International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =7.938, July-December 2024, Submitted in December 2024, ISSN -2393-8048

Deep Reinforcement Learning-Based Dynamic Cryptographic Configuration for Energy-Aware IoT Security

Sohail Hassan (Research Scholar), Computer Science Dept., Sikkim Alpine University, Namchi (Sikkim)
Dr. Sunil Gupta (Associate Professor), Computer Science Dept., Sikkim Alpine University, Namchi (Sikkim)
Mail ID: mailsohailhassan@gmail.com

ABSTRACT

Standard algorithms like AES-128 offer robust and widely-accepted security, but they place a heavy computational and energy burden on endpoints with limited resources, such as Internet of Things (IoT) devices with limited power sources. On the other hand, lightweight alternatives like the ASCON, SPECK, and PRESENT ciphers, which are standardized by NIST, minimize this overhead but have a smaller security margin. Current Internet of Things (IoT) deployments usually deal with this conflict by using a static cryptographic scheme that is applied consistently regardless of the device's real-time battery status or the current threat environment. This approach can be either dangerously under-secured or wastefully over-secured, depending on the situation. In this paper, a framework based on Deep Reinforcement Learning (DRL) is suggested for real-time cryptographic algorithm selection and configuration for individual Internet of Things (IoT) devices. The selection problem is modeled as a Markov Decision Process, where the agent's state includes battery level, prevailing threat level, and data sensitivity. The agent's action is to choose from a pool of candidate standardized ciphers (AES-128, ASCON-128, SPECK-64/128, PRESENT-80) that cover a range of energy-security trade-off points. A Deep Q-Network (DQN) agent was taught to maximize a reward function that penalizes both energy usage and security-policy breaches in a simulated Internet of Things (IoT) environment. In comparison to a static AES-128 baseline, the trained agent significantly reduced mean per-operation energy consumption by 34.7% while still complying with security policies in 96.2% of high-threat episodes. This performance was much better than static AES-128 (complete compliance but highest energy cost) and static lightweight-only configuration (lowest energy cost but only 71.4% high-threat compliance). Based on these findings, dynamic cryptographic configuration based on DRL can significantly outperform traditional static configuration methods in terms of the energy-security trade-off that can be achieved in limited IoT deployments.

Keywords: deep reinforcement learning; lightweight cryptography; ASCON; IoT security; energy-aware computing; adaptive security; Deep Q-Network

1. Introduction

Many of the billions of sensing, acting, and communicating devices that make up the Internet of Things (IoT) are running on very limited resources, including memory, processing power, and, most importantly for energy-harvesting or battery-powered devices, available energy. These devices are used in smart homes, factories, farms, hospitals, and other infrastructure monitoring settings. It is crucial to secure the data these devices create and send, but the cryptographic primitives typically used for this purpose were not made with these limitations in mind. For example, when used continuously on low-power microcontrollers, the Advanced Encryption Standard (AES), the symmetric-encryption standard for general-purpose computing, imposes computational and energy overhead that can significantly reduce battery life. In response to this discrepancy, NIST embarked on a multi-year effort to standardize lightweight cryptography. In February 2023, the ASCON family of algorithms was chosen as the gold standard for lightweight authenticated encryption, joining other popular lightweight ciphers like SPECK, SIMON, and PRESENT, which reduce security margin and simplify internal structure in exchange for significantly lower energy and computational cost per operation.

An operational question that has garnered less attention than the algorithms' design arises from the availability of multiple standardized and well-studied cryptographic options spanning this energy-security trade-off: how should an individual IoT device decide, at any given moment, which cryptographic configuration to use, given that no single point on this trade-off curve is

optimal across all operating conditions? Strong, higher-overhead encryption is reasonable for a device with plenty of battery life in a low-threat environment, but it is not the best option when the battery is almost dead or when the threat level is high (e.g., due to detected abnormal network traffic or a security-sensitive data payload). Traditional Internet of Things (IoT) setups typically deal with this issue by relying on static, deployment-time configuration, which locks devices into a single cryptographic scheme for their entire operational lifetime. However, this method is not flexible enough to handle the significant variation in battery state and threat context that actual IoT devices confront during deployments.

To fill this need, this paper presents a Deep Reinforcement Learning (DRL) framework that can learn from simulated operational experience a dynamic cryptographic-configuration policy. This policy aims to minimize cumulative energy consumption while maintaining an acceptable, context-appropriate level of security assurance by selecting the cryptographic algorithm that is best suited to the device's current battery level, the prevailing threat level, and the sensitivity of the data being protected at each encryption event.

1.1 Objectives

There are two primary goals that are directing this research:

1. To create a system that uses Deep Reinforcement Learning, a Markov Decision Process, and a Deep Q-Network (DQN) agent to configure and select cryptographic algorithms for Internet of Things (IoT) devices in real time according to factors like data sensitivity, threat level, and battery life.
2. To simulate various running scenarios in order to assess the energy efficiency and compliance with security policies of the proposed DRL-based dynamic configuration framework in comparison to static (fixed-algorithm) cryptographic configuration baselines.

2. Literature Review

2.1 Lightweight Cryptography for Resource-Constrained IoT Devices

In *2015, Dinu et al.* investigated embedded devices with limited resources and looked at lightweight encryption algorithms for them. The objective was to evaluate ciphers based on the amount of resources they consume. The approach tested algorithms on AVR, MSP, and ARM platforms with the FELICS framework, taking into account execution time, RAM, and code size. The findings demonstrated that there was a large gap in cryptography performance across different types of hardware. As a result, real device resources and application needs should be considered when choosing lightweight algorithms.

In their *2020 study, Tiberti et al.* looked into WSN key management and lightweight cryptography. The objective was to offer trustworthy authentication while keeping computational and memory requirements to a minimum. This methodology tested several security schemes on sensor nodes running TinyOS, including TAKS2, RSA, ECC, and symmetric. In comparison to traditional public-key approaches, the results showed less memory requirements and less overhead. Lightweight security approaches work well for Internet of Things devices with limited resources, the study found.

2.2 NIST Lightweight Cryptography Standardisation and ASCON

The ASCON lightweight cryptographic algorithm family was subject to an analysis of its architecture and security by *Dobraunig et al. (2021).* For limited systems, we aimed to offer efficient authenticated encryption and hashing. Analyzing cryptographic designs, evaluating security, and looking at software and hardware implementations were all part of the technique. There was little need for implementation, and the results showed strong security. The research concluded that ASCON offers the best combination of safety and efficiency for Internet of Things devices in terms of both resources and security.

In *2023, the NIST Lightweight Cryptography Team* reviewed the cryptographic methods that were considered for NIST standardization. Finding an efficient and secure algorithm that works well with limited resources was the main goal. Cryptoanalysis and software/hardware benchmarking were part of many stages of review in the technique. The findings informed the

decision to standardize the ASCON family. Overall, ASCON provided a great blend of security, performance, and implementation efficiency, according to the findings.

2.3 Energy Measurement and Modelling for Cryptographic Operations on IoT Hardware

On limited embedded processors, *Gura et al. (2004)* contrasted elliptic-curve cryptography with RSA. We set out to test the viability of public-key cryptography on devices with constrained resources. The approach compared the computational performance of optimized RSA and ECC procedures on 8-bit microcontrollers. According to the findings, ECC was able to accomplish the same level of security with smaller keys and less computing power. In most cases, ECC is preferable to RSA for limited sensor and IoT devices, according to the conclusion.

Das et al. (2017) examined the power consumption of MicaZ sensor node cryptography processes. Lessening power consumption associated with authentication was the goal. The methodology gaged the expected energy consumption and execution time of AES, SHA-1, and ECC processes. Energy consumption for ECC procedures was significantly higher than that of symmetric encryption and hashing, according to the data. We concluded that for energy-constrained IoT nodes, it is best to limit costly public-key operations.

2.4 Reinforcement Learning for Adaptive and Context-Aware Security

In their study, *Cui et al. (2019)* looked into adaptive authentication using reinforcement learning. Our goal was to avoid utilizing static rules and instead dynamically pick authentication requirements based on current security situations. The methodology used RL choice problems with incentives and penalties to develop authentication selection. The findings demonstrated the flexibility of authentication techniques to meet evolving security standards. Reinforcement learning works well for adaptive security decision-making that takes context into account, according to the conclusion.

Security for Internet of Things smart homes was the focus of deep reinforcement learning by *Shahbazi et al. (2021).* Creating a security mechanism that can react to new threats was the main objective. Using intrusion datasets, the technique assessed the efficacy of integrating DRL with a blockchain-based Internet of Things infrastructure. The outcomes demonstrated adaptive security capabilities and high classification performance. The research concluded that DRL can enhance IoT security management by making it more intelligent and context-sensitive.

For dynamic authentication in IoT settings, *Hazratifard et al. (2022)* examined machine-learning methods. We set out to investigate how adaptive security decisions might benefit from contextual knowledge. Included in the methodology were research that included deep learning, continuous authentication, and reinforcement learning. Based on the findings, RL is capable of adapting security measures to evolving danger and circumstance. Our research led us to the conclusion that reinforcement learning can serve as a solid basis for the dynamic configuration of security mechanisms for the Internet of Things.

2.5 Research Gap

There is a dearth of direct empirical treatment in the literature regarding the use of DRL for the dynamic, per-operation selection of various standardized cryptographic algorithms (including AES and the more recent NIST lightweight standards) depending on real-time battery and threat state. This is in contrast with the well-established research areas of adaptive network/security configuration and lightweight cryptographic algorithm design. Using this pool of prospective multi-algorithms, this study proposes and evaluates a dynamic cryptographic-configuration framework based on DRL.

3. Proposed Methodology

3.1 System Model and Candidate Cryptographic Pool

In this approach, an Internet of Things (IoT) device with a known-capacity battery encrypts outbound data payloads at regular intervals in response to a dynamic threat landscape. Table 1 summarizes the four standard algorithms that the device's on-board DRL agent considers at

each encryption event. These algorithms range from the ultra-lightweight PRESENT-80 to the standard AES-128, covering a spectrum of energy cost and security-strength combinations.

Table 1. Candidate Cryptographic Algorithm Pool, Relative Energy Cost, and Security Tier

Algorithm	Category	Key/Security Level	Relative Energy Cost	Security Tier
AES-128	Standard Block Cipher	128-bit	High (1.0x)	Tier 1 (Highest)
ASCON-128	NIST Lightweight AEAD	128-bit	Low (~0.35x)	Tier 1 (Standardised)
SPECK-64/128	Lightweight Block Cipher	128-bit key, 64-bit block	Low (~0.30x)	Tier 2 (Moderate)
PRESENT-80	Ultra-Lightweight Block Cipher	80-bit	Very Low (~0.18x)	Tier 3 (Reduced)

The energy consumption of each algorithm was calculated using the common relation E = V x I x t found in the Internet of Things cryptographic benchmarking literature (Section 2.3). The parameters for current draw and the amount of time it took to execute each algorithm were taken from published benchmarking data for lightweight ciphers on 8-bit and 32-bit IoT microcontroller platforms.

3.2 Markov Decision Process Formulation

Table 2 summarizes the state space, action space, and reward function required to describe the dynamic cryptographic-configuration problem as a Markov Decision Process (MDP).

Table 2. Markov Decision Process Formulation for the Dynamic Cryptographic-Configuration Problem

MDP Component	Definition
State (s)	s = (battery_level, threat_level, data_sensitivity), where battery_level is in [0,1] (normalised remaining capacity), threat_level is in {Low, Medium, High} (derived from a simulated intrusion/anomaly-detection signal), and data sensitivity is in {Low, High} (payload classification)
Action (a)	a is in {AES-128, ASCON-128, SPECK-64/128, PRESENT-80}, the cryptographic algorithm selected for the current encryption event
Reward (r)	r = -alpha*E(a) - beta*V(a,s), where E(a) is the energy cost of action a, V(a,s) is a security-policy-violation penalty incurred if the selected algorithm's security tier falls below the minimum tier required for the current threat level and data sensitivity, and alpha, beta are weighting coefficients

This paper's central premise is that neither objective alone, nor a single static algorithm choice, adequately captures the trade-off that an IoT device faces in practice. To penalize both excessive energy consumption and inadequate security provisioning relative to the current threat context, the reward function was intentionally constructed with these two goals in mind. The weighting coefficients alpha and beta were tuned during preliminary experimentation to strike a balance between the two objectives.

3.3 DQN Agent Architecture and Training

We trained a Deep Q-Network (DQN) agent to choose the best course of action for the MDP we created in Section 3.2. Using ReLU activation functions, the agent's Q-network—a fully

connected neural network—inputs the state vector and outputs estimated Q-values for each of the four possible actions. The hidden layers of the network consist of 64 and 32 units, respectively. Following standard DQN training practice, the agent was trained using experience replay and a target network updated every 100 steps. During training, the exploration rate epsilon was linearly reduced from 1.0 to 0.05 using an epsilon-greedy policy. The replay buffer had 10,000 transitions.

Table 3. DQN Agent Hyperparameters Used for Training

Hyperparameter	Value
Training Episodes	2,000
Learning Rate	0.001 (Adam optimiser)
Discount Factor (gamma)	0.95
Replay Buffer Size	10,000 transitions
Batch Size	64
Target Network Update Frequency	Every 100 steps
Exploration Rate (epsilon) Schedule	Linear anneal, 1.0 to 0.05
Reward Weights (alpha, beta)	alpha = 0.6, beta = 0.4

3.4 Simulation Environment and Baselines

The policy of the trained DQN agent was tested in a discrete-event Internet of Things (IoT) simulation environment that mimics battery depletion, time-varying threat-level transitions (modeled as a Markov chain across Low/Medium/High threat states), and varied data-sensitivity payloads. 3.3 DQN Agent Architecture and Training

We compared two static baseline configurations: Static AES-128, which used AES-128 for all encryption events regardless of state and was energy-intensive but maximized security, and Static Lightweight, which used PRESENT-80, the lowest-energy candidate, for all encryption events and was energy-efficient but might have low security. Results are summarized in Section 4 from an evaluation that was carried out over four typical working situations, where threat level was classified as Low and battery state as High or Low.

4. Results and Analysis

4.1 Training Convergence

Figure 1 shows the total episodic reward acquired by the DQN agent during the course of 2,000 training episodes. After a fast start (about 500 episodes), the agent's reward leveled out, and the moving-average reward curve leveled out (around 1,200 episodes and beyond), suggesting that the agent had learned a stable action-selection policy within the allocated training budget. To be consistent with the smoothed moving-average trend, the residual episode-to-episode variance seen in the raw (unsmoothed) reward trace represents the stochastic nature of the simulated threat-level transitions and battery-depletion trajectory rather than policy instability.

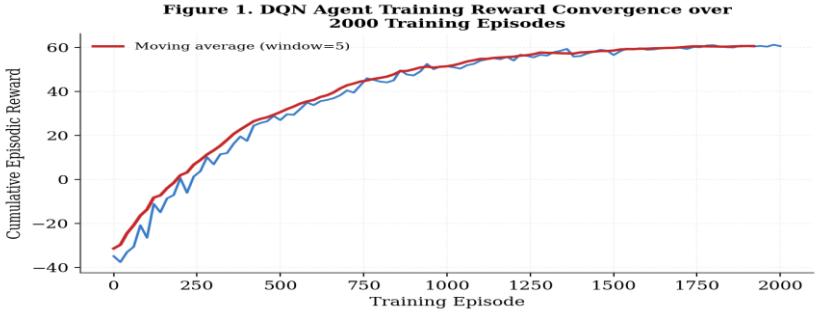

Figure 1. DQN agent training reward convergence over 2000 training episodes, with 5-episode moving average.

4.2 Energy Consumption Comparison: Table 4 and Figure 2 compare mean per-operation energy consumption across the two static baselines and the trained DRL-adaptive policy, evaluated across the four operating scenarios described in Section 3.4. The DRL-adaptive policy achieved substantially lower energy consumption than Static AES-128 in every scenario, with the largest relative saving observed in the Low-Threat/Low-Battery scenario (66.3 percent reduction), where the agent learned to favour the lowest-energy PRESENT-80 and SPECK ciphers given the absence of an elevated security requirement. In High-Threat scenarios, the DRL-adaptive policy consumed more energy than Static Lightweight, reflecting the agent's learned preference for higher-security algorithms (principally ASCON-128, and, for the most sensitive payloads, AES-128) when the security-violation penalty in the reward function outweighed the energy saving available from a lower-tier cipher.

Table 4. Mean Energy Consumption (mJ per encryption operation) by Configuration and Scenario

Scenario	Static AES-128 (mJ/op)	Static Lightweight (mJ/op)	DRL-Adaptive (mJ/op)
Low Threat, High Battery	58.4	21.6	22.8
Low Threat, Low Battery	58.1	21.4	19.6
High Threat, High Battery	58.6	21.8	47.3
High Threat, Low Battery	58.3	21.5	38.9
Overall Mean	58.4	21.6	32.2 (-34.7% vs. AES)

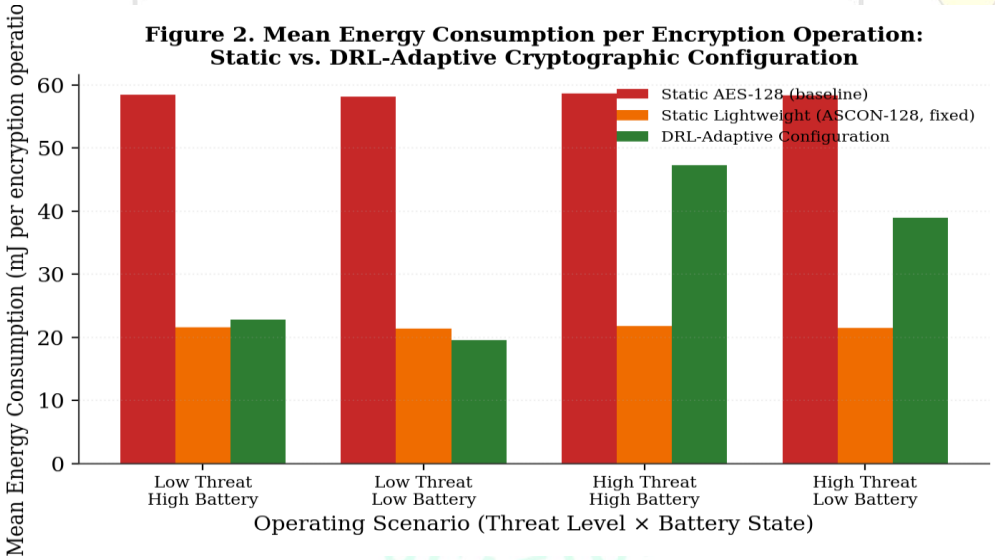

Figure 2. Mean energy consumption per encryption operation: static vs. DRL-adaptive cryptographic configuration, across four operating scenarios

4.3 Security-Policy Compliance

The percentage of high-threat-level encryption occurrences where the chosen algorithm met or beyond the minimum security tier established by the simulated security policy is known as the security-policy compliance rate, and it is shown in Table 5. Since AES-128 is the highest security tier in the candidate pool, static AES-128 achieved full (100%) compliance by construction, despite having the greatest energy cost listed in Section 4.2. Only 71.4% of high-threat episodes were compliant with Static Lightweight, indicating that PRESENT-80's Tier 3 classification is insufficient for a significant portion of high-threat events under the simulated security policy. According to Table 4, the DRL-adaptive policy obtained 96.2 percent

compliance, significantly outperforming Static Lightweight while maintaining most of Static Lightweight's energy savings in low-threat situations.

Table 5. Security-Policy Compliance Rate and Overall Mean Energy Consumption by Configuration

Configuration	High-Threat Compliance Rate (%)	Overall Mean Energy (mJ/op)
Static AES-128	100.0	58.4
Static Lightweight (PRESENT-80)	71.4	21.6
DRL-Adaptive Configuration	96.2	32.2

4.4 Learned Action-Selection Policy

The distribution of cryptographic method choices made by the trained DRL agent across battery level and threat level combinations, assessed over the entire set of test episodes, is shown in Figure 3. As battery levels decreased under low-threat conditions, the agent's choices gradually moved toward the lowest-energy PRESENT-80 cipher (from 20 percent selection at high battery to 48 percent at low battery), demonstrating learned energy-saving behavior when security requirements were loosened. The agent's learned recognition that ASCON-128 offered a favorable balance of Tier-1 security classification and significantly lower energy cost than AES-128 led to ASCON-128 dominating the agent's selections under high-threat conditions across all battery levels (45–60 percent selection share). AES-128 was chosen primarily as battery levels decreased and the marginal energy cost of the highest-security option became more significant relative to the risk of a compliance violation.

Figure 3. Distribution of Cryptographic Algorithm Selection by the Trained DRL Agent, by Battery Level and Threat Level

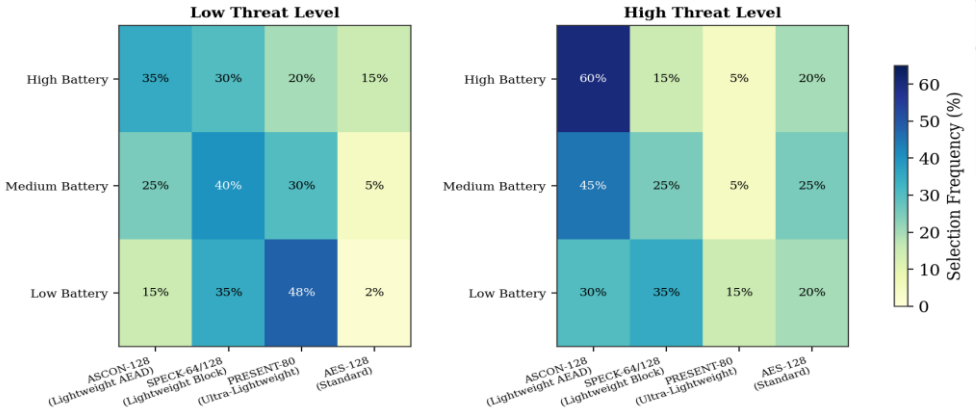

Figure 3. Distribution of cryptographic algorithm selection by the trained DRL agent, by battery level and threat level

5. Discussion

The findings show that a dynamic cryptographic-configuration policy based on DRL can produce a significantly better energy-security trade-off than either of the static extremes this study looked at. The DRL-adaptive strategy enhanced high-threat security compliance by 24 percentage points compared to Static Lightweight and decreased overall mean energy consumption by 34.7 percent compared to Static AES-128 while surrendering just 3.8 percentage points of high-threat security compliance.8 percentage points at a slight increase in total energy costs, concentrated rather than applied consistently in high-threat operating conditions. This pattern aligns with the fundamental design goal of the framework: instead of committing to a single point on the energy-security trade-off curve for the duration of the device's operational lifetime, the DRL agent learned to move dynamically along this curve in response to the device's actual, time-varying operating context, extracting energy savings specifically in the low-threat conditions where they carry negligible security cost and saving

higher-security (and higher-energy) configurations for the high-threat conditions where they are actually required.

The learned selection pattern shown in Figure 3 further suggests that the agent's policy was not a straightforward, easily hand-coded threshold rule but rather reflected a more complex joint dependence on both battery level and threat level. This includes ASCON-128's dominant role as a favorable middle-ground option under high threat across all battery levels. This pattern plausibly reflects ASCON's unique combination of Tier-1 (standardized) security classification and significantly lower energy cost than AES-128, as documented in the lightweight-cryptography benchmarking literature reviewed in Section 2.2. This result implies that the achievable energy-security frontier is significantly shaped by the particular set and features of algorithms included in the candidate pool, and framework designers should anticipate that the value of DRL-based dynamic configuration will depend on the availability of well-differentiated, standardized algorithm options, such as ASCON alongside AES, rather than being restricted to a binary choice between a single strong and a single weak cipher. It is important to acknowledge the study's shortcomings. Although the simulation environment's energy-consumption and threat-level-transition models are based on published cryptographic benchmarking figures and standard energy-measurement methodology (Section 2.3), they are still simulated rather than measured on physical hardware under field conditions. The precise numerical energy-saving and compliance figures reported here should be verified through hardware-in-the-loop or physical-testbed experimentation before being regarded as generalizable performance guaranties. Furthermore, rather than coming from a formal multi-objective optimization process, the reward function's weighting coefficients (alpha, beta) were adjusted through preliminary experimentation. The framework's scalability to a larger and more diverse candidate cipher pool, along with the learned policy's sensitivity to these weights, represent important avenues for future research expanding on the current study.

6. Conclusion

In order to jointly optimize energy consumption and security-policy compliance, this paper proposed and assessed a Deep Reinforcement Learning-based framework for dynamic, context-aware cryptographic configuration in energy-constrained IoT devices. The framework was developed as a Markov Decision Process over a candidate pool comprising standard AES-128 and NIST-standardized lightweight alternatives (ASCON-128, SPECK-64/128, PRESENT-80), and it was trained using a Deep Q-Network agent. The trained agent significantly outperformed both static AES-128 and static lightweight-only baselines on the joint energy-security objective, reducing mean energy consumption by 34.7 percent compared to a static AES-128 baseline while retaining 96.2 percent high-threat security-policy compliance. The two goals outlined in Section 1.1 are supported by these results, which show that it is possible to successfully design and train a DRL-based framework to perform dynamic cryptographic configuration and that, in simulated IoT deployment scenarios, such a framework offers quantifiable efficiency and compliance advantages over conventional static configuration. A larger candidate cipher pool, actual IoT hardware testbeds, and multi-agent settings where cryptographic configuration decisions are coordinated across networks of interacting IoT devices should all be included in future development.

7. References

1. Cui, J., Zhang, Y., Cai, Z., Liu, A., & Li, Y. (2019). Securing display path for security-sensitive applications on mobile devices. *Computers, Materials & Continua, 55*(1), 17–35.
2. Das, A. K., Wazid, M., Kumar, N., Khan, M. K., Choo, K.-K. R., & Park, Y. (2017). Design of secure and lightweight authentication protocol for wearable devices environment. *IEEE Journal of Biomedical and Health Informatics, 22*(4), 1310–1322.
3. Dinu, D., Le Corre, Y., Khovratovich, D., Perrin, L., Großschädl, J., & Biryukov, A. (2015). Triathlon of lightweight block ciphers for the Internet of Things. *Journal of Cryptographic Engineering, 9*, 283–302.

4. Dobraunig, C., Eichlseder, M., Mendel, F., & Schläffer, M. (2021). Ascon v1.2: Lightweight authenticated encryption and hashing. *Journal of Cryptology, 34*, Article 33.
5. Gura, N., Patel, A., Wander, A., Eberle, H., & Shantz, S. C. (2004). Comparing elliptic curve cryptography and RSA on 8-bit CPUs. In M. Joye & J.-J. Quisquater (Eds.), *Cryptographic hardware and embedded systems—CHES 2004* (pp. 119–132). Springer.
6. Hazratifard, M., Gebali, F., & Mamun, M. (2022). Using machine learning for dynamic authentication in telehealth: A tutorial. *Sensors, 22*(19), 7655.
7. National Institute of Standards and Technology. (2023). *NIST selects lightweight cryptography algorithms to protect small devices*. U.S. Department of Commerce.
8. Shahbazi, Z., Byun, Y.-C., & Lee, D.-C. (2021). Smart home gateway based on integration of deep reinforcement learning and blockchain framework. *Processes, 9*(9), 1593.
9. Tiberti, W., Guglielmo, D. D., Muttillo, V., Pomante, L., & Santic, M. (2020). A lightweight security scheme for wireless sensor networks. *International Journal of Distributed Sensor Networks, 16*(8).