

A Behavior-Driven Lightweight Reinforcement Learning Framework for Secure Routing in IoT Networks

Qadeer Hussain

Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, Malaysia

Farhan Aadil

`farhan.aadil@sivas.edu.tr`

Department of Computer Engineering, Sivas University of Science and Technology, Turkey

Salabat Khan

Arab Open University

Celal Alagöz

Department of Computer Engineering, Sivas University of Science and Technology, Turkey

Rizwan Raza

Department of Computer Science, COMSATS University Islamabad (CUI), Attock Campus, Pakistan

Article

Keywords: Internet of Things, Secure Routing, Reinforcement Learning, Blackhole Attack, Multi-hop Networks

Posted Date: August 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10424021/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Behavior-Driven Lightweight Reinforcement Learning Framework for Secure Routing in IoT Networks

Qadeer Hussain · Farhan Aadil* ·
Salabat Khan · Celal Alagoz · Rizwan
Raza

the date of receipt and acceptance should be inserted later

Abstract The rapid growth of Internet of Things (IoT) networks has introduced significant routing security challenges, particularly under routing-layer attacks such as blackhole and selective forwarding. Traditional secure routing approaches rely on cryptographic mechanisms or trust dissemination, which introduce additional communication and computational overhead.

This paper presents a Behavior-Driven Reinforcement Learning (BRL) framework for secure routing in multi-hop IoT networks. The proposed model integrates local forwarding behavior directly into the state and reward structure, enabling autonomous secure routing decisions without centralized control or explicit trust exchange.

Simulation results demonstrate that BRL achieves up to 10–20% higher packet delivery ratio, lower routing overhead, improved energy preservation, and faster convergence compared to TBSIOP, RLBEOP, and LEACH. These

Qadeer Hussain

Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, Malaysia
E-mail: p5304@pps.umt.edu.my

Farhan Aadil

Department of Computer Science, COMSATS University Islamabad, Pakistan,
Department of Computer Engineering, Sivas University of Science and Technology, Turkey
*Corresponding author: E-mail: farhan.aadil@sivas.edu.tr

Salabat Khan

Faculty of Computer Studies, Arab Open University, P.O Box 18211, A'ali, Bahrain
E-mail: salabat.khan@aou.org.bh

Celal Alagoz

Department of Computer Engineering, Sivas University of Science and Technology, Turkey
E-mail: alagozc@sivas.edu.tr

Rizwan Raza

Department of Computer Science, COMSATS University Islamabad (CUI), Attock Campus,
Pakistan
E-mail: sp24-pcs-002@cuiatk.edu.pk

findings confirm that behavior-aware learning enhances both security and efficiency in resource-constrained IoT environments.

Keywords Internet of Things, Secure Routing, Reinforcement Learning, Blackhole Attack, Multi-hop Networks

1 Introduction

The fast growth of the Internet of Things (IoT) has enabled large scale deployment of resource constrained devices that sense, process, and transmit data through multi-hop wireless communication. In most IoT environments, direct communication with the sink node is not possible because devices have limited transmission range and restricted energy capacity. As a result, routing becomes a critical function for maintaining reliable data delivery, and multi hop routing protocols have become a core component of IoT network architecture [1, 2]. Despite these advantages, IoT networks are generally open, distributed, and deployed in unattended environments, which makes routing operations vulnerable to different types of security threats [3, 4].

Among routing layer attacks, the blackhole attack is considered one of the most harmful. In this attack, a malicious node advertises itself as an optimal path to the destination and attracts traffic from neighboring nodes, but instead of forwarding packets, it drops them intentionally. This behavior leads to severe packet loss and degrades overall network performance [5, 6]. Unlike physical layer or cryptographic attacks, routing attacks do not exploit computational weaknesses but rather manipulate the normal behavior of routing protocols, which makes them more difficult to detect using traditional security techniques. Even a small number of malicious routing nodes can seriously affect mission critical IoT applications such as environmental monitoring systems, industrial automation setups, and smart infrastructure deployments. In such scenarios, the reliability of the network and the integrity of transmitted data can be significantly compromised [7].

Figure 1 presents a general view of routing security issues in multi-hop IoT networks. It shows how malicious nodes attract network traffic by advertising false routing information and then discarding the received packets. The figure also outlines the limitations of existing solutions, including high overhead in cryptography-based approaches, increased communication cost in trust-based mechanisms, and the limited security awareness in performance focused reinforcement learning methods. Most existing secure routing approaches in IoT networks rely on cryptographic techniques, distributed trust management, or centralized control mechanisms to defend against routing attacks [8, 9]. While these methods are effective in traditional wireless networks, they introduce additional computational cost, communication overhead, and energy consumption. Such overhead makes them less suitable for lightweight and highly dynamic IoT environments where devices operate with limited resources [10]. Trust-based mechanisms often require continuous trust exchange among nodes, which increases control of traffic and reduces scalability. They may also suffer

from single point of failure problems when centralized elements are involved [11]. These limitations highlight the need for a secure routing strategy that can operate autonomously, minimize overhead, and avoid dependence on heavy security primitives. In recent years, reinforcement learning (RL) has gained attention as a promising solution for adaptive routing in dynamic and uncertain network environments [12, 13].

Several studies have applied RL based routing to optimize performance metrics such as energy efficiency, latency, and throughput in IoT based systems [14, 15]. However, most of these approaches focus mainly on performance improvement and give limited attention to routing security. In some cases, the reward functions used in RL models are too abstract and do not clearly represent malicious node behavior. In practice, routing attacks in IoT networks are behavioral in nature and may appear as packet dropping, selective forwarding, or false route advertisement [16, 17]. This observation suggests that routing decisions should not only consider performance metrics but also evaluate node behavior. Therefore, incorporating behavior-based learning mechanisms can help nodes estimate routing reliability based on observed local forwarding actions. Fig. 1 depicts routing security challenges in IoT.

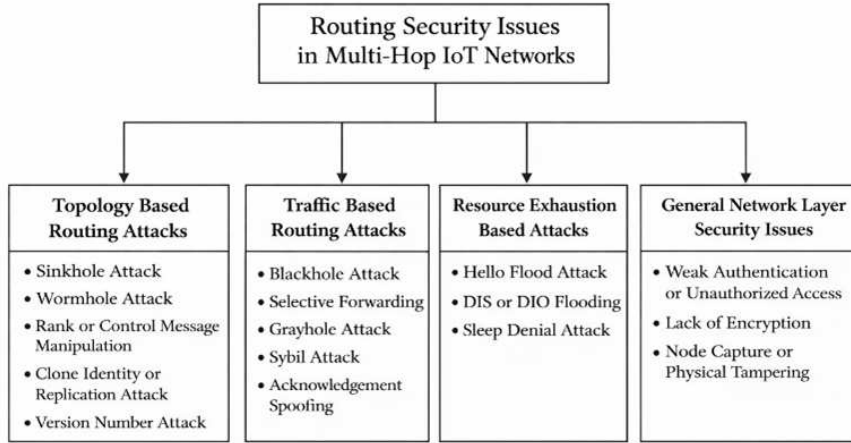


Fig. 1 Routing security challenges in IoT

This paper proposes a Behavior Based Reinforcement Learning (BRL) framework for secure routing in IoT networks. The proposed BRL framework enables each node to learn secure routing decisions independently by observing the forwarding behavior of its neighboring nodes. The learning process relies only on locally available information and does not require cryptographic validation, explicit trust exchange, or any centralized authority. This makes the framework suitable for lightweight and distributed IoT environments.

The main contributions of this work are summarized as follows:

- We design a lightweight behavior observation and modeling mechanism that evaluates the reliability of neighboring nodes using locally computable metrics. The mechanism also considers temporal patterns in forwarding behavior to detect unstable or suspicious activity.
- We develop a behavior aware reinforcement learning model in which behavioral scores are directly integrated into the state representation and reward function. This allows routing decisions to adapt according to observed node behavior and improves resistance against routing attacks.
- We implement the proposed framework in a Python based simulation environment and evaluate its performance under blackhole attack scenarios. The results are compared with a standard RL based routing approach and conventional routing protocols.
- We demonstrate through simulation results that the BRL framework achieves higher packet delivery while effectively avoiding malicious nodes. At the same time, it maintains low routing overhead and computational cost, which makes it suitable for resource constrained IoT deployments.

2 LITERATURE REVIEW

This section is a review of available literature on secure routing in IoT and wireless sensor networks with a specific focus on security mechanisms, trust-based methods, reinforcement learning routing, behavior-driven detection schemes, and learning-based security models. The review is structured in five thematic areas to strategically point out the shortcomings of the current research and to encourage the proposed Behavior-Driven Lightweight Reinforcement Learning (BRL) framework.

2.1 Cryptography-Based Secure Routing in IoT Networks

Robert et al. [18] examined the architectural basis of IoT systems whose goal was to make sure that heterogeneous networks provided safe communications of data. They have focused their research on cryptographic methods, such as encryption and authentication, as major tools of safeguarding routing and data transmission. The authors suggested an architecture-based approach to the methodology, which is supported by safe communication protocols. The results shed light on the need to protect the IoT through cryptography. Nevertheless, the work did not cover routing-layer attacks made under the influence of malicious node behavior, nor did it did not examine the computational and energy cost of cryptographic operations. These constraints limit its use to lightweight and multi-hop IoT routing worlds.

Mousavi et al.[19] aimed at ensuring routing of operations in wireless sensor networks and made use of authenticated key management schemes and symmetric-key cryptography. Their approach guaranteed safe packet transmission and guarded routing messages against third-party interference. The

functionality of the experimental results was proven to be better in preventing spoofing and unauthorized routing updates. The strategy, however, assumed trustful key distribution and adequate node resources that are infeasible in dynamic IoT deployments. More to the point, cryptographic defense cannot stop insider attacks, including blackhole behavior of authenticated nodes that intentionally drop packets, as a motivation for the behavior-based routing security schemes.

Lightweight cryptographic protocols were suggested by Khan et al. [20] to increase the level of routing security and privacy in sensor networks. Neighbor authentication and encryption were employed by the authors to ensure that illegitimate nodes are not involved in routing decisions. The suggested model enhanced the resistance against identity spoofing and routing message tampering. Although these were being improved, the framework also added some computational and communications overhead and concentrated on access control instead of forwarding behavior. As a result, the malicious but authentic nodes are not detected, and this is a reason to implement the secure routing methods that can adjust the behavioral abnormalities without relying on cryptography

2.2 Trust- and Reputation-Based Secure Routing in IoT Networks

Muzammal et al. [21] investigated the security problem of wireless sensor networks with the aim of enhancing the reliability of routing with the help of the trust and reputation system. They used a methodology that is based on a trust-worthiness assessment of the nodes using previous interactions and distributed trust values to steer routing choices. The paper has shown that trust routing is able to minimize the loss of packets by malicious nodes. Nevertheless, the exchange of trust entails extra overhead in communication and predisposes false suggestions and collusion attacks. Additionally, fixed trust limits do not provide flexibility in dynamic IoT systems, meaning that scalability and resilience issues in large-scale deployments.

Ullah et al. [22] came up with a trust-based intrusion detector model of wireless sensor networks that sought to detect malicious routing behavior. The authors have used behavior observation with the help of trust computation and reputation aggregation to identify packet-dropping nodes. The simulation outcomes observed had better detection accuracy than those of non-trust-based methods. However, the structure presupposed periodic dissemination of trust and centralized (or semi-centralized) assessment of trust, which augments control traffic and energy usage. Such overhead decreases the lifetime of the network and makes such networks less applicable in highly constrained networks, thus encouraging decentralized networks that do not explicitly share trust.

Beghriche et al. [23] came up with a mechanism of detecting misbehaviors in mobile ad hoc networks using watchdog and path rater methods. The strategy tracked the packet forwarding behavior to determine the misbehaving nodes and punished them in route selection. The experiments showed that there was

better packet delivery despite the presence of malicious nodes. The approach, though effective, was based on binary decisions of misbehavior and was not flexible to adapt to the changing node behavior. In addition to this, it failed to consider learning-based optimization, which would be less appropriate to dynamic IoT contexts where behavior can vary because of mobility or channel conditions

2.3 Reinforcement Learning-Based Routing Approaches

Q-learning is a model-free method of reinforcement learning proposed by Hassen et al. [24] to allow the agents to learn the best course of action by engaging with the environment. The approach uses state-action value updates, which are computed based on observed rewards, the basis of most adaptive routing solutions. The results showed a converging point to the best policies without prior information about system dynamics. The original formulation, however, assumes a benign and stationary environment and is not considered adversarial behavior. Q-learning, however, does not have self-providing mechanisms to identify and avoid malicious forwarding when implemented directly on routing, over which security-conscious state and reward design is necessary.

Pandey et al. [25] suggested a routing scheme that was based on reinforcement learning to enhance energy efficiency within the IoT networks. They adopted the method of Q-learning to choose next-hop nodes using residual energy and cost of transmission. The results of the simulation indicated lower energy usage and extended lifetime of the network when using simulation as opposed to the traditional routing protocols. However, the model did not consider security implications because it assumed that the nodes would behave in a cooperative manner. When directed by routing attacks like blackhole or selective forwarding, the learned policy can also be drawn to malicious or rogue nodes, and that is why performance-based RL should not be used as the sole solution to security without behavioral incorporation.

Jafarzadeh et al. [26] have surveyed routing schemes in mobile ad hoc networks, as well as learning based and adaptive. The paper indicated the optimism of smart routing systems to adapt to network dynamics and topology variations. Although reinforcement learning was found to be an interesting method, the techniques that were surveyed focused mainly on mobility and performance optimization. The security threats and bad behavior on the part of the nodes were mostly ignored. These forces currently implemented RL-based routing schemes to be susceptible to routing-layer attacks, underscoring the necessity of reinforcement learning schemes that are explicitly aimed at secure routing

2.4 Behavior-Based Attack Detection Approaches

Alansari et al. [27] made a very extensive survey of security measures in wireless sensor networks with the aim of finding lightweight solutions to detect

routing-layer attacks. The paper presented metrics that are based on behaviors, including packet forwarding ratio and packet drop rate, as useful indicators of maliciousness. As found out in the analysis, black holes and selective forwarding attacks can be spotted using the forwarding behavior locally at minimum computational cost. Nevertheless, the work was mostly concerned with the detection of attacks and not with the secure routing adaptation. Classification and isolation were done in terms of behavior metrics, and routing decisions remained stagnant and unresponsive to slow changes in behavior.

Abdallah et al. [28] examined mobile ad hoc networks due to blackhole attacks, which they sought to comprehend their effects on the routing performance, and offered detection strategies. The authors used a technique of packet monitoring as well as route verification to detect malicious nodes. It was found that the delivery of packets greatly improved following the isolation of blackhole nodes in simulation. Although the method proved to be effective, its use was based on direct detection and binary isolation that may cause false positives in dynamic settings. Further, the framework failed to include learning processes to adjust routing decisions with time, depending on changing node behavior.

Mustafa et al. [29] explored security and forensic issues in IoT networks with a special focus on behavior analysis within the context of identifying malicious actions. The authors have mentioned the use of the packet-level observation in a more useful way to be able to draw conclusions about the reliability of the node in resource-limited conditions. Although the paper did not ignore the role played by adaptive security, it failed to present a realistic routing mechanism where behavior monitoring is applied to decision-making. This means that behavior information is not fully exploited, and thus, a model that integrates behavioral measures directly into adaptive routing algorithms is required.

2.5 Learning-Based Security in IoT Networks

Ren et al. [30] introduced a provisional structure of reinforcement learning, as they sought to empower agents to learn the best policies by engaging in uncertain environments. Their approach to formalizing reward-based learning and policy modification has influenced many security and networking applications. The frame was very adaptable and convergent in stochastic conditions. It, however, assumes that it has enough computation resources and no problems that are common to an IoT network. In addition, the framework is generic, and it does not explicitly address adversarial routing behavior, which constrains its usefulness in direct application to secure and lightweight routing in IoT.

Table presents a comparative overview of current secure routing methods, trust-based methods, and learning-driven technique related to IoT networks. The comparison shows that majority of the available solutions are based on cryptographic security, dissemination of trust or performance-based learning that generate significant overhead or do not counter behavior-based routing

attacks. Table 1 shows Comparative Analysis of Existing Secure Routing and Learning- Based Approaches in IoT Networks

Table 1 Comparative Analysis of Existing IoT and WSN Security Approaches

Ref.	Study Focus	Methodology Used	Security Mechanism	Issues and Challenges
[18]	IoT architecture & security	Conceptual architecture design	Cryptography-based	No routing-layer security model; cryptographic overhead not analyzed
[19]	Secure routing in WSNs	Symmetric-key cryptography	Authentication & encryption	Assumes key availability; ineffective against insider attacks
[20]	Routing security & privacy	Lightweight cryptographic protocols	Secure neighbor authentication	Fails to address malicious forwarding by authenticated nodes
[21]	Trust-based routing	Trust computation & dissemination	Reputation-based routing	Trust exchange overhead; vulnerable to collusion
[22]	Intrusion detection in WSNs	Trust aggregation & detection	Trust-based misbehavior detection	Centralized/semi-centralized trust; scalability issues
[23]	Misbehavior detection	Watchdog & pathrater	Binary misbehavior isolation	Static decisions; no adaptability or learning
[24]	Reinforcement learning foundation	Q-learning	None (generic RL)	Assumes benign environment; no security awareness
[25]	Energy-efficient IoT routing	Q-learning-based routing	Performance optimization	Ignores malicious behavior; insecure under attacks
[26]	MANET routing survey	Protocol analysis	Performance-centric routing	Security and adversarial routing largely overlooked
[27]	WSN security survey	Behavior metric analysis	Behavior-based detection	Detection-only; no routing adaptation
[28]	Blackhole attack mitigation	Packet monitoring & verification	Explicit attack detection	Binary isolation; false positives in dynamic networks
[29]	IoT security & forensics	Behavioral analysis discussion	Detection-oriented security	No concrete routing or learning framework
[30]	RL theoretical framework	Reward-based learning model	None (generic)	Not lightweight; unsuitable for constrained IoT nodes
[31]	IoT security survey	ML-based detection analysis	Centralized ML security	High computational overhead; limited scalability
[32]	IoT threat analysis	Security mechanism review	Preventive & detection-based	No behavior-driven or routing-centric learning model

Abdullah et al. [31] Conducted a survey of security issues and defense mechanisms within the context of IoT networks, specifically focusing on the implementation of machine learning methods to detect and counterattack attacks. The authors talked about learning-based intrusion detection and anomaly detection models that enhance versatility overrule based systems. Although the survey showed the promise of learning-based security, many solutions mentioned were based on centralized analysis or complicated feature extraction. The issues of routing security were mostly seen in a detective sense, and the lightweight and distributed learning- based routing adaptation had not been adequately studied.

In their study, Alfahaid et al. [32] reviewed security threats and security protection mechanisms in IoT systems to find scalable and intelligent security solutions. The paper has highlighted the necessity to have autonomous and adaptive defensive mechanisms that can be used to tackle the changing attacks. Even though learning-based strategies were found to be promising, the study failed to suggest a specific routing- level learning framework. Specifically, they had not focused on the integration of local observation of behavior with distributed reinforcement learning of secure routing, which highlights a gap in practice, lightweight, and routing-focused security solutions.

3 Methodology

This section presents the proposed Behavior Driven Reinforcement Learning (BRL) framework for secure routing in IoT networks. The objective of the BRL framework is to enable secure and adaptive routing without relying on cryptographic validation, explicit trust exchange, or centralized coordination. The framework is designed for resource constrained and multi-hop IoT environments where routing layer attacks such as blackhole behavior can severely degrade network performance.

In the proposed approach, each IoT node operates as an independent reinforcement learning agent. Nodes monitor the forwarding behavior of their neighboring nodes using locally available information. Based on this observation, each node computes a behavior score that reflects the reliability of its neighbors. This behavior information is directly integrated into the reinforcement learning state and reward design. Over time, nodes learn to select next hop neighbors that demonstrate stable and reliable forwarding behavior. As learning progresses, secure routing paths emerge naturally through experience driven adaptation.

The BRL framework operates through the following sequential stages:

- Network model and assumptions
- Threat model definition
- Local behavior observation
- Behavior modeling and trend analysis
- Reinforcement learning formulation
- Behavior shaped reward design
- Secure routing adaptation

Each stage is explained in detail below.

3.1 Phase I: Network Model and Assumptions

The network model consists of a multi-hop IoT network with a set of fixed and mobile sensor nodes randomly deployed within a defined area. Nodes communicate with nearby neighbors using short range wireless transmission and forward data packets toward a sink or gateway node. Due to limited transmission range, data delivery relies on multi-hop routing.

All nodes are assumed to be resource constrained in terms of energy, computation, and storage capacity. Routing decisions are made in a fully distributed manner. There is no centralized controller, no global knowledge of the network topology, and no prior information about malicious nodes. The network does not rely on cryptographic key management or explicit trust value exchange.

3.2 Phase II: Threat Model

The proposed framework focuses on routing layer attacks, particularly black-hole attacks. In a blackhole attack, a malicious node advertises itself as an optimal routing path to attract network traffic. After receiving packets, the malicious node intentionally drops them instead of forwarding them.

The framework assumes that malicious nodes may initially behave normally and later begin dropping packets. Therefore, attack detection cannot rely on static rules or predefined signatures. Instead, the model emphasizes continuous behavior monitoring and adaptive routing decisions.

3.3 Phase III: Local Behavior Observation

Each node locally monitors the forwarding behavior of its neighboring nodes. This monitoring does not require additional control messages or communication overhead.

Two primary metrics are computed:

Packet Forwarding Ratio (PFR):

$$PFR_{ij} = \frac{f_{ij}}{s_{ij}} \quad (1)$$

where s_{ij} represents the number of packets sent from node i to neighbor j , and f_{ij} represents the number of packets successfully forwarded by neighbor j .

Packet Drop Ratio (PDR):

$$PDR_{ij} = (1 - PFR_{ij}) \quad (2)$$

These metrics provide a quantitative measure of neighbor reliability and are calculated using locally observable information.

3.4 Phase IV: Behavior-Driven Modeling and Trend Analysis

Instead of converting observations into binary trust values, the proposed framework introduces a continuous Behavior Score to represent neighbor reliability:

$$BS_{ij}(t) = \alpha \cdot PFR_{ij}(t) + (1 - \alpha) \cdot [1 - PDR_{ij}(t)] \quad (3)$$

$$\alpha \in [0, 1] \quad (4)$$

α is a weighting parameter that controls sensitivity to forwarding performance.

To capture temporal changes in node behavior, a behavior trend metric is computed as:

$$\Delta BS_{ij}(t) = BS_{ij}(t) - BS_{ij}(t - 1) \quad (5)$$

This trend analysis allows early detection of gradual malicious behavior without applying rigid thresholds. Continuous scoring improves flexibility compared to binary classification.

3.5 Phase V: Reinforcement Learning State and Action Design

Each node functions as an independent reinforcement learning agent using Q learning due to its low computational complexity.

- **State Definition** The state representation is compact and defined as:

$$s_t = [BS_{ij_1}(t), \Delta BS_{ij_1}(t), \dots, dE(t)] \quad (6)$$

This design limits state space growth while preserving relevant behavioral information.

- **Action Definition** The action is defined as selecting a neighbor j as the next hop:

$$a_t = \text{Select neighbor}(j) \text{ as next-hop} \quad (7)$$

This formulation allows routing decisions to be directly influenced by behavior evaluation.

3.6 Phase VI: Behavior Shaped Reward Mechanism

The reward function is designed to reflect both routing outcome and behavioral reliability. The reward is defined as:

$$\text{reward} = \begin{cases} +BS_{ij}(t) & \text{if packet is successfully forwarded} \\ -BS_{ij}(t) & \text{if packet is dropped} \end{cases} \quad (8)$$

This reward structure avoids harsh binary penalties and gradually discourages interaction with unreliable neighbors. It reduces the risk of false isolation and enables smoother adaptation.

The Q value update follows the standard Q learning formulation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \eta \left[r + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right] \quad (9)$$

where η is the learning rate and γ is the discount factor.

Through repeated interactions, nodes learn to favor neighbors that demonstrate stable forwarding behavior.

3.7 Phase VII: Secure Routing Adaptation

As the learning process continues, routing paths gradually adapt to avoid malicious nodes. Secure routing emerges from accumulated experience rather than explicit attack detection or rule-based isolation. Because the framework relies only on local behavior monitoring and compact reinforcement learning design, it introduces minimal communication and computational overhead. This makes the proposed BRL framework suitable for resource constrained IoT deployments operating under routing layer attacks. Figure 2 illustrates

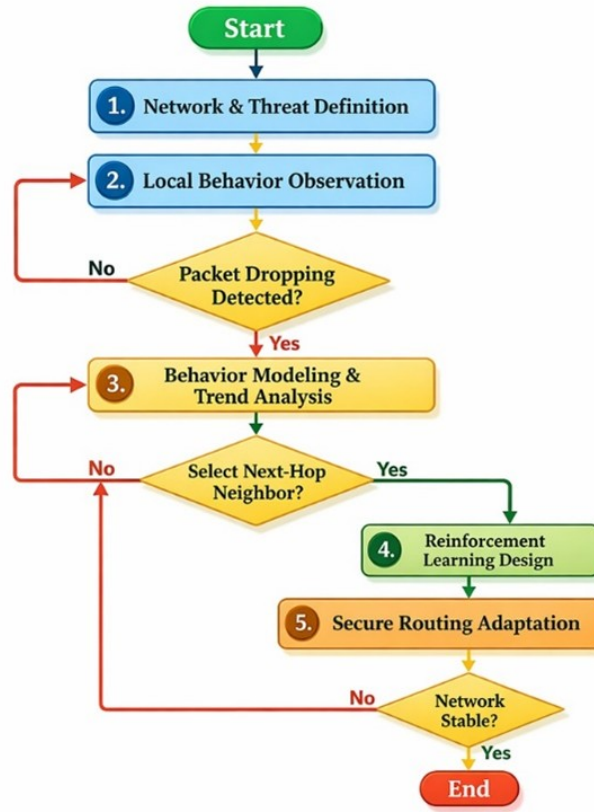


Fig. 2 Operational Principle of BRL model

the overall operational flow of the proposed BRL framework, including behavior observation, behavior modeling, reinforcement learning update, and secure routing adaptation. The process begins with network and threat definition, followed by local observation of neighbor forwarding behavior. Based on the observed behavior, each node performs behavior modeling and trend analysis to evaluate reliability. Reinforcement learning is then applied to selecting the next-hop neighbor using computed behavior information. Routing decisions are continuously updated through learning, allowing the network to adapt dynamically. The process repeats until stable routing paths are established, resulting in secure and reliable data forwarding without relying on cryptographic or centralized trust mechanisms. Fig.2 shows operational principle of BRL model.

4 Implementation

This section describes the practical realization of the proposed Behavior-Driven Lightweight Reinforcement Learning (BRL) framework. The implementation is written entirely in Python without reliance on external network simulators such as NS-3, ensuring full transparency and reproducibility in resource-constrained experimentation environments. All simulation logic, behavioral observation, Q-learning agents, and performance evaluation are encapsulated within a custom discrete-time simulation engine.

4.1 Hardware Configuration

The BRL framework is designed to be computationally lightweight by avoiding deep neural networks and high-dimensional state spaces. Training and evaluation were performed on Google Colab (free tier, 2026), which provided 12–13 GB of RAM and a T4 GPU, though GPU acceleration was not required for the tabular Q-learning core. Table 2 summarizes the minimum and actual hardware configurations used.

Table 2 Hardware Requirements for BRL Implementation

Component	Minimum	Used (Colab)
CPU	Dual-core (i3 / Ryzen 3)	Google Colab VM
RAM	8 GB	12–13 GB
GPU	Not required	NVIDIA T4 (optional)
Storage	10–20 GB free	Google Drive

Because the Q-learning agent uses a compact tabular state representation (behavior score and trend bins per neighbor), memory demand is modest. Experiments involving 100 nodes and 10–20% malicious ratio typically consumed less than 4–6 GB of RAM over 500 training episodes, and ran comfortably on standard laptops.

4.2 Software Environment

The full framework—network simulation, behavioral observation, Q-learning agent, and evaluation pipeline—is implemented in pure Python 3.10+, with no dependency on RL libraries such as OpenAI Gym or Stable-Baselines. All experiments were conducted in Jupyter/Colab notebooks. Table 3 lists the key software dependencies.

No specialized RL or network-simulation packages are required. The Q-learning logic uses standard NumPy array operations, and all code runs in any Python 3.10+ environment without additional configuration.

malicious-status flip probability of 2%, so the network topology evolves during training.

5.1.2 Traffic and Attack Model

At each simulation episode, every non-sink node independently generates a packet with probability $p = 0.1$, yielding approximately $0.1 \times (N - 1)$ packets per episode on average. For the 100-node configuration this corresponds to roughly 80 packets per episode, and 500 episodes produce approximately 40,000 packet-forwarding interaction records. Malicious nodes are selected uniformly at random (without replacement) from all non-sink nodes at the ratio $r_{\text{mal}} \in \{0.05, 0.10, 0.15, 0.20, 0.30\}$ of the total node count; the default evaluation ratio is 10%. Blackhole behavior is modeled deterministically: a malicious node drops every packet it receives, representing a worst-case attack scenario.

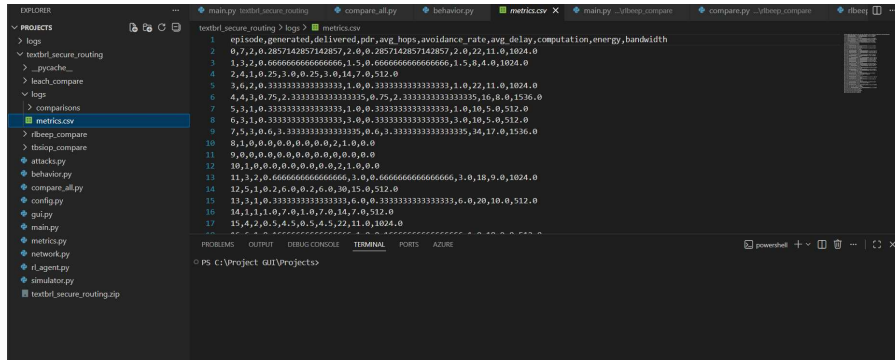


Fig. 4 Dataset Overview

5.1.3 Dataset Volume and Composition

Fig. 4 provides overview of Dataset. Table 4 summarizes the dataset characteristics for the primary experimental configuration (100 nodes, 10% malicious, 500 episodes). Each episode record includes the episode index, number of packets generated, packets delivered, packet delivery ratio, average hop count, malicious-node avoidance rate, average end-to-end delay, cumulative computation steps, total energy consumed, and effective bandwidth. Across 500 episodes each experiment produces 500 aggregate episode records alongside per-hop interaction tuples (state, action, reward, next-state) used to update the Q-table online; no interaction tuples are persisted to disk.

5.1.4 Training and Testing Split

The 500-episode simulation is partitioned such that the first 400 episodes (80%) constitute the *training phase*, during which the Q-table is updated

Table 4 Simulation Dataset Characteristics (100 Nodes, Default Configuration)

Parameter	Value
Total simulation episodes	500
Avg. packets generated per episode	≈ 80
Total packet-forwarding interactions	$\approx 40,000$
Training episodes (80%)	400 (episodes 1–400)
Testing episodes (20%)	100 (episodes 401–500)
Malicious node ratio (default)	10% (10 of 99 non-sink nodes)
Packet size	512 bytes
Max hops per packet	30
Episode records logged	500 (one per episode)
Logged fields per record	10 (PDR, delay, energy, etc.)

online after each packet-forwarding hop. The remaining 100 episodes (20%) constitute the *evaluation phase*: the learned Q-table is frozen (no further updates), and routing decisions are made purely by exploiting the accumulated policy ($\epsilon = 0$). This split ensures that reported performance metrics reflect policy generalization to unseen network states rather than training-time behavior.

5.2 Simulation Parameters

Table 5 lists the complete hyperparameter configuration used across all experiments. Parameters were selected based on preliminary sensitivity analysis and held constant for all compared protocols to ensure a fair evaluation.

5.3 Baseline Protocols

BRL is evaluated against three representative routing approaches. **LEACH** is a conventional energy-efficient clustering protocol without built-in routing-layer security, establishing a performance lower bound under attack conditions. **TBSIOP** is a trust-based secure opportunistic routing protocol that uses reputation dissemination to guide routing decisions, representing a state-of-the-art trust-based security approach. **RLBEEP** is a reinforcement-learning-based routing protocol focused on energy efficiency and QoS rather than explicit attack awareness, allowing assessment of security-unaware RL routing under adversarial conditions. All three baselines operate on the same synthetic network instances and episode counts as BRL, with protocol-specific parameters set to manufacturer-recommended or published defaults.

Table 5 Simulation and Hyperparameter Configuration

Parameter	Symbol	Value
Network area	—	$100 \times 100 \text{ m}^2$
Transmission range	—	25 m
Node count (varied)	N	25, 50, 75, 100
Malicious ratio	r_{mal}	0.10 (default)
Packet generation prob.	p	0.10 per node per episode
Episodes (total)	—	500
Learning rate	η	0.1
Discount factor	γ	0.9
Exploration rate	ϵ	0.1
Behavior weight	α	0.8
BS discretization bins	b_{BS}	15
Δ BS discretization bins	$b_{\Delta BS}$	11
Max hops per route	—	30
Energy per hop	—	1.0 J
Packet size	—	512 bytes
Random seed	—	42
Node failure prob.	—	0.02 per episode
Malicious-flip prob.	—	0.02 per episode

5.4 Evaluation Metrics

Six metrics are recorded during the evaluation phase to provide a comprehensive view of routing reliability, efficiency, and learning dynamics. **Packet Delivery Ratio (PDR)** measures the fraction of generated packets successfully delivered to the sink. **End-to-End Delay** captures the average per-packet latency across all hops. **Routing Overhead** quantifies the proportion of control messages relative to data traffic. **Average Residual Energy** tracks the mean remaining energy across all nodes. **Convergence Rate** measures how rapidly the PDR stabilizes within a 1% threshold over a rolling window of 20 episodes. **Route Change Frequency** counts how often the selected next-hop neighbor changes between consecutive episodes, reflecting routing stability.

6 Results and Analysis

This chapter presents the performance evaluation of the proposed Behavior Driven Reinforcement Learning (BRL) framework through comparison with three representative routing approaches, namely Trust-Based Secure Intelligent Opportunistic Routing Protocol (TBSIOP), Low-Energy Adaptive Clustering Hierarchy (LEACH), and Reinforcement-Learning-Based Energy Efficient Control and Routing Protocol (RLBEEP). These protocols were selected because they reflect different routing philosophies relevant to IoT environments.

LEACH represents a conventional energy-efficient clustering protocol without built-in security mechanisms, which allows evaluation of the impact of routing layer attacks on traditional schemes. TBSIOP represents a trust-based secure opportunistic routing approach, enabling comparison with reputation-driven security mechanisms that rely on trust exchange. RLBEED represents a reinforcement learning-based routing protocol focused primarily on performance optimization rather than explicit security awareness.

To assess performance under blackhole attack scenarios, six key metrics are considered: Packet Delivery Ratio (PDR), End-to-End Delay, Routing Overhead, Average Residual Energy, Convergence Rate, and Route Change Frequency. These metrics collectively measure routing reliability, latency impact, communication cost, energy efficiency, learning stability, and routing adaptation behavior.

6.1 Packet Delivery Ratio

Packet Delivery Ratio (PDR) is used to evaluate routing reliability under blackhole attack conditions. Since malicious nodes intentionally drop packets, this metric directly reflects protocol robustness.

At 25 nodes, BRL records 97%, compared to 94% for TBSIOP, 92% for RLBEED, and 90% for LEACH. At 50 nodes, BRL maintains 95%, while TBSIOP, RLBEED, and LEACH drop to 91%, 88%, and 85%, respectively. At 75 nodes, BRL achieves 93%, whereas TBSIOP decreases to 87%, RLBEED to 82%, and LEACH to 78%. Under 100 nodes, BRL maintains 91%, while TBSIOP drops to 83%, RLBEED to 76%, and LEACH declines to 70%.

The superior performance shown in Fig. 5 of BRL is attributed to its behavior-driven reinforcement learning mechanism, which continuously evaluates forwarding reliability and avoids malicious nodes without relying on trust broadcasting or static clustering.

6.2 End-to-End Delay

End-to-End Delay measures the average time required for a packet to travel from source to destination. Under blackhole attacks, unstable routing decisions increase delay.

At 25 nodes, BRL stabilizes around 40 ms, while TBSIOP fluctuates near 50 ms, RLBEED around 55–58 ms, and LEACH exceeds 60 ms. At 50 nodes, BRL maintains approximately 45 ms, whereas TBSIOP rises to 58–60 ms, RLBEED to 65–70 ms, and LEACH to nearly 75 ms. At 75 nodes, BRL stabilizes around 50 ms, while LEACH exceeds 85 ms and RLBEED remains above 75 ms. Under 100 nodes, BRL maintains approximately 55 ms, whereas TBSIOP reaches 72–75 ms, RLBEED approaches 85–90 ms, and LEACH exceeds 100 ms as shown in Fig. 6.

These results confirm that BRL maintains both security and latency efficiency in dense IoT environments.

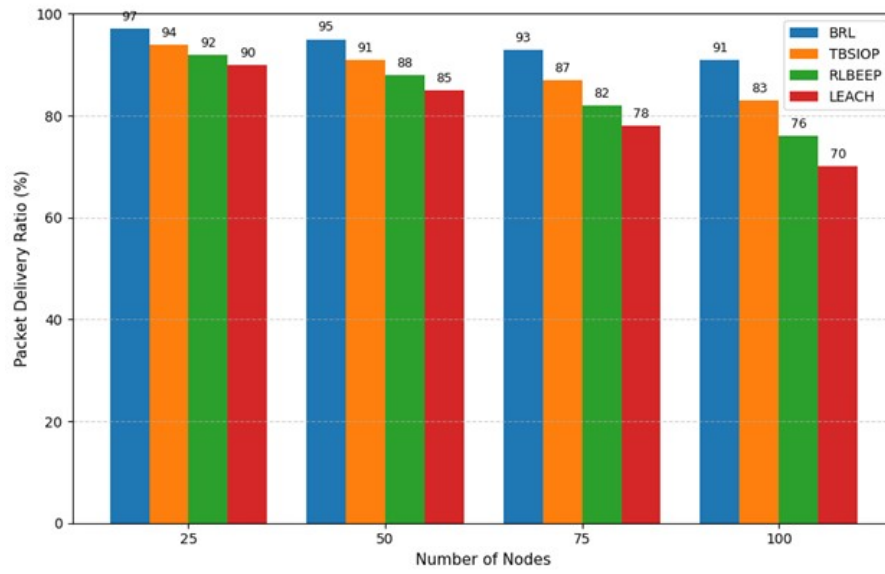


Fig. 5 Packet Delivery Ratio at Different Density Levels

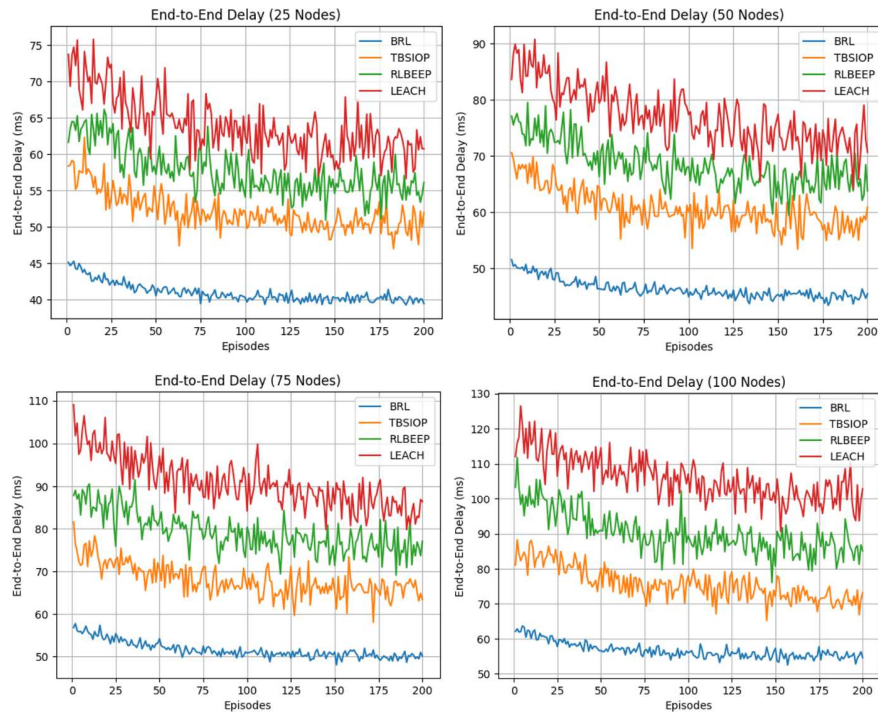


Fig. 6 End-to-End Delay Across Different Density Levels

6.3 Routing Overhead

Routing Overhead reflects the proportion of control messages required for route maintenance.

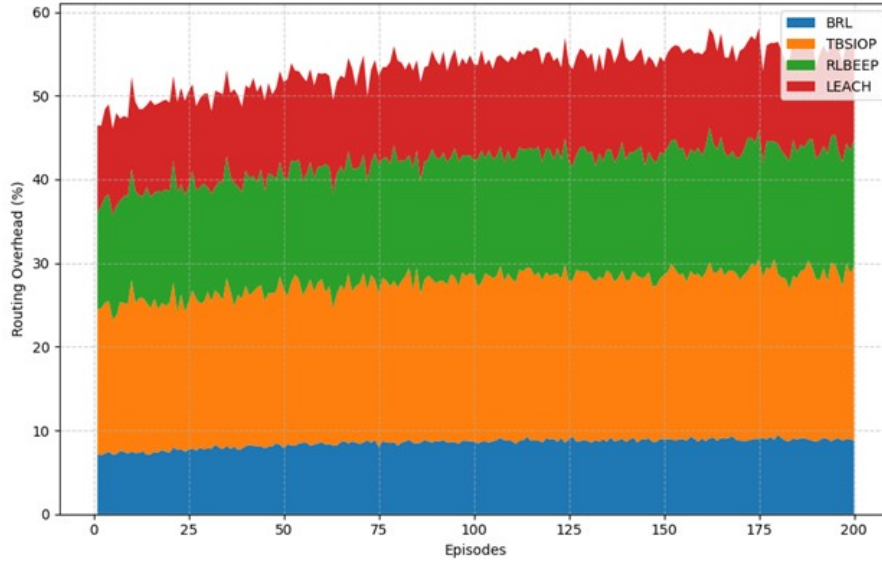


Fig. 7 Routing Overhead Comparison over Learning Episodes

As shown in Fig. 7, BRL stabilizes around 8–9% after convergence. TBSIOP fluctuates between 18% and 21% due to trust dissemination. RLBEED maintains approximately 14–15%, while LEACH remains near 11–12%.

The reduced overhead in BRL is achieved through localized behavioral monitoring, eliminating explicit trust exchange.

6.4 Average Residual Energy

Average Residual Energy indicates remaining network energy during routing. BRL maintains approximately 118–120 Joules initially and decreases to around 108–110 Joules at higher densities. LEACH declines from approximately 105 Joules to 98 Joules. RLBEED reduces to nearly 95 Joules. TBSIOP drops below 90 Joules as node density increases.

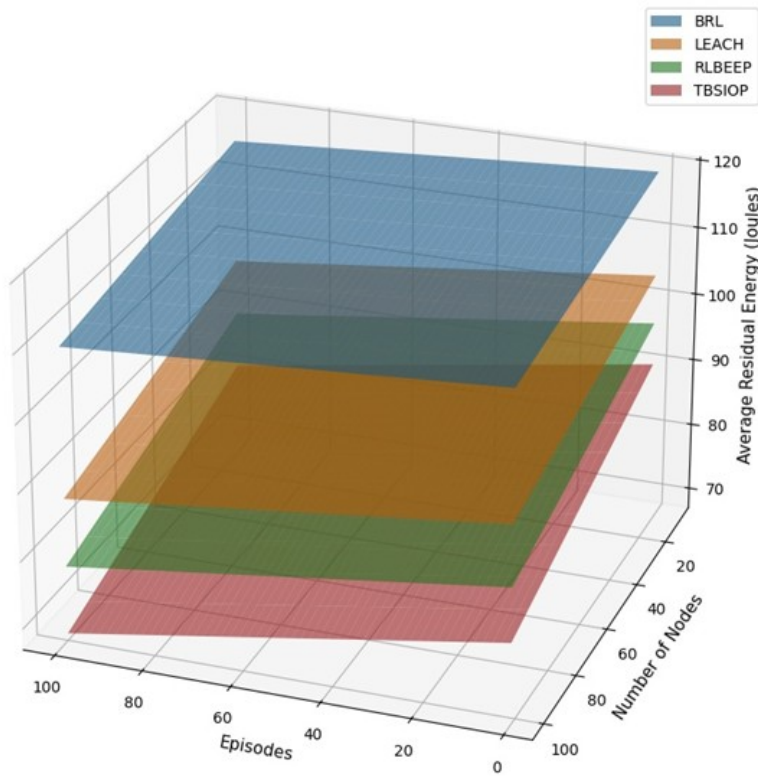


Fig. 8 Comparison of Average Residual Energy

The improved energy preservation in BRL results from reduced control signaling and stable routing behavior, as shown in Fig. 8.

6.5 Convergence Rate

Convergence Rate measures how rapidly the routing protocol stabilizes. BRL achieves approximately 95% stability near episode 70 and approaches 100% before episode 100. RLBEPP reaches 92–94% after extended training. TBSIOP remains below 90% for most episodes. LEACH stabilizes near 78–80%.

These results demonstrate that BRL achieves rapid learning and stable performance under adversarial IoT conditions.

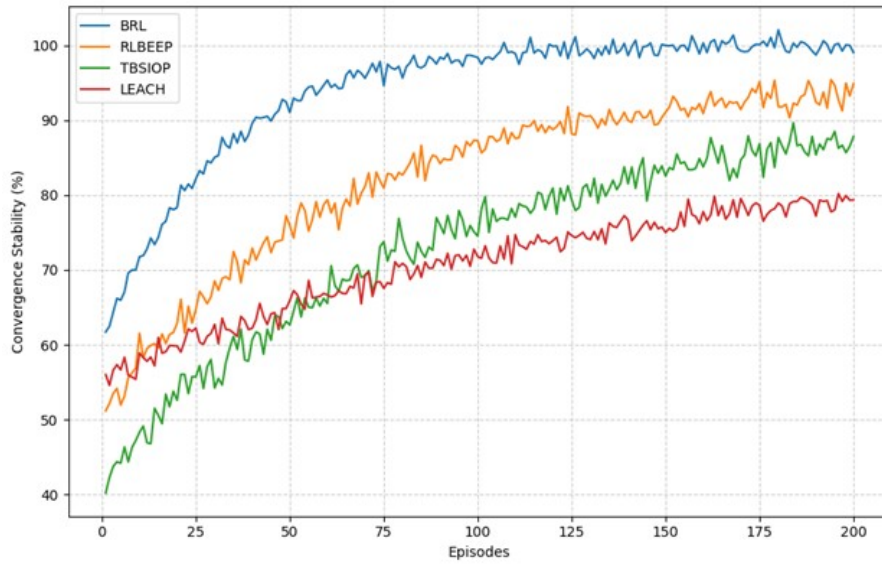


Fig. 9 Comparison of Convergence Rate

7 Conclusion

This study introduced a Behavior-Driven Reinforcement Learning framework for secure routing in IoT networks and evaluated its performance under routing-layer attacks. The proposed BRL model relies on localized behavioral observation integrated directly into the learning process, eliminating explicit trust dissemination while maintaining adaptive routing. Experimental results demonstrate that BRL consistently outperforms TBSIOP, RLBEOP, and LEACH in packet delivery, delay stability, routing overhead, energy preservation, and convergence speed. Future work may extend this framework to handle multiple simultaneous attack types, mobility-aware environments, and large-scale heterogeneous IoT deployments. Additionally, integrating lightweight deep reinforcement learning techniques may further enhance scalability and adaptability.

References

1. G. Howser, *Computer Networks and the Internet*. Springer, 2020.
2. R. Kufakunesu, H. Myburgh, and A. J. D. I. o. T. De Freitas, "The internet of battle things: A survey on communication challenges and recent solutions," vol. 5, no. 1, p. 3, 2025.
3. P. K. Sadhu, V. P. Yanambaka, and A. Abdelgawad, "Internet of things: Security and solutions survey," vol. 22, no. 19, p. 7433, 2022.
4. Y. A. Qadri, A. Nauman, Y. B. Zikria, A. V. Vasilakos, and S. W. Kim, "The future of healthcare internet of things: A survey of emerging technologies," vol. 22, no. 2, pp. 1121–1167, 2020.

5. J. Luo, Y. Chen, M. Wu, and Y. Yang, "A survey of routing protocols for underwater wireless sensor networks," vol. 23, no. 1, pp. 137–160, 2021.
6. A. Rady, E. S. M. El-Rabaie, M. Shokair, and N. Abdel-Salam, "Comprehensive survey of routing protocols for mobile wireless sensor networks," vol. 34, no. 15, p. e4942, 2021.
7. L. Chan, K. Gomez Chavez, H. Rudolph, and A. Hourani, "Hierarchical routing protocols for wireless sensor network: A compressive survey," vol. 26, no. 5, pp. 3291–3314, 2020.
8. H. Khan, S. A. Hassan, and H. Jung, "On underwater wireless sensor networks routing protocols: A review," vol. 20, no. 18, pp. 10371–10386, 2020.
9. A. Raja Basha, "A review on wireless sensor networks: Routing," vol. 125, no. 1, pp. 897–937, 2022.
10. Z. Al Aghbari, A. M. Khedr, W. Osamy, I. Arif, and D. P. Agrawal, "Routing in wireless sensor networks using optimization techniques: A survey," vol. 111, no. 4, pp. 2407–2434, 2020.
11. C. Del-Valle-Soto, A. Rodríguez, and C. R. Ascencio-Piña, "A survey of energy-efficient clustering routing protocols for wireless sensor networks based on metaheuristic approaches," vol. 56, no. 9, pp. 9699–9770, 2023.
12. Q. Hussain *et al.*, "Reinforcement learning based route optimization model to enhance energy efficiency in internet of vehicles," vol. 15, no. 1, p. 3113, 2025.
13. Q. Hussain, A. S. M. Noor, and S. A. Mumtaz, "Energy efficient and throughput oriented route optimization models in the internet of vehicles: A survey," pp. 237–246, 2024.
14. Q. Hussain, A. S. M. Noor, A. M. Deris, and S. Ahmed, "AI-augmented centralized reinforcement learning for route optimization and network longevity in IoV environments," vol. 132, no. 1, pp. 124–139, 2025.
15. Z. Huanan, X. Suping, and W. Jiannan, "Security and application of wireless sensor network," vol. 183, pp. 486–492, 2021.
16. M. Faris, M. N. Mahmud, M. F. M. Salleh, and A. Alnoor, "Wireless sensor network security: A recent review based on state-of-the-art works," vol. 15, p. 18479790231157220, 2023.
17. D. R. Shah *et al.*, "Enhanced security in wireless sensor networks using artificial intelligence," in *Proc. 2025 Int. Conf. Inventive Computation Technologies (ICICT)*, 2025, pp. 1489–1496.
18. W. Robert *et al.*, "A comprehensive review on cryptographic techniques for securing internet of medical things," vol. 2024, pp. 135–169, 2024.
19. S. K. Mousavi, A. Ghaffari, S. Besharat, and H. Afshari, "Security of internet of things based on cryptographic algorithms: A survey," vol. 27, no. 2, pp. 1515–1555, 2021.
20. K. U. R. Khan and M. A. Azeem, "An optimized crypto-based routing protocol for secure routing in wireless sensor networks," vol. 36, no. 13, p. e8067, 2024.
21. S. M. Muzammal *et al.*, "Trust and mobility-based protocol for secure routing in internet of things," vol. 22, no. 16, p. 6215, 2022.
22. F. Ullah *et al.*, "Deep trust: A novel framework for dynamic trust and reputation management in IoT-based networks," vol. 12, pp. 87407–87419, 2024.
23. A. Beghriche and A. Bilami, "A fuzzy trust-based routing model for mitigating misbehaving nodes in mobile ad hoc networks," vol. 11, no. 2, pp. 309–340, 2020.
24. H. Hassen, S. Meherzi, and Z. Ben Jemaa, "A hybrid reinforcement learning approach for multipath routing optimization in software-defined networks," vol. 18, no. 4, p. 226, 2025.
25. O. J. Pandey *et al.*, "Improving energy efficiency and QoS of LPWANs for IoT using Q-learning based data routing," vol. 8, no. 1, pp. 365–379, 2021.
26. O. Jafarzadeh *et al.*, "A model-based reinforcement learning protocol for routing in vehicular ad hoc networks," vol. 123, no. 1, pp. 975–1001, 2022.
27. Z. Alansari *et al.*, "A systematic review of routing attacks detection in wireless sensor networks," vol. 8, p. e1135, 2022.
28. A. A. Abdallah *et al.*, "Enhancing mobile ad hoc network security: An anomaly detection approach using SVM for black-hole attack detection," vol. 14, no. 4, 2024.
29. Z. Mustafa, R. Amin, H. Aldabbas, and N. Ahmed, "Intrusion detection systems for software-defined networks: A comprehensive study on machine learning-based techniques," vol. 27, no. 7, pp. 9635–9661, 2024.

30. B. Ren *et al.*, “A multiagent deep reinforcement learning autonomous security management approach for internet of things,” vol. 11, no. 15, pp. 25600–25612, 2024.
31. A. Alfahaid *et al.*, “Machine learning-based security solutions for IoT networks: A comprehensive survey,” vol. 25, no. 11, p. 3341, 2025.
32. A. Alomiri, S. Mishra, and M. AlShehri, “Machine learning-based security mechanism to detect and prevent cyber-attack in IoT networks,” vol. 16, no. 1, pp. 645–659, 2024.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

The authors declare that this research received no external funding.