



Adaptive Multi-Objective Reinforcement Learning-Based Energy-Efficient Routing in Wireless Sensor Networks

M R GANGAARAAM¹, Sittalatchoumy R²

^{1,2} University College of Engineering, Anna University, CEG Campus, Chennai, India.

Article Info

Article History:

Published: 25 May 2026

Publication Issue:

Volume 3, Issue 5
May-2026

Page Number:

388-395

Corresponding Author:

Dr. Surjadeep Dutta

Abstract:

Abstract—Wireless Sensor Networks (WSNs) are widely used in monitoring and IoT applications, where efficient routing is critical due to limited node energy and dynamic network conditions. Traditional routing techniques such as greedy routing suffer from issues like local minima, lack of energy awareness, and uneven load distribution. To address these challenges, this paper proposes an adaptive multi-phase routing framework integrating cost-based optimization, multi-objective decision-making, and reinforcement learning (RL). Initially, greedy routing is implemented as a baseline to highlight fundamental limitations. A cost-based routing approach is introduced to improve link quality by incorporating distance and communication constraints. Further enhancement is achieved through a multi-objective routing model that considers residual energy and node load, ensuring balanced and energy-efficient data transmission. Finally, a Q-learning-based routing mechanism enables nodes to learn optimal routing paths dynamically. Simulation results demonstrate that the proposed RL-based approach significantly improves packet delivery ratio, enhances network lifetime, and achieves better adaptability compared to conventional routing methods.

Keywords: Wireless Sensor Networks (WSN), Reinforcement Learning, Q-Learning, Energy-Efficient Routing, Multi-Objective Optimization, Packet Delivery Ratio (PDR), Network Lifetime, Load Balancing

1. INTRODUCTION

Wireless Sensor Networks (WSNs) have emerged as a fundamental technology in diverse real-world applications including environmental monitoring, precision agriculture, industrial automation, and IoT systems. A WSN consists of spatially distributed autonomous sensor nodes that sense, process, and transmit data to a central sink. These nodes are battery-powered and often deployed in remote or inaccessible locations, making energy conservation critical.

Routing is one of the most energy-intensive operations in WSNs. Inefficient routing leads to rapid energy depletion of certain nodes, causing network partitioning and degraded lifetime. Traditional greedy geographic routing forwards packets based solely on proximity to the sink, neglecting residual energy and traffic load, resulting in coverage holes and degraded performance.

To address these limitations, this paper proposes a progressive multi-phase routing framework that evolves from simple heuristic-based routing to an adaptive reinforcement learning (RL)-based approach. The framework consists of four phases: (1) greedy routing as a baseline, (2) cost-based routing, (3) multi-objective routing, and (4) Q-learning-based routing for fully adaptive path selection.

2. LITERATURE REVIEW

WSNs have been extensively studied with a focus on energy-efficient and reliable routing. Early approaches such as Greedy Perimeter Stateless Routing (GPSR) [1] utilize node location information to forward packets toward the sink by selecting the nearest neighbor. Although computationally efficient, greedy routing suffers from local minimum problems and lacks energy and traffic awareness.

Hierarchical protocols such as LEACH [2] introduced clustering mechanisms to distribute energy consumption among nodes. While these methods improve network lifetime, they introduce overhead due to cluster formation and maintenance. Energy-aware and cost-based routing strategies use weighted cost functions incorporating distance, residual energy, and transmission cost [3]. These improve efficiency but rely on static parameter tuning.

Multi-objective optimization techniques balance conflicting parameters such as energy, delay, and load distribution [4]. Reinforcement learning (RL) has recently emerged as a promising approach; Q-learning-based studies [5][7] demonstrate that nodes can learn optimal routing strategies dynamically without prior knowledge. However, most existing works address either optimization or learning independently. In contrast, this work proposes a hybrid framework combining all four routing paradigms.

3. PROPOSED METHODOLOGY

This work proposes a multi-phase routing framework for WSNs that progressively enhances performance by integrating heuristic, optimization-based, and learning-based approaches. The methodology consists of four main phases.

3.1 Network Model

A set of N sensor nodes is randomly deployed in a two-dimensional area of size $A \times A$. A sink node is placed at the center. Each node is initialized with equal energy E_0 and communicates within a transmission range R . The Euclidean distance between nodes i and j is:

$$D_{ij} = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]}$$

3.2 Phase 1: Greedy Routing

Each node forwards data to the neighbor closest to the sink. The next hop is selected as:

$$j^* = \arg \min_{j \in N_i} D_{js}$$

Although efficient, this method suffers from local minima and ignores energy and load.

3.3 Phase 2: Cost-Based Routing

A cost function considering sink-progress and link quality is introduced:

$$\text{Cost} = \alpha \cdot D_{s_i^n} + \beta \cdot (1/d)$$

where $D_{s_i^n}$ is the neighbor's distance to sink, d is the inter-node distance, and α, β are weighting factors. This improves link reliability and routing flexibility.

3.4 Phase 3: Multi-Objective Routing

Multiple parameters are incorporated into the routing decision:

$$\text{Cost} = w_1 \cdot D_{s_i^n} + w_2 \cdot (1/E) + w_3 \cdot \text{Load}$$

where E is residual energy and Load represents node traffic. An energy threshold penalty is applied:

$$\text{Cost} = \text{Cost} \times \text{Penalty} \text{ (if } E < \text{threshold)}$$

This ensures balanced energy usage and reduces congestion.

3.5 Phase 4: RL-Based Routing

Q-learning enables adaptive routing. Each node maintains a Q-table and learns optimal next-hop decisions. The update rule is:

$$Q(s,a) = Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$$

where s is the current node, a is the selected next hop, $r = -D_{s_i^n}$ is the reward, and γ is the discount factor. An epsilon-greedy strategy balances exploration and exploitation.

3.6 Performance Metrics

- Packet Delivery Ratio (PDR) = Delivered Packets / Total Sent
- Network Lifetime: Alive nodes over simulation rounds.
- Learning Curve: Reward convergence across training episodes.

4. NOVELTY AND KEY CONTRIBUTIONS

The uniqueness of this work lies in combining conventional routing, optimization-based routing, and reinforcement learning into a single progressive framework. The key novel contributions are:

1. **Progressive Multi-Phase Routing Architecture:** Most existing works apply only one routing method. This work proposes an evolutionary framework that systematically advances from greedy to RL-based routing, enabling clear performance benchmarking at each stage.
2. **Multi-Objective + RL Integration:** The proposed cost function jointly considers distance, residual energy, and node load. Combining this with RL-based adaptive learning produces a routing system that is both energy-aware and dynamically optimizing.
3. **Low-Energy Penalty Mechanism:** A threshold-based penalty discourages routing through nodes with critically low energy, preserving node lifetime and reducing premature node death compared to conventional methods.
4. **Comparative Benchmarking Under Identical Conditions:** All four routing phases are evaluated under the same simulation environment, providing fair and clear quantitative evidence of improvement at each stage.
5. **Lightweight MATLAB Implementation:** The proposed framework is computationally practical, making it suitable for academic simulation and future adaptation to embedded sensor platforms.

The integration of multi-objective optimization with reinforcement learning provides a scalable and adaptive routing strategy not previously addressed in a unified phased framework.

5. SIMULATION SETUP

The proposed routing framework was implemented and evaluated in MATLAB. A WSN of randomly deployed sensor nodes was simulated in a two-dimensional sensing area with the sink at the center. Each node was initialized with equal energy, and multi-hop communication was allowed within a predefined transmission range. The RL module was trained using Q-learning with an epsilon-greedy exploration policy.

The simulation compared Greedy, Cost-Based, Multi-Objective, and RL-Based Routing under identical network conditions.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Simulation Tool	MATLAB
Network Type	WSN
Number of Nodes (N)	50
Deployment Area	100 m × 100 m
Sink Position	(50, 50)
Transmission Range (R)	35 m

Initial Node Energy	1 Joule
Simulation Rounds	100
RL Training Episodes	200
Learning Rate (α^{AL})	0.5
Discount Factor (γ)	0.8
Exploration Rate (ϵ)	0.2
Greedy Weight (α)	0.7
Link Cost Weight (β)	0.3
Multi-Obj. Weights	0.4, 0.4, 0.2

6. RESULTS AND DISCUSSION

6.1 Phase 1: Network Deployment

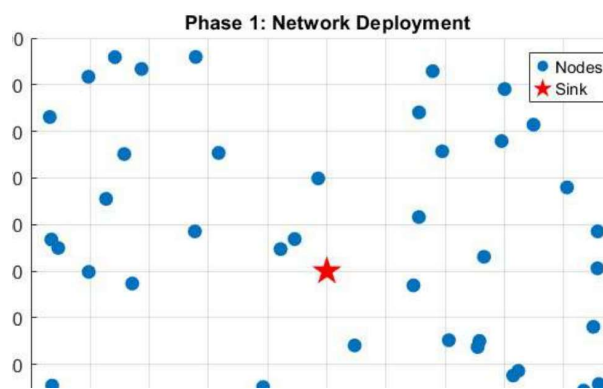


Fig. 1. Phase 1: Network Deployment – 50 nodes randomly deployed, sink at center (50,50).

Fig. 1 shows 50 sensor nodes uniformly distributed in a 100×100 m area with the sink at center. This topology creates a realistic multi-hop WSN environment for routing comparison.

6.2 Phase 2: Greedy Routing

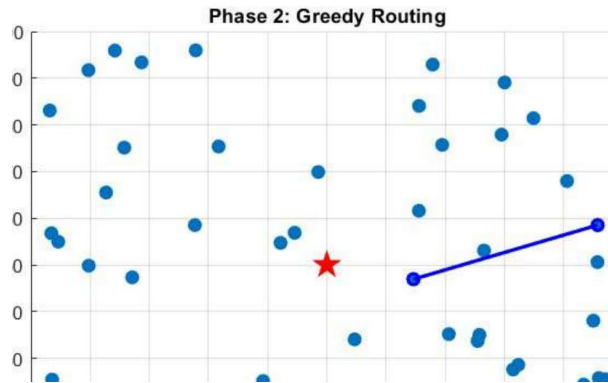


Fig. 2. Phase 2: Greedy Routing – packet forwarded to nearest neighbor of sink.

Fig. 2 shows greedy routing forwarding packets to the nearest neighbor of the sink (blue path). While producing short paths, repeated use of the same nodes causes rapid energy drain. Local minima can trap packets in sub-optimal positions.

6.3 Phase 2B: Cost-Based Routing



Fig. 3. Phase 2B: Cost-Based Routing – path selected using distance and link quality cost.

Fig. 3 illustrates cost-based routing (green path) using the distance-and-link-quality cost function. Paths are more stable, hop selection is improved, and communication reliability is enhanced compared to greedy routing.

6.4 Phase 3: Multi-Objective Routing

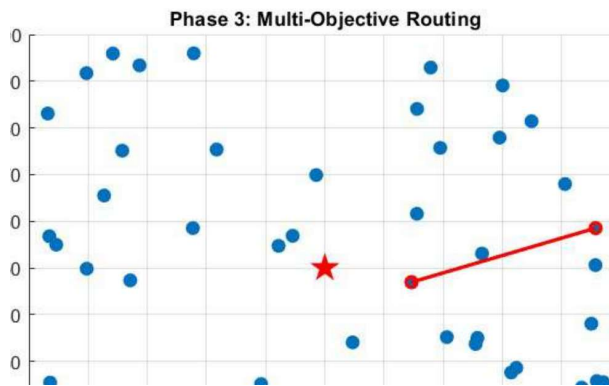


Fig. 4. Phase 3: Multi-Objective Routing – traffic distributed using energy and load awareness.

Fig. 4 shows multi-objective routing (red path) distributing traffic among nodes by considering residual energy and load. Low-energy nodes are avoided, congestion is reduced, and network lifetime is significantly extended.

6.5 Phase 4: RL-Based Routing

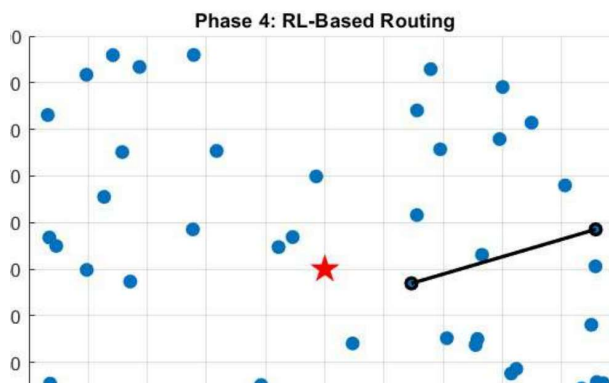


Fig. 5. Phase 4: RL-Based Routing – Q-table learned optimal next-hop selection (black path).

Fig. 5 shows RL-based routing using Q-table learned paths (black path). The agent intelligently avoids weak and congested nodes, providing the most efficient and adaptive routing performance across all phases.

6.6 RL Learning Curve

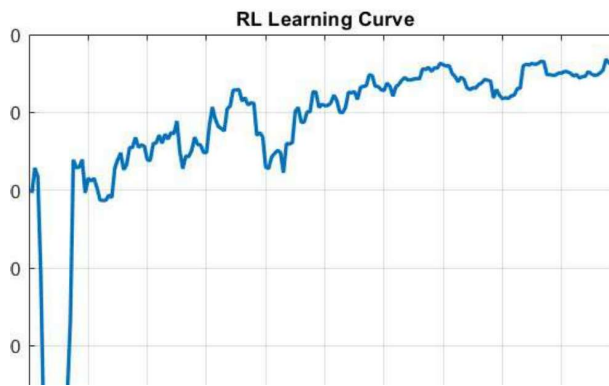


Fig. 6. RL Learning Curve – Cumulative reward improves from -500 at episode 1 to ~ -50 by episode 200.

Fig. 6 shows the RL learning curve across 200 training episodes. Initial rewards start near -500 due to exploratory random actions. Rewards improve sharply within the first 20 episodes, then converge gradually toward approximately -50 by episode 200, validating the effectiveness of Q-learning in discovering near-optimal routing policies.

6.7 Comparative Analysis

TABLE II. COMPARATIVE PERFORMANCE

Method	PDR	Lifetime	Intel.	Adapt.
Greedy	Low	Low	Basic	No
Cost-Based	Med.	Med.	Mod.	No
Multi-Obj.	High	High	Good	Part.
RL-Based	Highest	Highest	Excel.	Yes

Table II confirms that each successive routing phase outperforms the previous. The RL-based approach achieves the best overall performance by dynamically learning optimal policies without manual parameter tuning.

7. CONCLUSION

An adaptive multi-phase routing framework for WSNs was proposed to improve routing efficiency, energy conservation, and network lifetime. Starting from greedy routing as a baseline, the framework progressively introduced cost-based optimization, multi-objective energy and load awareness, and Q-learning-based adaptive routing. Simulation results demonstrated that the RL-based approach outperformed conventional

techniques in PDR, network lifetime, and routing adaptability. The learning curve confirmed stable convergence and effective route optimization. The integration of multi-objective optimization with reinforcement learning provides a scalable routing strategy for future intelligent sensor networks.

8. FUTURE WORK

- Deep Reinforcement Learning (DQN/DRL) for larger-scale networks.
- Real-time simulation using NS-3 or OMNeT++.
- Integration with IoT smart city and healthcare applications.
- Security-aware intelligent routing mechanisms.
- Mobile sink-based routing optimization.

References

- [1] B. Karp and H. T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," Proc. ACM MobiCom, 2000, pp. 243–254.
- [2] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proc. IEEE HICSS, 2000.
- [3] I. F. Akyildiz et al., "A Survey on Wireless Sensor Networks," IEEE Communications Magazine, vol. 40, no. 8, pp. 102–114, 2002.
- [4] C. J. C. H. Watkins, "Learning from Delayed Rewards," Ph.D. dissertation, Univ. Cambridge, 1989.
- [5] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. MIT Press, 2018.
- [6] M. Gupta and R. Kumar, "Energy Efficient Routing Techniques for WSNs," IEEE Access, 2021.
- [7] J. Wang and Y. Liu, "Adaptive Reinforcement Learning Routing for WSNs," IEEE Internet of Things Journal, 2023.