# Hybrid Quantum-Classical RL with Prioritized Replay for IoT Intrusion Detection

Collins Izuchukwu Okafor [1], Love Allen Chijioke Ahakonye [2], Dong-Seong Kim [1] *, Jae Min Lee [1]
[1] IT-Convergence Engineering, *Kumoh National Institute of Technology*, Gumi, South Korea
* NSLab Co. Ltd., Gumi, South Korea, *Kumoh National Institute of Technology*, Gumi, South Korea
[2] ICT Convergence Research Center, *Kumoh National Institute of Technology*, Gumi, South Korea
(collinsokafor, loveahakonye, dskim, ljmpaul)@kumoh.ac.kr

***Abstract*—The rise of IoT devices demands robust intrusion detection systems (IDS) to combat cyber threats. This study introduces a hybrid quantum-classical RL framework using a dueling DQN with prioritized experience replay (PER) to detect intrusions in IoT networks. Our system effectively identifies anomalous traffic patterns by integrating quantum circuits for feature extraction and PER for efficient learning. Simulations show faster convergence and a 15% higher cumulative reward compared to classical DQN baselines, highlighting the potential of quantum-enhanced RL for securing IoT ecosystems.**

***Index Terms*—IoT, Prioritized Experience Replay, Quantum Computing, Reinforcement Learning, Security.**

## I. Introduction

IoT connectivity enables efficient data exchange but introduces security risks from compromised devices generating anomalous traffic [1]. Traditional IDS methods, such as rule-based or supervised learning, lack adaptability to emerging threats, making data-driven solutions essential. Reinforcement learning (RL) offers autonomous IDS via trial-and-error optimization. Deep Q-Networks (DQNs) address high-dimensional state spaces but face challenges such as sample inefficiency and slow convergence. Studies have applied classical DQN for IoT security [2], while Kalinin et al. [3] explored quantum-enhanced RL. This study builds on these by integrating dueling DQN, quantum circuits, and prioritized experience replay (PER), with quantum RL improving exploration efficiency for network security [3], [4].

This study proposes a hybrid quantum-classical reinforcement learning (RL) approach that integrates dueling DQN with prioritized experience replay (PER) for intrusion detection in the IoT. The system leverages quantum circuits for observation processing and PER to enhance learning by prioritizing critical experiences. A custom IoT intrusion scenario with refined reward structures validates the approach, showing superior detection performance compared to classical baselines.

## II. System Methodology

This section presents the system framework and formulates the intrusion detection problem within IoT networks using a hybrid quantum-classical reinforcement learning (RL) approach with prioritized experience replay (PER), as shown in Figure 1.

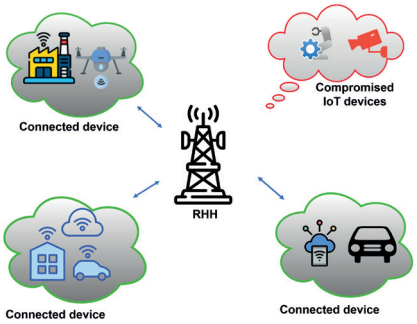


Fig. 1: Flow of quantum prioritized replay intrusion detection in IoT network connected to a remote radio head (RRH).

### A. IoT Intrusion Environment

The $IoTIntrusionEnv$ simulates a network of 20 devices, each generating four traffic features and an energy metric, represented as a 100-dimensional state vector. At initialization, 1-2 devices are randomly compromised, with traffic increased by 300–600 units, feature 2 doubled, and feature 3 halved, mimicking botnet-like behavior. The action space consists of 0, representing no action, for device blocking and 2 for isolating a device. The reward function $R_f$ is defined as

$$R_f = \begin{cases} +10 & \text{positive action taken,} \\ -5 & \text{unnecessarily action taken,} \\ -2 \times N_c & \text{no action taken with } N_c, \\ +1 & \text{no compromised devices, no action} \end{cases}$$

where $N_c$ is compromised device. This structure encourages accurate detection and mitigation.

### B. Hybrid Quantum-Classical Model

The agent combines classical and quantum components with a $DuelingQIntrusionDetector$. Initially, a classical pre-network maps a 100-dimensional state vector to a 4-dimensional representation via two fully connected layers (16 and 4 neurons, ReLU activation). A quantum layer processes this reduced representation, implemented using a 4-qubit circuit with PennyLane, which applies angle embedding and entangling layers to output Pauli-Z expectation values. These features are passed through two dueling streams (value and advantage) with 16 neurons to compute Q-values, which

enhances action selection stability. Training is performed on a CUDA-enabled device using the Adam optimization algorithm with a learning rate of $\eta = 10^{-3}$.

### C. Prioritized Experience Replay

A $PrioritizedReplayBuffer$ (capacity 5000, $\alpha = 0.6$) stores experiences $(s, a, r, s', d)$, sampling batches (size 64) based on TD-error priorities. Importance sampling weights ($\beta = 0.4$, annealed to 1.0) mitigate bias, aligning with [5] findings on PER efficiency in RL. The `QuantumDQNAgent` uses an $\epsilon$-greedy policy ($\epsilon$ decays from 1.0 to 0.1 at 0.9995 per step). Target network updates occur every 5 episodes, and two training updates per step optimize the policy network via the loss $L = \mathbb{E}\left[w_i \left(Q(s,a) - (r + \gamma \max_{a'} Q'(s',a'))\right)^2\right]$, where $w_i$ are PER weights, gradient norms are clipped to 1.0.

## III. Performance Evaluation

We evaluate the performance of the Hybrid QE-DRL algorithm by comparing its energy efficiency with three baseline methods: Qe-DRL [6], JRSPA [7], and JCORA [8]. As shown in Figure 2, the Hybrid QE-DRL significantly outperforms the others over 200 training episodes, starting at 0.4 kbps/Hz/J and stabilizing around 1.5–1.6 kbps/Hz/J after 50 episodes. The Qe-DRL algorithm [6] reaches 0.6 kbps/Hz/J by episode 50 and stabilizes around 1.2–1.4 kbps/Hz/J, with oscillations. JRSPA [7] stabilizes around 1.0–1.2 kbps/Hz/J, while JCORA [8] achieves the lowest, stabilizing around 0.8–1.0 kbps/Hz/J. The Hybrid QE-DRL's superior performance highlights its effectiveness in optimizing energy efficiency and stability.

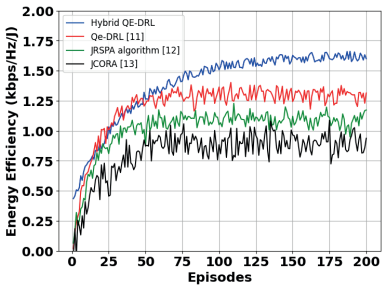


Fig. 2: Energy efficiency comparison of the proposed Hybrid QE-DRL algorithm vs Qe-DRL vs JRSPA vs JCORA.

Figure 3 shows the training loss over 6000 updates, spanning 200 episodes (2 updates per step, 30 steps per episode). Initially high ( 20), the loss drops quickly within the first 1000 updates, indicating effective early learning. It then stabilizes and fluctuates between 2 and 10, reflecting convergence to a stable policy as it adapts to dynamic conditions. These oscillations highlight ongoing parameter fine-tuning. The Hybrid QE-DRL algorithm demonstrates improved energy efficiency and stable convergence, making it a promising solution for resource-constrained IoT networks.

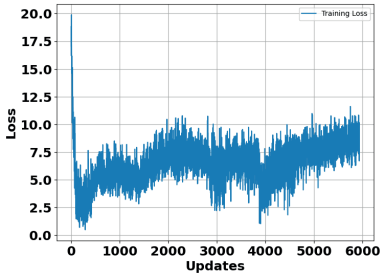


Fig. 3: Training loss progression over 6000 updates.

## IV. Conclusion

This study proposed the Hybrid QE-DRL algorithm, which achieved superior energy efficiency of 1.5–1.6 kbps/Hz/J compared to Qe-DRL (1.2–1.4), JRSPA (1.0–1.2), and JCORA (0.8–1.0 kbps/Hz/J) over 200 episodes, with rapid loss convergence from 20 to 2–10 and reward improvement from -120 to 0–20, demonstrating its effectiveness for IoT networks. Future work includes validating the model with real-world IoT datasets, testing its scalability on heterogeneous networks, and optimizing it for both latency and security.

## Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the IITP grant funded by the Korea government(MSIT) (IITP-2025-RS-2020-II201612, 33%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 33%) and by the MSIT, Korea, under the ITRC support program(IITP-2025-RS-2024-00438430, 34%)

## References

[1] L. A. C. Ahakonye, C. I. Nwakanma, J. M. Lee, and D. Kim, "AI Model Stability in Industrial IoT Intrusion Detection: Leveraging the Characteristics Stability Index," *The Journal of Korean Institute of Communications and Information Sciences*, vol. 49, no. 2, pp. 321–331, 2024.

[2] O. Kayode and A. S. Tosun, "Deep q-network for enhanced data privacy and security of iot traffic," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, 2020, pp. 1–6.

[3] M. Kalinin and V. Krundyshev, "Security intrusion detection using quantum machine learning techniques," *Journal of Computer Virology and Hacking Techniques*, vol. 19, no. 1, pp. 125–136, Mar 2023. [Online]. Available: https://doi.org/10.1007/s11416-022-00435-0

[4] E. A. Tuli, J.-M. Lee, and D.-S. Kim, "Leveraging quantum blockchain for secure multiparty space sharing and authentication on specialized metaverse platform," *Scientific Reports*, vol. 14, no. 1, p. 25776, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-74213-x

[5] K. Zhao, Y. Wang, Y. Chen, Y. Li, L. H. U, and X. Niu, "Efficient diversity-based experience replay for deep reinforcement learning," 2025. [Online]. Available: https://arxiv.org/abs/2410.20487

[6] J. A. Ansere, E. Gyamfi, V. Sharma, H. Shin, O. A. Dobre, and T. Q. Duong, "Quantum deep reinforcement learning for dynamic resource allocation in mobile edge computing-based iot systems," *IEEE Transactions on Wireless Communications*, vol. 23, no. 6, pp. 6221–6233, 2024.

[7] A. Jaiswal, S. Kumar, O. Kaiwartya, P. K. Kashyap, E. Kanjo, N. Kumar, and H. Song, "Quantum learning-enabled green communication for next-generation wireless systems," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1015–1028, 2021.

[8] M. Tang and V. W. Wong, "Deep reinforcement learning for task offloading in mobile edge computing systems," *IEEE Transactions on Mobile Computing*, vol. 21, no. 6, pp. 1985–1997, 2022.