT," *Data in Brief*, vol. 54, June 2024, Art. no. 110496, https://doi.org/10.1016/j.dib.2024.110496.

[18] W. Fedus *et al.*, "Revisiting Fundamentals of Experience Replay," in *Proceedings of the 37th International Conference on Machine Learning*, Vienna, Austria, 2020, pp. 3061–3071.

[19] A. K. Cook *et al.*, "Multiple-Choice Questions in Small Animal Medicine: An Analysis of Cognitive Level and Structural Reliability, and the Impact of these Characteristics on Student Performance," *Journal of Veterinary Medical Education*, vol. 47, no. 4, pp. 497–505, July 2020, https://doi.org/10.3138/jvme.0918-116r.

[20] M. Chahoud *et al.*, "Reward shaping in DRL: A novel framework for adaptive resource management in dynamic environments," *Information Sciences*, vol. 715 Oct. 2025, Art. no. 122238, https://doi.org/10.1016/j.ins.2025.122238.