Received: 19 January 2026 | Revised: 9 April 2026 | Accepted: 11 July 2026 | Published online: 18 August 2026

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–13
DOI: 10.47852/bonviewJCCE62029141

Trust-Gated Reinforcement Learning-Based Secure Routing for Wireless Sensor Networks

Rui Zheng[1], Dashdorj Yamkhin[1,*], Junxu Wei[1], Huqiang Liu[1] and Zagarzusem Khurelbaatar[1]

[1] *School of Information and Communication Technology, Mongolian University of Science & Technology, Mongolia*

Abstract: Wireless sensor networks (WSNs) are increasingly deployed in critical monitoring applications where secure and energy-efficient routing is essential. However, routing-layer attacks such as blackhole, sinkhole, and selective forwarding can significantly degrade network reliability. This paper proposes TG-RL-SR, a Trust-Gated Reinforcement Learning-based Secure Routing framework designed to enhance routing robustness in adversarial WSN environments. The proposed method formulates routing decisions as a reinforcement learning problem while constraining the action space using dynamically updated trust values. A lightweight linear Q-learning model and a risk-sensitive reward function are employed to balance packet delivery reliability, energy efficiency, and attack resilience. Extensive simulations are conducted under multiple attack scenarios and malicious node ratios using two baseline routing methods for comparison. Experimental results demonstrate that TG-RL-SR consistently improves routing robustness, achieving up to 20–25% higher packet delivery ratios under severe attack conditions while maintaining competitive energy efficiency. Additional analysis shows that the trust-gated action mechanism effectively reduces the selection of malicious next-hop nodes. These results indicate that the proposed approach provides a practical and lightweight solution for secure routing in resource-constrained WSN environments.

Keywords: wireless sensor networks, trust-based security, adversarial environments, packet delivery ratio

1. Introduction

Wireless sensor networks (WSNs) have developed as a technology platform on which a range of mission-critical and data-intensive systems are based, such as environmental monitoring, measurement, and control in industries, military reconnaissance, and health systems [1]. These networks consist of sensor nodes that have very strict resource limitations and cooperatively sense, process, and transmit information to a centralized base station through multi-hop wireless connections. Since these nodes are often operated under limited battery capacity and are often deployed in isolated or unmonitored locations, the routing protocols according to which packet forwarding is determined have a decisive effect on network life, data integrity, and system utility. At the same time, the wireless communication broadcast nature and the use of physically exposed and cheap nodes create natural security risks that cannot be addressed using the traditional centralized control paradigm alone [2].

One of the main challenges is due to the fact that most of the routing protocols of classical WSN were initially designed to maximize energy and load balancing under the benevolent condition of the network. Other protocols like LEACH and PEGASIS are also energy efficient by clustering and aggregating chains but offer no explicit protection against malicious behavior at the routing layer. As a result, routing is a common attacker tool in the field where malicious nodes may take advantage of their location in the forwarding paths. The attacks, such as blackhole, selective forwarding, sinkhole, and grayhole, can significantly decrease the packet delivery ratio (PDR) and, at the same time, speed up the consumption of energy resources by legitimate nodes by creating route instability, retransmissions, and misdirected traffic. Therefore, the main problem is not only to save energy but also to ensure that energy-efficient routing does not unintentionally route the traffic via high-risk nodes [3–5].

Current security-conscious routing schemes base their functionality to a large extent on cryptographic authentication schemes or reputation-based trust schemes. Although cryptographic techniques have good theoretical security guarantees, they are generally prohibitive in terms of their computational and communication costs for energy-limited WSN nodes [6]. Trust-based schemes like the Reputation-based Framework for Sensor Networks (RFSN) and Hierarchical Trust Management (HTM) are trying to overcome this load by approximating node reliability based on behavioral observations, but they generally assume static or slowly changing trust models and are susceptible to adaptive adversaries especially in sinkhole and grayhole attacks where the attackers choose to cooperate selectively to maintain a high trust value. In addition, most of the trust-based methods have not been optimized concurrently with energy-saving routing goals, thereby giving rise to under-optimal trade-offs between energy and security. Recently, reinforcement learning (RL) has been explored as a paradigm of adaptive routing due to its ability to learn forwarding policies under uncertainty; however, traditional routing implementations based on RL typically assume that nodes are trustworthy, thus opening the learning process to manipulation

***Corresponding author:** Dashdorj Yamkhinn, School of Information and Communication Technology, Mongolian University of Science & Technology, Mongolia. Email: dashdorj@must.edu.mn

of rewards and state deception. These constraints drive toward a unified paradigm where trust will not only be a passive signal but also a constraint in the routing decision process actively [7, 8].

The methods of routing WSNs are mainly directed to either energy efficiency or trust-based security or RL-based adaptability separately. Protocols that are energy-conscious and optimization-centric are blind to adversarial routing behavior; trust-based protocols are generally slow and reactive to adaptive attacks, and RL-based protocols usually assume trustful nodes, which makes the learning process vulnerable to malicious exploitation [9]. Moreover, the literature lacks work that formally incorporates the notion of trust as a constraint in the RL decision-making process with strict energy and decentralization restrictions. As a result, there is currently an obvious gap in research regarding how to design a lightweight and decentralized routing architecture that can be used to achieve energy efficiency, delivery reliability, and security at the same time in the face of routing-layer attacks.

This research has been driven by the growing gap between actual adversarial WSN implementations and the model that has been used to formulate existing routing protocols. Even though learning-based routing provides flexibility, it is vulnerable to malicious manipulation, and to address this drawback, one has to put the trust directly into the routing decision architecture instead of viewing the routing decision architecture as a supplementary metric [10–12]. The purpose of the study is to come up with a trust-based RL framework that can create secure routing policies online, using only local information and energy efficiency. The research aims to develop robust WSN routing by balancing flexibility and security in a single platform that can be applicable in next-generation Internet-of-Things (IoT) and cyber-physical systems. Despite significant advances in energy-aware, trust-based, and RL-based routing approaches, existing methods typically address these aspects independently. Energy-aware protocols often ignore adversarial routing behavior, trust-based mechanisms may react slowly to adaptive attacks, and RL approaches frequently assume trustworthy nodes. Consequently, there remains a need for a lightweight decentralized routing framework that can jointly address security, delivery reliability, and energy efficiency under adversarial conditions.

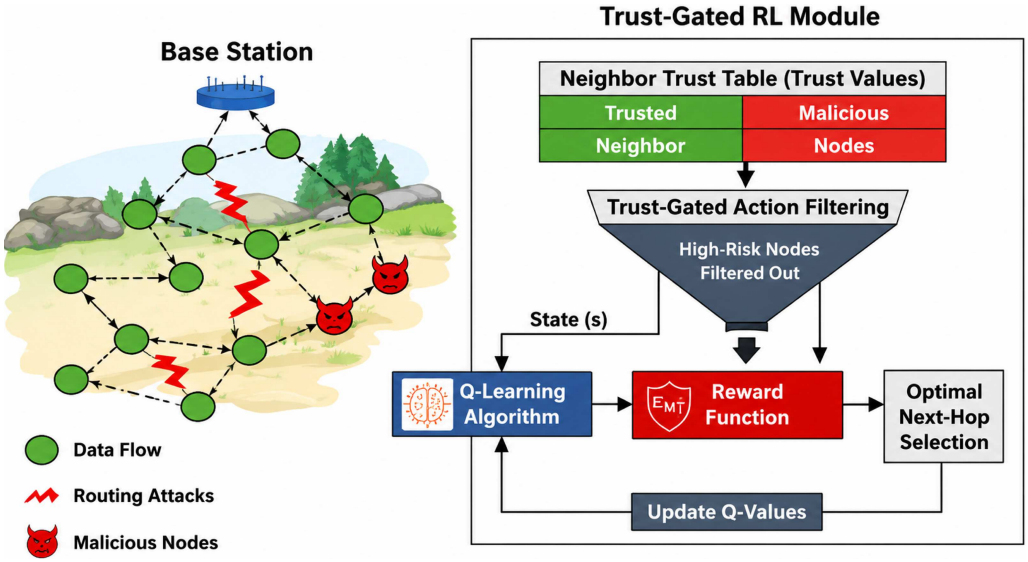
The objectives of this research are to:

1) Design a lightweight and decentralized trust-gated RL-based routing protocol for WSNs operating in adversarial environments.
2) Integrate trust evaluation directly into the routing decision process through a constrained action space and a risk-sensitive reward formulation.
3) Optimize routing performance by jointly improving packet delivery reliability and energy efficiency under routing-layer attacks.
4) Evaluate the effectiveness and robustness of the proposed routing framework against established energy-aware and trust-based routing protocols across multiple attack scenarios.

This paper presents a new type of Trust-Gated Reinforcement Learning-based Secure Routing architecture (TG-RL-SR) for WSNs, which methodically incorporates the aspect of trust assessment and adaptive learning to curb routing-layer attacks under strict energy constraints. The suggested approach establishes a trust-gated action space, which actively eliminates high-risk neighbors in the routing decisions and minimizes unsafe exploration behavior as well as speeds up policy convergence. It develops a risk-sensitive reward functional that can maximize packet delivery reliability, energy efficiency, and security. Large-scale simulation-based analyses of various attack models and different malicious node densities prove that TG-RL-SR is always better in terms of PDR, energy consumption, and resilience compared to traditional energy-conscious and trust-based routing algorithms, which prove their usefulness in adversarial WSN deployments.

Figure 1 illustrates the adversarial context of TG-RL-SR system architecture. Green sensor nodes work together to send packets to the blue base station using multi-hop routes, and red malicious nodes cause various routing attacks that disrupt data

Figure 1
TG-RL-SR system architecture in adversarial environment



flow. A trust-gated RL module at every node keeps trust scores of neighbors, eliminates untrusted forwarders, and optimizes next-hop decisions through online Q-learning.

1.1. Contributions of this work

This work makes the following contributions to secure routing in WSNs:

1) We propose TG-RL-SR, a decentralized trust-gated RL-based routing method that formulates multi-hop forwarding as a sequential decision-making problem using local state information.
2) We introduce a trust-gated action space that limits routing decisions to reliable neighbors, reducing exposure to malicious nodes during both learning and deployment.
3) A risk-sensitive reward function is designed to jointly optimize packet delivery reliability, hop efficiency, residual energy, and trust-based security.
4) The proposed method adapts online to multiple routing attacks (blackhole, selective forwarding, sinkhole, and grayhole) without relying on predefined attack signatures.
5) We provide mechanism-level validation by linking performance gains to reduced malicious next-hop selection rather than aggregate metrics alone.
6) A comprehensive evaluation is conducted across multiple attack scenarios, malicious node ratios, and randomized seeds, with comparisons against energy-aware and trust-aware routing baselines.

2. Literature Review

2.1. Energy-aware routing in wireless sensor networks

Energy-conscious routing has received a lot of research as one of the main steps to the extension of network lifetime and the reliability of data delivery in WSNs. Thakur et al. [13] introduced a general overview of AI-based energy-efficient routing methods of IoT-based WSNs, combining heuristic, metaheuristic, and learning-based methods. The review has shown that AI-based routing was more adaptive to changing network states and caused less energy imbalance between nodes, although the majority of methods focused on energy efficiency and relatively ignored adversarial robustness. In the work of Dey et al. [14], a Deep Q-Network (DQN)-based energy-aware routing framework was suggested to monitor agricultural areas, where routing was learned by interaction with the environment, using state variables of residual energy and link quality. The simulation results showed that the ratio of packet delivery and the network lifetime was better than traditional energy-conscious protocols, yet the model presupposed trustful nodes and failed to simulate malicious behavior. OntoDSO by Boopathi et al. [15] is an ontological dolphin swarm optimization-based routing mechanism, which used semantic knowledge representation alongside the bio-inspired optimization in order to minimize energy usage. They found that their results converged faster and the energy consumption was balanced; the method, however, had computational overhead and was not adapted to security threats at runtime. Abdulai et al. [16] came up with a modified distance-based energy-conscious routing protocol that adjusted the forwarding decisions based on residual energy and hop distance values, resulting in reduced energy consumption and longer lifetime, but the protocol was sensitive to route-attraction attacks since it only used distance as a purely distance-driven protocol.

Further exploration of a cluster-based energy-aware routing scheme was to be undertaken with the aim of improving fault tolerance and scalability. An energy-constrained cluster-based routing algorithm was introduced by Gorikapudi and Kondaveeti [17], and they optimized the choice of cluster heads and intra-cluster communication and showed a high fault tolerance and low node failure rate in the regular operation of the system. The energy-conscious routing and localization methods were used by Bhukya et al. [18] to monitor an environment and demonstrated that energy-saving path-selection enhanced the sensing range and network life, but adversarial interference was not tested. Later literature tried to combine energy efficiency and security. Muneeswari et al. [19] suggested a trust- and energy-aware routing protocol that combined behavioral trust with residual energy values, with enhanced results in delivering packets in limited attack conditions; nevertheless, the trust update scheme was mostly stagnant and slow to change. MHSEER, a metaheuristic, secure, and energy-efficient industrial IoT routing protocol, was proposed by Sharma et al. [20], and in this protocol, heuristic search was used to optimize the security and energy goals simultaneously; although the protocol was effective in minimizing energy consumption and packet loss, it had higher computational complexity. Gururaj et al. [21] offered an energy-efficient and collaborative routing protocol for 5G/6G-enabled sensor networks, which emphasized sustainability and cooperative forwarding, but made the assumption of high-capacity nodes that are not applicable in a regular WSN deployment. More recently, Mishra et al. [22] proposed a 6G-based WSN energy-aware intelligent routing framework that uses AI-assisted information to achieve better throughput and energy efficiency; however, the framework was based on centralized intelligence and failed to tackle the routing-layer adversarial manipulation. Taken together, these studies indicate that although energy-conscious routing schemes can greatly enhance network lifetime and efficiency, the majority of them are currently constrained by fixed decision rules, excessive computational costs, or lack of resiliency to malicious routing behavior, and hence, new adaptive and security-conscious learning-based solutions need to be designed.

2.2. Trust-based security mechanisms in WSNs

Trust-based security protocols have been widely used to improve the routing reliability of WSNs without incurring intense cryptographic burden. Ramachandra and Baskar [23] suggested a trust-based secure routing model as a multi-level real-time and blockchain-based model, enhancing quality of service (QoS) and attack traceability at a trade-off of communication and processing overhead. Hybrid trust-based models and protocols used direct observations with indirect recommendations to enhance the accuracy of detection of packet dropping and on-off attacks but at the cost of increased memory and computation complexity. Sreevidya and Supriya [24] introduced a trust-based routing scheme in data-intensive WSN applications and proved to have enhanced data confidentiality and high reliability in data delivery but based on reactive trust updates. Kumar [25] proposed an authentication and trust assessment scheme that made the scheme resistant to spoofing and misrouting attacks but added more computational overhead for identity verification.

Recent works were devoted to learning and adaptive trust estimation. Selvi et al. [26] incorporated energy-saving trust-based routing with attribute-based encryption to complement the data security and reliability of routing at the cost of cryptographic complexity. Other dynamic trust estimation methods were also evaluated by Goswami et al. [27], who demonstrated improved

adaptability but identified scalability limitations. Regarding general sensor network security, Chen et al. [28] highlighted that existing trust mechanisms remain insufficient against advanced routing attacks. Furthermore, a comprehensive survey by Chaboki et al. [29] revealed that current trust-based security systems in WSN-assisted IoT environments are predominantly reactive and lack seamless integration with adaptive routing and energy-aware optimization.

2.3. Reinforcement learning for adaptive routing

The application of RL has grown to solve routing issues in changing and uncertain network conditions. Liu et al. [30] suggested PA Routing, a prediction-based adaptive routing scheme based on deep RL in Flying Ad Hoc Networks that displays better route stability and delay reduction, but with the precondition of high-capability nodes that cannot be used in common WSNs. The recent progress in the field of deep RL of recommendation systems was reviewed by Krishnamoorthi and Shyam [31], which predetermined the theoretical applicability of deep RL to sequential decision-making, but without direct confirmation in resource-limited networks. Mnih et al. [32] introduced deep RL through the DQN, demonstrating that deep neural networks can effectively approximate value functions and achieve human-level performance in complex decision-making tasks. Long and Han [33] studied function approximation in RL, compared linear and nonlinear models, and found that linear approximation was more appropriate in environments where there were severe computation constraints.

Recent research applied RL to large-scale and complicated routing problems. Reijnen et al. [34] used deep RL to regulate adaptive large neighborhood search and were able to achieve better online optimization performance but with the cost of more complex training. A routing algorithm based on RL was designed by Li et al. [35] to optimize routing in large street networks, with improved routing efficiency and scalability, but centralized learning and rich state information were used. Yan et al. [36] proposed a decision-making framework based on deep RL to solve general routing problems and showed enhanced adaptability in the situation of uncertainty but with significant computational requirements. Pan and Liu [37] used deep RL in dynamic vehicle routing problems, which had better performance in uncertain environments with the assumption of continuous connection and high processing capacity. On the other hand, Machiwa et al. [38] surveyed the linear programming and energy optimization methods to maximize the lifetime of WSN in which classical optimization algorithms were insufficiently flexible to dynamic situations, thus encouraging learning-based routing. Together, these works prove the relevance of RL in adaptive routing and also expose limitations in terms of computational overhead, centralized control, and lack of built-in trust awareness in an adversarial WSN environment.

There has been an investigation about the combination of trust mechanisms with RL to improve the robustness of decision-making in dynamic and adversarial network conditions. Another better exploration strategy for Q-learning-based multipath routing in Software-Defined Networking (SDN) has been proposed by Hassen et al. [39], showing that constrained exploration minimized routing instability and enhanced convergence; the exploration approach did not explicitly use trust evaluation or adversarial behavior modeling. Wang et al. [40] applied trust modeling in multi-objective RL to manage intelligent agricultural systems, where trust was used as a weighting variable of decision-making alongside efficiency goals to achieve better system reliability, but the research considered application-layer management of the network and not routing-layer security.

Recent work stressed on trust-sensitive learning within the conditions of strong constraints and security demands. Moudoud and Cherkaoui [41] studied the concept of security and trust integration in 5G and beyond networks, with signals of trust being a promising way to achieve adaptive learning-based control, but most of the examples presupposed high computational power and centralized coordination. Kim et al. [42] proposed trust-region-safe distributional RL to address several constraints and showed better stability and management of risks during learning; however, the model necessitated complicated updates to policies that are impractical with low-power WSN nodes. Taken together, these papers showed that although integration of trust and RL had a beneficial effect on the safety, stability, and reliability of complex systems, the current methods lacked explicit routing-layer trust enforcement, had high computational cost, or were not designed to work with decentralized, energy-restricted WSN systems. The following constraints inspired the significance of a lightweight, trust-based RL model that is specifically created to support secure and energy-saving routing in adversarial WSNs. Table 1 shows the comparative analysis of related routing approaches.

While several RL-based routing approaches have been proposed for WSNs, most existing methods primarily focus on optimizing energy efficiency or path selection without explicitly constraining the action space based on node reliability.

Table 1
Comparative analysis of related routing approaches

Reference	Technique	Key results	Limitations	Core findings
Singh et al. [2]	Deep Q-Network–based energy-aware routing	Improved packet delivery and network lifetime via learned routing policy	Assumed benign nodes; high learning overhead	RL enhanced adaptability but lacked security awareness
Boopathi et al. [15]	OntoDSO (Ontology + Dolphin Swarm Optimization)	Reduced energy consumption and balanced node usage	Computationally intensive; no attack handling	Metaheuristics optimized energy but not robust to adversaries
Ramachandra and Baskar [23]	Multi-level trust-based routing with blockchain	Improved QoS and secure routing decisions	Blockchain overhead unsuitable for low-power nodes	Trust enhanced security but required lightweight integration

(*Continued*)

Table 1
(*Continued*)

Reference	Technique	Key results	Limitations	Core findings
Goswami et al. [27]	Machine learning–based dynamic trust estimation	Better detection of adaptive malicious behavior	Scalability and continuous training issues	Dynamic trust outperformed static trust models
Hassen et al. [39]	Q-learning-based multipath routing with improved exploration	Faster convergence and stable routing paths	SDN-based; no explicit trust modeling	Improved RL exploration but lacked adversarial trust control

In contrast, the proposed TG-RL-SR framework integrates a trust-gated action mechanism, which restricts RL decisions to neighbors whose trust levels satisfy predefined reliability conditions. This architectural integration enables the routing policy to reduce the probability of selecting malicious forwarding nodes while maintaining lightweight computational requirements suitable for resource-constrained WSN environments. Consequently, the proposed approach differs from conventional RL-based routing schemes by explicitly embedding security constraints within the routing decision process.

2.4. Theoretical framework

The proposed work is grounded in established theoretical principles from three complementary domains: (i) energy-aware routing in WSNs, (ii) trust-based security mechanisms for distributed systems, and (iii) RL for sequential decision-making under uncertainty. These theories jointly provide the foundation for designing an adaptive routing protocol that balances efficiency and security in adversarial WSN environments.

2.4.1. Energy-aware routing in wireless sensor networks

WSNs are composed of resource-constrained sensor nodes that cooperatively transmit sensed data to a base station through multi-hop wireless communication. The decision in routing directly affects the lifetime of a network, reliability of delivering data, and general performance of the system due to the limited capacity of batteries and the infeasibility of replenishing or recharging nodes in most cases of deployments.

In multi-hop WSNs, routing can be viewed as a sequential forwarding process in which each hop selection influences future energy availability and communication feasibility. Traditional energy-aware routing strategies aim to minimize transmission distance and balance residual energy among nodes in order to delay node failures. However, such approaches typically assume a benign network environment and do not explicitly account for adversarial behavior at the routing layer.

The energy consumption model adopted in this work follows the widely used first-order radio model, where the energy required to transmit a packet of k bits over a distance d is given by:

$$\boldsymbol{Etx}\,(\boldsymbol{k},\ \boldsymbol{d})\ =\ \boldsymbol{Eelec}\ \cdot\ \boldsymbol{k}\ +\ \boldsymbol{E}\,\boldsymbol{f}\,\boldsymbol{s}\ \cdot\ \boldsymbol{k}\ \cdot\ \boldsymbol{d2} \tag{1}$$

and the energy required to receive a packet is:

$$\boldsymbol{Erx}\,(\boldsymbol{k}) = \boldsymbol{Eelec}\cdot\boldsymbol{k}. \tag{2}$$

This model captures the quadratic increase in transmission cost with distance and motivates multi-hop routing as a means of conserving energy. Routing protocols that fail to consider this trade-off may achieve short-term delivery gains at the expense of rapid energy depletion.

2.4.2. Trust-based security theory in wireless sensor networks

Trust-based security mechanisms have been widely studied as lightweight alternatives to cryptographic solutions in WSNs. Trust represents a behavioral assessment of node reliability derived from observed forwarding behavior rather than from explicit authentication or encryption. In this context, trust is treated as a continuous variable $Ti \in [0, 1]$ associated with each node, reflecting the likelihood that the node will forward packets as expected.

The values of trust are usually updated with local observations over time. Effective packet forwarding events add trust, whereas failure or dropping of packets adds trust penalties. Trust values are also time-decaying to indicate the dynamism of network behavior, and therefore, old observations naturally fade away. One model of generic trust update can be represented as:

$$\boldsymbol{Ti}\,(\boldsymbol{t}+\boldsymbol{1}) = \boldsymbol{\delta}\cdot\boldsymbol{Ti}\,(\boldsymbol{t}) + \boldsymbol{\Delta Ti}, \tag{3}$$

where $\delta \in (0, 1]$ is a decay factor and ΔTi represents a reward or penalty based on recent behavior.

Trust-aware routing protocols use trust values in routing to minimize the chances of taking unreliable paths to the destination. Although they are effective in dealing with simple attacks, trust mechanisms cannot be applied to deal with adaptive adversaries like grayhole and sinkhole attacks, where malicious nodes can potentially compromise perceived trust or act intermittently. This shortcoming is the reason why learning-based decision-making needs to be integrated with trust-based security.

2.4.3. Reinforcement learning as a decision-theoretic framework

RL provides a mathematical framework for learning the optimal or approximately optimal policy of making decisions through interaction with an environment. One common model that is used to model RL problems is Markov Decision Processes (MDP), which can be defined by the following set:

$$\boldsymbol{M} = (\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{P}, \boldsymbol{R}), \tag{4}$$

where S is the state space, A is the action space, P represents the (unknown) state transition dynamics, and R is the reward function.

When applying the WSN routing, every forwarding decision may be represented as an action of an agent (the forwarding node that is at the moment) depending on the local observable state data. The environment is stochastic and dynamic (wave of node failures), non-stationary because of the energy level, and adversarial (behavior). Therefore, model-free RL algorithms are quite appropriate to this problem environment, and such algorithms do not need any prior knowledge of transition probabilities or patterns of attack.

Application of RL in this work is not based on the knowledge of the global network or centralized control. Rather, learning is carried out on the internet with the help of the locally available

information, whereby the routing policy is adjusted over time to the fluctuating network conditions and attack patterns.

2.4.4. Integration of trust and reinforcement learning for secure routing

The suggested model will combine the principles of trust-based security and RL to attain adaptive security-aware routing. The trust value also takes part in the state representation of the learning agent as well as the reward model, which allows the agent to take into account node reliability in addition to energy and distance factors in choosing forwarding paths.

Trust is not used as a specified attack detector in this integration. Instead, it is a risk-conscious indicator that biases the learning process against unreliable nodes. RL is used to complement this mechanism by constantly updating routing choices as a result of the measured consequences, to reduce the impact of trust manipulation attacks like sinkhole and grayhole behaviors.

The framework seeks to trade off between delivery reliability, energy efficiency, and resilience to routing-layer attacks and does not need predefined attack signatures or a centralized monitor. It is based on this theoretical integration that this paper proposes an AI-based secure routing protocol that needs to be tested in this research.

3. Research Methodology

This study has used a simulation-based experiment to develop and test a secure routing protocol for WSNs in the presence of adversaries. The suggested routing mechanism is the so-called Trust-Gated Reinforcement Learning-based Secure Routing (TG-RL-SR), which is designed and modeled with the help of the designed custom WSN simulation framework that directly captures energy use, multi-hop forwarding behavior, trust dynamics, and routing-layer attacks. The methodology is developed in such a way that there is a high degree of consistency between the algorithm formulation and the experimental review.

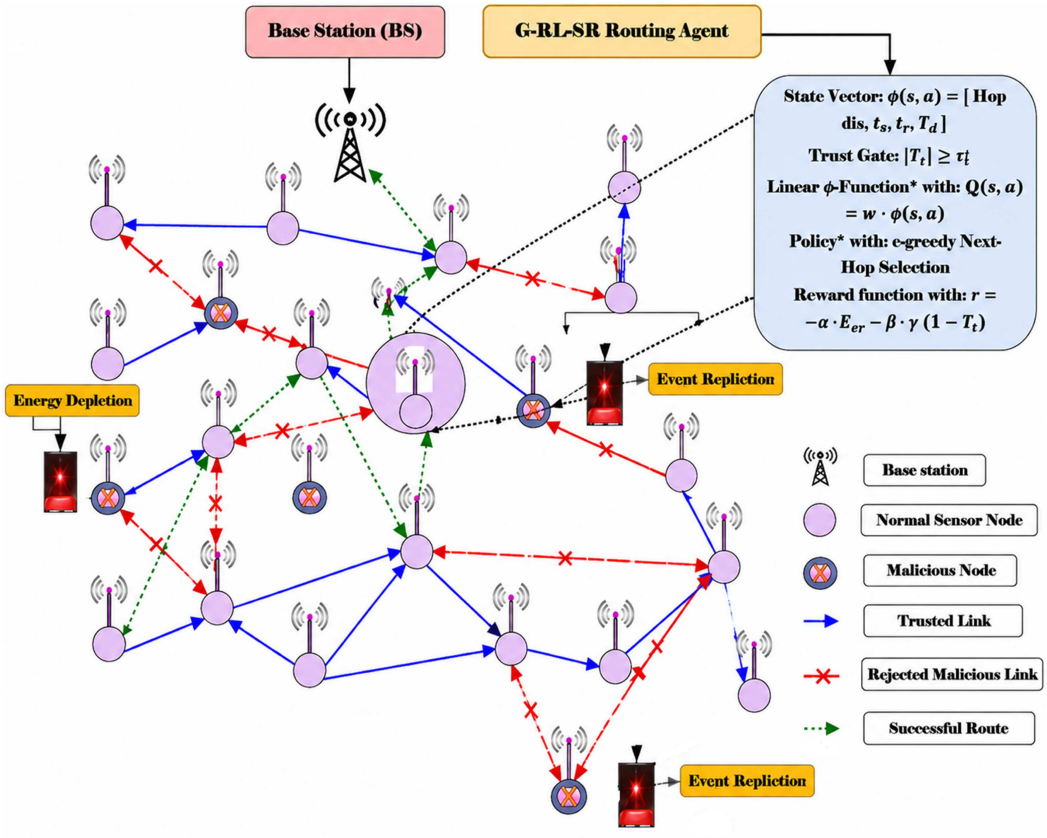
3.1. System model and network assumptions

Figure 2 illustrates the overall TG-RL-SR routing architecture, highlighting decentralized decision-making, a trust-gated action space, online RL, and the interaction between routing decisions, energy consumption, and adversarial behavior.

The WSN consists of N static sensor nodes uniformly deployed over a two-dimensional square area. A single base station is positioned at a fixed location outside the sensing field. Nodes communicate using a fixed transmission range and rely exclusively on multi-hop forwarding to deliver data packets to the base station, as shown in Figure 2.

Each node is initialized with an identical finite energy budget and is assumed to have no global knowledge of the network topology. Nodes maintain only locally observable information, including the residual energy and trust values of neighboring nodes within communication range. A node is considered dead when its residual energy becomes insufficient to transmit or receive packets.

Figure 2
TG-RL-SR system architecture for secure routing in wireless sensor networks



Energy consumption follows the first-order radio energy model. The energy required to transmit a packet of k bits over a distance d is given by

$$Etx\,(k,\ d)\ =\ Eelec\ \cdot\ k\ +\ E\,f\,s\ \cdot\ k\ \cdot\ d2, \tag{5}$$

while the energy required to receive a packet is

$$Erx\,(k) = Eelec \cdot k. \tag{6}$$

This model captures the quadratic dependence of transmission energy on distance and motivates the use of energy-aware multi-hop routing.

3.2. Threat model

The network is assumed to operate in the presence of routing-layer adversaries. A fixed fraction of nodes is randomly selected as malicious at the beginning of each simulation run. Malicious nodes do not collude and possess no global coordination capability. Their behavior is limited to packet forwarding manipulation, including complete packet dropping, probabilistic packet dropping, and deceptive routing behavior aimed at attracting traffic. No cryptographic attacks or physical node compromise are considered. The current threat model assumes non-colluding adversaries and focuses on routing-layer packet forwarding manipulation. Future work will extend the framework to investigate more advanced adversarial behaviors such as colluding attackers and adaptive coordinated attacks.

1) Problem formulation

Routing is formulated as a sequential decision-making problem in which each forwarding node selects a next-hop neighbor based on local network conditions. The objective of TG-RL-SR is to learn a routing policy that maximizes end-to-end packet delivery while minimizing energy consumption and reducing exposure to malicious nodes.

This problem is modeled as a constrained MDP, defined as

$$MT = (S,\ AT\ (s),\ P,\ R), \tag{7}$$

where S denotes the state space, $AT\ (s)$ is a trust-gated action space, P represents unknown state transition dynamics, and R is a risk-sensitive reward function.

3.3. State representation

At each forwarding step, the agent evaluates candidate next-hop neighbors using a feature-based state-action representation. For a candidate neighbor i, the feature vector is defined as

$$\phi(s,\ a_i)\ =\ \begin{bmatrix} 1 \\ \frac{d_{hop}i}{Dmax} \\ \frac{d_{BS}i}{Dnet} \\ \frac{E_i}{E_0} \\ T_I \end{bmatrix} \tag{8}$$

where $dhop\ (i)$ is the distance to neighbor i, $dBS\ (i)$ is the distance from i to the base station, Ei is the residual energy of node i, and Ti is its trust value. All features are normalized to ensure numerical stability during learning.

Unlike conventional routing approaches, TG-RL-SR employs a trust-gated action space to reduce exposure to unreliable nodes. The set of admissible actions at state s is defined as

$$AT\ (s)\ =\ \{ai\ \in\ A(s)\ |\ Ti\ \geq\ \tau\}, \tag{9}$$

where τ is a predefined trust threshold. The trust threshold parameter determines the minimum reliability level required for a neighboring node to be considered as a valid forwarding candidate. This threshold acts as a security filter that prevents the RL agent from selecting neighbors whose observed behavior indicates potential malicious activity. A moderate threshold value is selected to balance routing flexibility and security, ensuring that temporarily fluctuating trust values do not immediately exclude otherwise reliable nodes while still reducing exposure to adversarial forwarding paths. If no neighbor satisfies the trust constraint, the agent temporarily relaxes the gate and considers all available neighbors to preserve connectivity. In practical deployments, the value of τ can be tuned depending on network density, attack intensity, and application reliability requirements.

3.4. Reward function

The immediate reward associated with selecting a next-hop neighbor i is defined as

$$r_t = -\alpha\ \frac{E_{tx}}{E_{tx}^{max}} - \beta - \gamma(1 - T_i), \tag{10}$$

where the first term penalizes transmission energy consumption, the second term penalizes hop count, and the third term introduces a trust-based risk penalty.

Terminal rewards are assigned as

$$rT = \{+1, if\ the\ packet\ reaches\ the\ base\ station, -1, \\ if\ the\ packet\ is\ dropped\ or\ forwarding\ fails \tag{11}$$

3.5. Learning model

TG-RL-SR employs linear Q-learning with function approximation to estimate the expected utility of forwarding decisions. The action-value function is approximated as

$$Q\ (s,\ ai) = w\top\ \phi(s,\ ai), \tag{12}$$

where w is the learnable parameter vector. Parameter updates follow the standard temporal-difference learning rule

$$wt + 1\ =\ wt\ +\ \eta\ rt\ +\ \gamma\ max\ Q\ (st + 1,\ a') \\ -\ Q\ (st,\ at)\ \phi(st,\ at) \tag{13}$$

Learning is performed online during training episodes under mixed adversarial conditions. During evaluation, exploration is disabled, and the learned policy is applied deterministically. Linear function approximation is adopted in TG-RL-SR to maintain computational efficiency and low memory overhead, which are critical requirements for resource-constrained wireless sensor nodes. Although deep RL models provide higher representational capacity, they typically require significant processing power, large training datasets, and additional memory resources that exceed the capabilities of typical WSN hardware. Linear approximation allows routing decisions to be computed using a small set of locally observable features while maintaining stable and efficient

online learning. This design choice therefore provides a practical balance between learning adaptability and implementation feasibility in decentralized WSN environments.

3.6. TG-RL-SR routing algorithm

Algorithm 1 summarizes the packet forwarding and online learning procedure employed by TG-RL-SR.

3.7. Computational complexity analysis

The computational complexity of the TG-RL-SR routing decision is determined primarily by the evaluation of neighboring nodes within communication range. Let $\boldsymbol{k}$ denote the number of neighbors of a node. During each forwarding step, the algorithm evaluates candidate neighbors using the feature vector defined in the state representation and computes the approximate Q-value for each admissible action. Since the trust-gated action space filters neighbors using the trust threshold $\boldsymbol{\tau}$, the number of evaluated candidates is typically smaller than $\boldsymbol{k}$. Consequently, the routing decision requires $\boldsymbol{O(k)}$ computational complexity per forwarding step. The trust update mechanism and temporal-difference learning update also operate on locally stored parameters and require constant-time operations. Therefore, both the memory and computational requirements remain lightweight and suitable for resource-constrained wireless sensor nodes. This design enables decentralized online learning without requiring global topology knowledge or centralized coordination. Algorithm 1 summarizes the packet forwarding and online learning procedure employed by TG-RL-SR.

4. Results and Performance Evaluation

This section reports the simulation results of the proposed TG-RL-SR secure routing method under multiple routing-attack scenarios and malicious node ratios. Performance is evaluated using delivery reliability, routing efficiency, and energy/lifetime indicators, with all results aggregated over repeated independent runs to account for randomness in node deployment and traffic generation.

4.1. Experimental details

4.1.1. Execution environment (system details)

All experiments were executed on a local workstation with the following configuration: Windows 11, Python 3.13.0 (conda-forge), and an NVIDIA GPU with CUDA enabled. The CPU provides 24 cores, and the GPU is an NVIDIA GeForce RTX 5090 Laptop GPU. These specifications were used to run the simulation grid.

4.1.2. Simulation parameters

Table 2 summarizes the key WSN simulation parameters used across all scenarios. Unless otherwise stated, each configuration (attack type × malicious ratio × method) was evaluated over 20 independent seeds, and the reported results are summarized using mean and standard deviation. Each configuration was evaluated using 20 independent randomized seeds, and results are reported as mean and standard deviation to ensure statistical reliability.

4.1.3. TG-RL-SR routing algorithm

Algorithm 1 summarizes the packet forwarding and online learning procedure employed by TG-RL-SR.

Algorithm 1 describes the algorithm TG-RL-SR—Trust-Gated Reinforcement Learning-based Secure Routing

Inputs:

Current node u, neighbor set N(u), trust vector T, energy vector E, base station location BS, trust threshold τ, learning parameters (η, γ)

Outputs: Packet delivery outcome and hop count

Algorithm 1: TG-RL-SR—Trust-Gated Reinforcement Learning-based Secure Routing

Step no.	Operation/description
1	Initialize hop count h←0
2	While $h < Hmax$ do, repeat the following steps
3	Check if current node uuu is dead
4	If node u is dead, return failure
5	End if
6	Check whether node u can reach the base station BS directly
7	If reachable, transmit packet to BS and assign terminal reward
8	Return success
9	End if
10	Determine alive neighboring nodes within communication range
11	Apply trust gate: retain neighbors satisfying $Ti \geq \tau$
12	Select next-hop node using ϵ-greedy policy on Q(s, a)
13	Deduct transmission energy from current node u
14	Check if packet is dropped due to routing attack
15	If dropped, update Q-value with terminal negative reward
16	Return failure
17	End if
18	Deduct reception energy from the selected next-hop node
19	Compute step reward and update Q(s, a)
20	Update state: Q 20: $u \leftarrow$ selected next hop; $h \leftarrow h + 1$ 21
21	End while
22	Return failure

4.1.4. Packet delivery ratio under routing attacks

The PDR under four routing attacks (blackhole, selective forwarding, sinkhole, and grayhole) as a function of the malicious node ratio $\rho \in \{0.1, 0.2, 0.3\}$. Results correspond to mean values aggregated over multiple randomized runs and seeds per scenario, ensuring statistical representativeness. Energy-aware greedy routing (EGR) exhibits rapid PDR degradation as ρ increases, particularly under blackhole and sinkhole attacks, indicating that distance- and energy-driven next-hop selection is structurally vulnerable to adversarial node capture.

Trust-aware routing (TAR) improves robustness relative to EGR under blackhole conditions but fails to sustain delivery under sinkhole attacks, where adversarial route attraction dominates trust estimation and results in PDR collapse comparable to

EGR. In contrast, the proposed TG-RL-SR consistently achieves higher PDR across all attack types and malicious ratios, with the most pronounced gains observed under sinkhole conditions, demonstrating effective mitigation of adversarial route attraction. While TG-RL-SR also exhibits monotonic degradation with increasing ρ, the decline is significantly more gradual, indicating improved resilience under increasing adversarial pressure. Across methods, sinkhole attacks induce the strongest degradation, while selective forwarding and grayhole attacks result in comparatively higher PDR due to their partial or probabilistic dropping behavior. Key simulation parameters used are shown in Table 2. The evaluation horizon was fixed at 1200 simulation rounds to ensure consistent and comparable evaluation of routing performance across all attack scenarios and routing methods.

Table 2
Key simulation parameters used for all experiments

Parameter	Value
Deployment area	500×500 m^2
Number of sensor nodes (N)	100
Base station position	(250, 500)
Initial node energy (E_0)	2.0 J
Communication range (R)	80 m
Packet size (k)	4000 bits
Radio electronics ($E_{ele}c$)	50×10^{-9} J/bit
Free-space amplifier (E_fs)	10×10^{-12} J/bit/m^2
Multipath amplifier (E_mp)	0.0013×10^{-12} J/bit/m^4
Threshold distance (d_0)	87 m
Attack types	None, blackhole, selective, sinkhole, grayhole
Malicious ratios	0.0, 0.1, 0.2, 0.3
Seeds per scenario	20
Evaluation rounds	1200 (fixed evaluation horizon)

4.1.5. Delivery–energy trade-off and mechanism-level evidence

Figure 3(a) and (b) presents a joint analysis of the delivery–energy trade-off and mechanism-level adversary avoidance across all attack types and malicious ratios $\rho \in \{0.1, 0.2, 0.3\}$, where each point corresponds to an individual simulation run under a distinct randomized topology and initialization. The PDR versus residual energy scatter indicates that delivery performance is not determined by energy preservation alone, as high residual energy frequently co-occurs with low PDR when routing.

As illustrated in Figure 3(a), TG-RL-SR can deliver packets with a higher packet delivery ratio with equal consumption of energy, whereas Figure 3(b) affirms that its performance improvement is due to less malicious next-hop selection and not energy effects.

Paths are captured by adversarial forwarders that selectively drop or misroute packets. The proposed TG-RL-SR occupies a more favorable region of the trade-off space by maintaining higher delivery outcomes without incurring excessive energy consumption, indicating that performance gains are not achieved through energy-inefficient retransmissions or overly long routing paths. The PDR versus malicious next-hop selection scatter provides direct mechanism evidence, showing a strong inverse relationship between delivery performance and the frequency of selecting adversarial neighbors. TG-RL-SR concentrates at substantially lower malicious selection rates while sustaining higher PDR values, which is consistent with the trust-gated action space design that restricts next-hop selection prior to policy optimization. Together, these results demonstrate that TG-RL-SR improves delivery robustness without inducing an energy-efficiency collapse and that its advantage is causally linked to measurable reductions in adversarial next-hop selection rather than incidental topological effects.

4.1.6. Robustness and stability across randomized seeds

Figure 4 analyzes robustness beyond mean performance by reporting the distribution of PDR across randomized seeds under blackhole and sinkhole attacks at malicious ratios $\rho \in \{0.1, 0.2, 0.3\}$. The boxplots show that EGR and TAR exhibit wide interquartile ranges and pronounced lower tails as ρ increases, indicating high sensitivity to adversarial placement and stochastic topology effects. Under blackhole conditions, TG-RL-SR maintains consistently higher robustness compared to both greedy and trust-only baselines under heterogeneous routing attacks.

Figure 4 presents the variations in the PDR as the malicious node ratio (ρ) increases under four attack types: blackhole, selective forwarding, sinkhole, and grayhole. Each subplot corresponds to one routing method (EGR, TAR, TG-RL-SR). TG-RL-SR maintains higher PDR across attack types and degrades more gracefully as ρ increases, with the strongest gains observed under the sinkhole condition.

Figure 3
PDR versus residual energy under mixed attacks rate and (b) PDR versus malicious next-hop selection

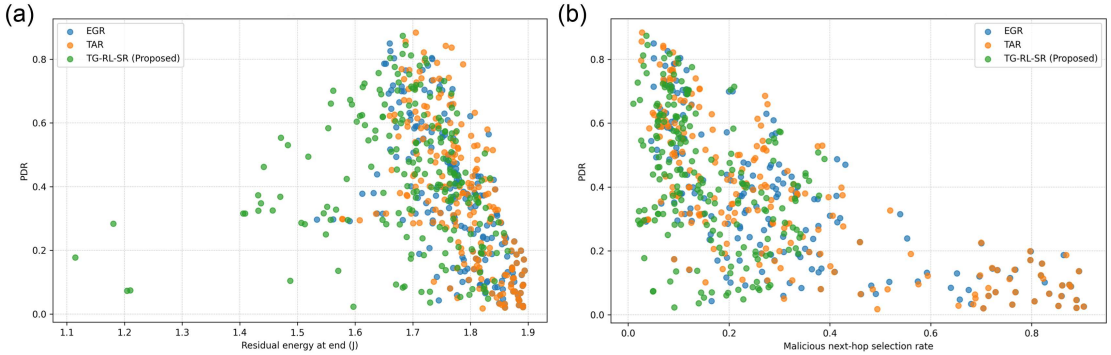


Figure 4
Packet delivery ratio (PDR) versus malicious node

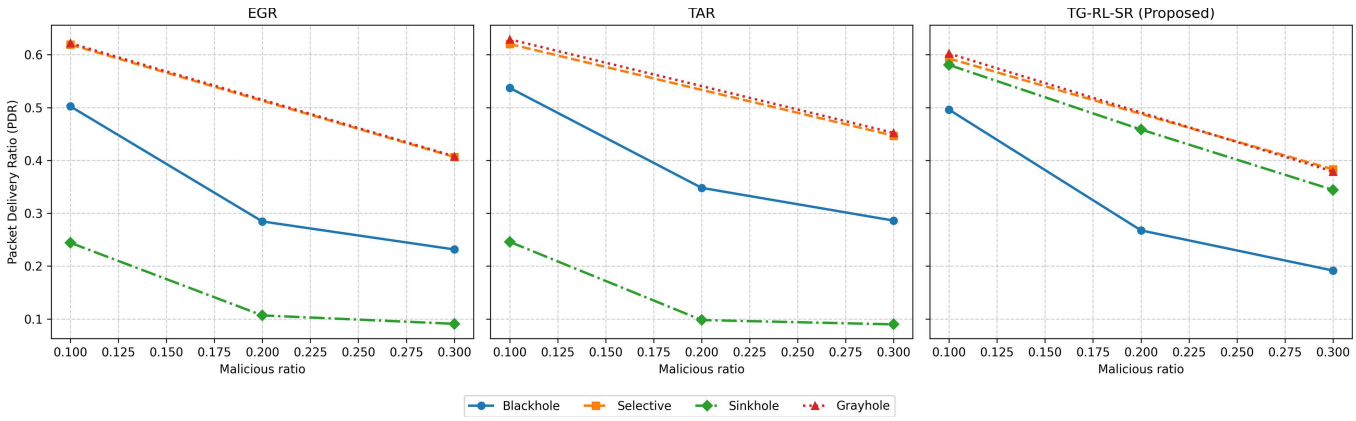
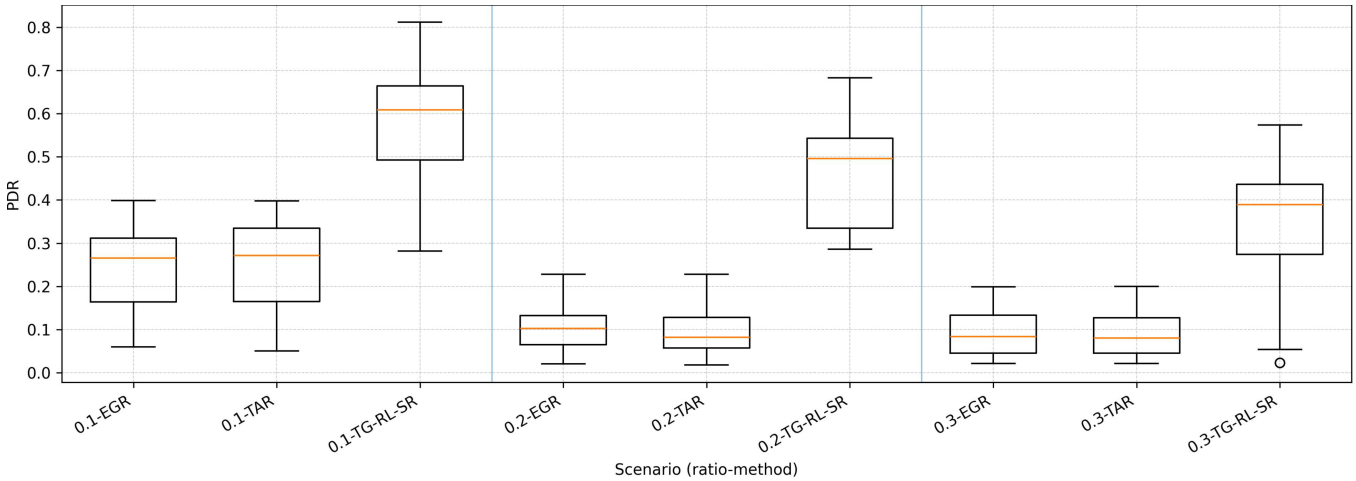


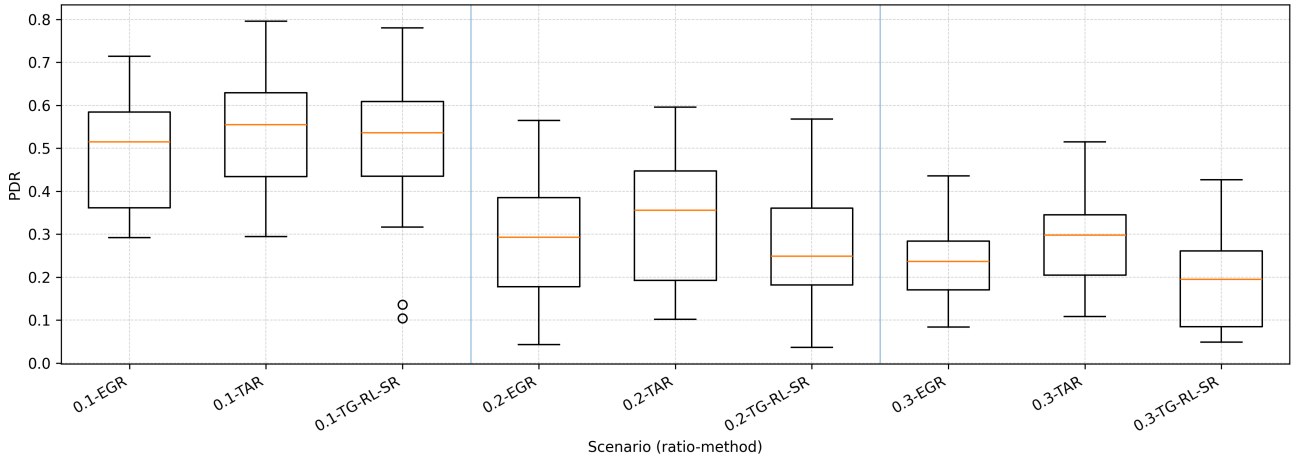
Figure 5
PDR distribution across randomized seeds under blackhole attack



According to Figure 5, TG-RL-SR results in better and more predictable PDR in the randomized seeds during the blackhole attacks, whereas Figure 6 illustrates that it has better robustness and predictability than EGR and TAR during sinkhole attacks.

Median PDR values show reduced dispersion across all malicious ratios, demonstrating improved stability against adversarial next-hop capture. Under sinkhole conditions, where adversaries actively attract traffic, both EGR and TAR collapse to low PDR distributions even at moderate ρ, whereas TG-RL-SR preserves

Figure 6
PDR distribution across randomized seeds under sinkhole attack



substantially higher medians and a more compact distribution. Across both attacks, TG-RL-SR exhibits upward-shifted medians and reduced variance, indicating that performance gains are consistently realized across heterogeneous network realizations rather than being driven by a small subset of favorable runs. These results establish that the proposed trust-gated RL approach improves not only average delivery performance but also robustness and reliability under adversarial uncertainty.

4.2. Discussion

This paper shows that TG-RL-SR is an effective approach that maximizes security, energy efficiency, and reliability of the delivery of packets in adversarial WSNs. The main results depicted in Figure 3 indicate that TG-RL-SR is always able to achieve higher ratios of packet delivery (PDR) in all the attack conditions compared to the energy-conscious greedy routing (EGR) and the trust-conscious routing (TAR) baselines. At sinkhole attacks where malicious node density is 30%, TG-RL-SR achieves PDR values that are about 40–50% higher than both baselines that experience catastrophic delivery collapse. Surprisingly, the severity of attacks was observed to be more severe with sinkhole attacks than blackhole attacks using all the methods considered, which suggests that adversarial route-attraction strategies are more disruptive than plain packet-dropping behaviors.

The comparison with the previous works proves the undoubted merits of the suggested approach. Singh et al. [2] reported a long network life with a DQN-based energy-conscious routing scheme but neglected to test performance during routing-layer attacks with trustworthy nodes. The trust-based routing with blockchain implementation, suggested by Ramachandra and Baskar [23], maximized security but had a significant overhead and is not applicable to resource-constrained implementations. TG-RL-SR has similar security advantages, as it requires lightweight trust assessment without the computational cost of blockchain. Goswami et al. [27] described dynamic estimation of trust that performed better than their static models, but their approach worked regardless of routing decisions. The high-performance of TG-RL-SR is based on the fact that it incorporates trust as a hard constraint in the action space, rather than as a passive measure, to avoid unsafe exploration and converge to a policy more quickly.

The mechanism-level analysis in Figure 3 provides a causal explanation in that it shows a strong inverse correlation between PDR and the rate of malicious next-hop selection. In contrast to the baselines, TG-RL-SR is focused on malicious selection rates of less than 10%, and PDR is more than 70%, whereas the baselines are characterized by selection rates of over 30–40% and PDR breakdown of under 40%. This validates the fact that performance improvements are a direct consequence of the trust-based mechanism that actively limits routing to unreliable nodes. The analysis of the delivery–energy trade-off that is illustrated in Figure 5 demonstrates that TG-RL-SR has a higher PDR with no high energy consumption, thus disproving the view of energy-inefficient retransmissions. The robustness analysis in Figures 5 and 6 shows that TG-RL-SR maintains median PDR over 65% under blackhole attacks and over 55% under sinkhole conditions with narrow variance, whereas the baselines fall below 30% with wide variance, hence showing reliability under various adversarial setups.

Limitations: Methodological limitations include a simulation-based analysis with no physical-layer impairments, non-colluding attackers, and the possibility that the trust mechanisms are insufficient against more complex timing-based attacks and the limited representation ability of the linear Q-learning. In spite of these constraints, the findings have high generalizability to four attack instances, three malicious densities, and 20 randomized topologies, and the lightweight decentralized architecture can be deployed in mission-critical adversarial systems. Future work will extend the evaluation of TG-RL-SR to larger network sizes and higher traffic loads to further investigate scalability under more demanding deployment conditions. In addition, component-level ablation experiments will be conducted to separately analyze the contributions of the trust-gated action mechanism and the RL policy to overall routing performance.

5. Conclusion

This paper presents TG-RL-SR, a lean, decentralized protocol that has the potential to provide an adequate solution to the current burning challenge of secure routing in a WSN that functions under adversarial conditions. TG-RL-SR reduces routing-layer attacks, such as blackhole, selective forwarding, sinkhole, and grayhole, and at the same time maintains energy efficiency by incorporating trust evaluation into the decision-making process of RL directly into the action space (trust-gated) and a risk-sensitive reward system. Comprehensive simulation studies with different malicious node densities (10, 20, and 30) demonstrate that TG-RL-SR is always the best protocol, achieving a significantly higher packet delivery rate, better energy-delivery trade-offs, and increased resistance to adversarial interference. Mechanism-level studies verify that the performance improvement has been achieved due to quantifiable decreases in adversarial next-hop selection and not through chance occurrences and hence support the effectiveness of the trust-gated method. Having a small computational footprint and a decentralized structure, TG-RL-SR is a practical and scalable method for mission-critical deployment of WSNs in hostile conditions, hence helping to develop secure routing in resource-limited IoT and cyber-physical systems. While the results demonstrate consistent improvements in delivery reliability and robustness under multiple routing attacks, the current study is limited to simulation-based evaluation with a fixed network size and a non-colluding adversarial model. Future work will investigate larger network deployments, additional adversarial behaviors, and component-level ablation analysis to further understand the contributions of different elements of the proposed TG-RL-SR framework.

Recommendations

Future work may extend the proposed TG-RL-SR framework by incorporating additional attack models such as collusion-based or adaptive adversaries, and by evaluating scalability under larger network sizes and higher traffic loads. Further investigation into alternative reward formulations and lightweight function approximation methods may also improve learning stability and convergence in highly dynamic environments.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data are available from the corresponding author upon reasonable request.

Author Contribution Statement

Rui Zheng: Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Project administration. **Dashdorj Yamkhin:** Conceptualization, Visualization, Supervision. **Junxu Wei:** Validation. **Huqiang Liu:** Software, Validation, Data curation. **Zagarzusem Khurelbaatar:** Resources.

References

[1] Yadav, S., & Anand, R. (2025). Analysis of advancing paradigms of smart grid innovations, applications, challenges, future trends and strategic implementations. *Discover Applied Sciences*, *7*(12), 1380. https://doi.org/10.1007/s42452-025-07905-2

[2] Singh, S. B., Rizvi, M. A., Saxena, K., Gupta, R., Tripathi, A. N., & Dewangan, N. K. (2026). An adaptive, energy-efficient and secure routing protocol for zone-related mobile Ad-hoc networks using reinforcement learning. *Scientific Reports*, *16*(1), 3002. https://www.nature.com/articles/s41598-025-32918-7

[3] Priyadarshi, R., Kumar, R. R., Ranjan, R., & Kumar, P. V. (2025). AI-based routing algorithms improve energy efficiency, latency, and data reliability in wireless sensor networks. *Scientific Reports*, *15*(1), 22292. https://doi.org/10.1038/s41598-025-08677-w

[4] Kumar, S. P., Garg, S., Alabdulkreem, E., & Ben Miled, A. (2024). Advanced generative adversarial network for optimizing layout of wireless sensor networks. *Scientific Reports*, *14*(1), 32139. https://doi.org/10.1038/s41598-024-83957-5

[5] Pathak, V., Singh, K., Khan, T., Shariq, M., Chaudhry, S. A., & Das, A. K. (2024). A secure and lightweight trust evaluation model for enhancing decision-making in resource-constrained industrial WSNs. *Scientific Reports*, *14*(1), 28162. https://doi.org/10.1038/s41598-024-75414-0

[6] Abujassar, R. S. (2025). Intelligent IoT-driven optimization of large-scale healthcare networks: The INRwLF algorithm for adaptive efficiency. *Discover Computing*, *28*(1), 93. https://doi.org/10.1007/s10791-025-09601-6

[7] Vishwas, H. N., & Ramesh, T. K. (2025). Recent trends in localization, routing, and security for wireless sensor networks. *IEEE Access*, *13*, 55290–55312. https://doi.org/10.1109/ACCESS.2025.3555280

[8] Karopoulos, V., Kambourakis, G., Sanchez, G., & Sanchez, I. (2024). *Policy analysis and recommendations on internet standards deployment*. Publications Office of the European Union. https://doi.org/10.2760/1817977

[9] Suresh Kumar, V., Deepa Priya, B. S., Valaboju, S., Karthik, B., Rashid, S. Z., & Kavitha, P. (2025). Energy-efficient routing protocols in wireless sensor networks a comprehensive survey and future directions. *ITM Web of Conferences*, *76*, 03007. https://doi.org/10.1051/itmconf/20257603007

[10] Ahmad, S. M., Samie, M., & Honarvar Shakibaei Asli, B. (2025). Building trust in autonomous aerial systems: A review of hardware-rooted trust mechanisms. *Future Internet*, *17*(10), 466. https://doi.org/10.3390/fi17100466

[11] Rasouli, N., Klein, C., & Elmroth, E. (2026). Resource management for mission-critical applications in edge computing: Systematic review on recent research and open issues. *ACM Computing Surveys*, *58*(3), 71. https://doi.org/10.1145/3762181

[12] Shihab, M. A., Marhoon, H. A., Ahmed, S. R., Radhi, A. D., & Sekhar, R. (2024). Towards resilient machine learning models: Addressing adversarial attacks in wireless sensor network. *Journal of Robotics and Control*, *5*(5), 1599–1617. https://doi.org/10.18196/jrc.v5i5.23214

[13] Thakur, S., Sarkar, N. I., & Yongchareon, S. (2025). AI-driven energy-efficient routing in IoT-based wireless sensor networks: A comprehensive review. *Sensors*, *25*(24), 7408. https://doi.org/10.3390/s25247408

[14] Dey, R., Thakurta, P. K. G., & Kar, S. (2024). Energy aware routing in wireless sensor networks for agricultural monitoring: A deep Q-network based framework. *SN Computer Science*, *5*(7), 804. https://doi.org/10.1007/s42979-024-03183-8

[15] Boopathi, M., Parikh, S., Awasthi, A., Malviya, A., Nachappa, M. N., Mishra, A., . . . , & Narula, G. S. (2024). OntoDSO: An ontological-based dolphin swarm optimization (DSO) approach to perform energy efficient routing in Wireless Sensor Networks (WSNs). *International Journal of Information Technology*, *16*(3), 1551–1557. https://doi.org/10.1007/s41870-023-01698-6

[16] Abdulai, J.-D., Adu-Manu, K. S., Katsriku, F. A., & Engmann, F. (2023). A modified distance-based energy-aware (mDBEA) routing protocol in wireless sensor networks (WSNs). *Journal of Ambient Intelligence and Humanized Computing*, *14*(8), 10195–10217. https://doi.org/10.1007/s12652-021-03683-y

[17] Gorikapudi, S., & Kondaveeti, H. K. (2024). Energy aware cluster based routing algorithm for optimal routing and fault tolerance in wireless sensor networks. *Journal of Network and Systems Management*, *32*(2), 30. https://doi.org/10.1007/s10922-024-09806-y

[18] Bhukya, S., Srinivasarao, D., & Saheb, K. (2023). Environmental monitoring with wireless sensor networks for energy-aware routing and localization. *Journal of Sensors, IoT & Health Sciences*, *1*(1), 27–39. https://doi.org/10.69996/jsihs.2023003

[19] Muneeswari, G., Ahilan, A., Rajeshwari, R., Kannan, K., & John Clement Singh, C. (2023). Trust and energy-aware routing protocol for wireless sensor networks based on secure routing. *International Journal of Electrical and Computer Engineering Systems*, *14*(9), 1015–1022. https://doi.org/10.32985/ijeces.14.9.6

[20] Sharma, A., Babbar, H., Rani, S., Sah, D. K., Sehar, S., & Gianini, G. (2023). MHSEER: A meta-heuristic secure and energy-efficient routing protocol for wireless sensor network-based industrial IoT. *Energies*, *16*(10), 4198. https://doi.org/10.3390/en16104198

[21] Gururaj, H. L., Natarajan, R., Almujally, N. A., Flammini, F., Krishna, S., & Gupta, S. K. (2023). Collaborative energy-efficient routing protocol for sustainable communication in 5G/6G wireless sensor networks. *IEEE Open Journal of the Communications Society*, *4*, 2050–2061. https://doi.org/10.1109/OJCOMS.2023.3312155

[22] Mishra, P., Kandpal, M., Panigrahy, S., Rautaray, J., & Dash, R. K. (2025). Energy-aware intelligent routing framework for 6G-based wireless sensor networks. *Procedia Computer*

Science, *258*, 1285–1294. https://doi.org/10.1016/j.procs.2025.04.362

[23] Ramachandra, S., & Baskar, M. (2025). Real-time multi-level trust-based secure routing for improved QoS in WSN using blockchain. *Results in Engineering*, *26*, 104732. https://doi.org/10.1016/j.rineng.2025.104732

[24] Sreevidya, B., & Supriya, M. (2024). Trust-based routing: A novel approach for data security in WSN-based data-critical applications. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *15*(1), 27–41. https://doi.org/10.58346/JOWUA.2024.I1.003

[25] Kumar, V. (2025). Authentication and trust evaluation method for wireless sensor networks (WSNs). *International Journal of Emerging Knowledge Studies*, *4*(7), 1004–1015. https://doi.org/10.70333/ijeks-04-09-002

[26] Selvi, M., Santhosh Kumar, S. V. N., Thangaramya, K., & Abdul Gaffar, H. (2025). Energy efficient trust aware secure routing algorithm with attribute based encryption for wireless sensor networks. *Scientific Reports*, *15*(1), 19724. https://doi.org/10.1038/s41598-025-03558-8

[27] Goswami, P., Khan, T., Pathak, V., & Alabdultif, A. (2025). Machine learning based dynamic trust estimation framework for securing wireless sensor networks. *Scientific Reports*, *15*(1), 35821. https://doi.org/10.1038/s41598-025-19768-z

[28] Chen, X., Makki, K., Yen, K., & Pissinou, N. (2009). Sensor network security: A survey. *IEEE Communications Surveys & Tutorials*, *11*(2), 52–73. https://doi.org/10.1109/SURV.2009.090205

[29] Chaboki, B., Rafsanjani, M. K., & Fanian, F. (2025). Trust-based security management in WSN-assisted IoT systems: A comprehensive review. *International Journal of Software Science and Computational Intelligence*, *17*(1), 1–39. https://doi.org/10.4018/IJSSCI.395861

[30] Liu, C., Wang, Y., & Wang, Q. (2023). PARouting: Prediction-supported adaptive routing protocol for FANETs with deep reinforcement learning. *International Journal of Intelligent Networks*, *4*, 113–121. https://doi.org/10.1016/j.ijin.2023.05.002

[31] Krishnamoorthi, S., & Shyam, G. K. (2024). Design of recommendation systems using deep reinforcement learning: Recent advancements and applications. *Journal of Theoretical and Applied Information Technology*, *102*(7), 2908–2923.

[32] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., …, & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*, 529–533. https://doi.org/10.1038/nature14236

[33] Long, J., & Han, J. (2023). Reinforcement learning with function approximation: From linear to nonlinear. *Journal of Machine Learning*, *2*(3), 161–193. https://doi.org/10.4208/jml.230105

[34] Reijnen, R., Zhang, Y., Lau, H. C., & Bukhsh, Z. (2024). Online control of adaptive large neighborhood search using deep reinforcement learning. *Proceedings of the Thirty-Fourth International Conference on Automated Planning and Scheduling*, *34*(1), 475–483. https://doi.org/10.1609/icaps.v34i1.31507

[35] Li, D., Zhang, Z., Alizadeh, B., Zhang, Z., Duffield, N., Meyer, M. A., …, & Behzadan, A. H. (2024). A reinforcement learning-based routing algorithm for large street networks. *International Journal of Geographical Information Science*, *38*(2), 183–215. https://doi.org/10.1080/13658816.2023.2279975

[36] Yan, D., Guan, Q., Ou, B., Yan, B., Zhu, Z., & Cao, H. (2025). A deep reinforcement learning-based decision-making approach for routing problems. *Applied Sciences*, *15*(9), 4951. https://doi.org/10.3390/app15094951

[37] Pan, W., & Liu, S. Q. (2023). Deep reinforcement learning for the dynamic and uncertain vehicle routing problem. *Applied Intelligence*, *53*(1), 405–422. https://doi.org/10.1007/s10489-022-03456-w

[38] Machiwa, E. J., Masanja, V. G., Kisangiri, M. F., & Matiko, J. W. (2024). A comprehensive survey on linear programming and energy optimization methods for maximizing lifetime of wireless sensor network. *Discover Computing*, *27*(1), 21. https://doi.org/10.1007/s10791-024-09454-5

[39] Hassen, H., Meherzi, S., & Jemaa, Z. B. (2024). Improved exploration strategy for Q-learning based multipath routing in SDN networks. *Journal of Network and Systems Management*, *32*(2), 25. https://doi.org/10.1007/s10922-024-09804-0

[40] Wang, Z., Jang, W., Ruan, B., Wang, J., & Xiao, S. (2026). Developing and integrating trust modeling into multi-objective reinforcement learning for intelligent agricultural management. *Smart Agricultural Technology*, *14*, 102145. https://doi.org/10.1016/j.atech.2026.102145

[41] Moudoud, H., & Cherkaoui, S. (2023). Empowering security and trust in 5G and beyond: A deep reinforcement learning approach. *IEEE Open Journal of the Communications Society*, *4*, 2410–2420. https://doi.org/10.1109/OJCOMS.2023.3313352

[42] Kim, D., Lee, K., & Oh, S. (2023). Trust region-based safe distributional reinforcement learning for multiple constraints. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 19908–19939.

How to Cite: Zheng, R., Yamkhin, D., Wei, J., Liu, H., & Khurelbaatar, Z. (2026). Trust-Gated Reinforcement Learning-Based Secure Routing for Wireless Sensor Networks. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE62029141