CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Reinforcement Learning Based Resource Utilization for Wireless Sensor Networks

by

Abdul khaliq

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Engineering
Department of Electrical Engineering

2026

CERTIFICATE OF APPROVAL

Reinforcement Learning Based Resource Utilization for Wireless Sensor Networks

by

Abdul khaliq

MEE231002

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Muhammad Sajid	MUST, AJK
(b)	Internal Examiner	Dr. Noor M. Khan	CUST, Islamabad

Dr. Imtiaz Ahmad Taj
Thesis Supervisor
June, 2026

Dr. Noor Muhammad Khan
Head
Dept. of Electrical Engineering
June, 2026

Dr. Imtiaz Ahmad Taj
Dean
Faculty of Engineering
June, 2026

Author's Declaration

I, **Abdul khaliq** hereby state that my MS thesis titled "**Reinforcement Learning Based Resource Utilization for Wireless Sensor Networks**" is my work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time, if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Abdul khaliq)

Registration No:MEE231002

Plagiarism Undertaking

I hereby declare that research work presented in this thesis titled "**Reinforcement Learning Based Resource Utilization for Wireless Sensor Networks**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged, and the complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I, as the author of the above-titled thesis, declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Abdul khaliq)

Registration No:MEE231002

Acknowledgement

In the name of Allah, the Most Gracious, the Most Merciful. All praise and thanks are due to Allah (SWT), the Creator and Sustainer of all that exists, who granted me the strength, knowledge, and opportunity to undertake this research. May peace and blessings be upon our beloved Prophet Muhammad (PBUH), his family, and his companions, who is a mercy to mankind and a guide for all humanity.

I am deeply grateful for the opportunity to complete this thesis on the performance evaluation of RL-EBRP and EH-RL-AEBRP protocols in wireless sensor networks. My heartfelt gratitude goes to those who have supported me throughout this journey.

I deeply appreciate my thesis advisor, Dr. Imtiaz A Taj, for their invaluable guidance, scholarly insights, and continuous encouragement. Their expertise and constructive feedback were pivotal in shaping the direction and quality of this research.

My deepest appreciation goes to my family and friends for their unwavering support, patience, and prayers throughout this endeavor. Their encouragement motivated me to persevere through challenges and complete this research.

This work would not have been possible without the blessings of Allah (SWT) and the support of everyone mentioned above. I remain eternally grateful for their contributions.

Abdul khaliq

Abstract

This dissertation addresses the fundamental challenge of energy constraints in Wireless Sensor Networks (WSNs) that limits their operational lifetime in Internet of Things (IoT) applications. The research develops two innovative routing protocols that leverage machine learning to optimize energy utilization: (1) RL-AEBRP (Reinforcement Learning-based Adaptive Energy Balancing Routing Protocol) for conventional battery-powered WSNs, and (2) EH-RL-AEBRP (Energy-Harvesting RL-AEBRP) for networks with energy harvesting capabilities.

The proposed framework introduces three key innovations: (i) a Q-learning based adaptive routing mechanism that dynamically adjusts to node energy levels and network conditions, (ii) an integrated energy harvesting prediction model, and (iii) a mathematical formulation for energy consumption optimization. Extensive simulations in MATLAB demonstrate significant improvements over state-of-the-art protocols (RL-EBRP) across multiple metrics:

The protocols particularly excel in dynamic IoT environments, maintaining stable performance under varying network densities and traffic loads. Theoretical analysis proves the convergence of the Q-learning algorithm, while simulation results validate the framework's effectiveness in realistic scenarios. This work provides both theoretical foundations and practical solutions for sustainable WSN deployments, with direct applications in smart city infrastructure, precision agriculture, and remote health monitoring systems. The proposed energy-aware routing paradigm opens new research directions for machine learning in resource-constrained networks.

Contents

List of Figures

List of Tables

Abbreviations

BFS	Breadth first search
BS	Base station
EH	Energy harvesting
EHRLAEBRP	Energy harvesting reinforcement learning adaptive energy balancing routing protocol
EHWSN	Energy harvesting wireless sensor network
FND	First node death
HWSN	Hybrid wireless sensor network
HND	Half Nodes death
LND	Last Node death
ML	Machine learning
MWSN	Mobile wireless sensor network
RL	Reinforcement learning
SN	Sensor Node
WSn	Wireless sensor node
WSN	Wireless sensor network
WBAN	Wireless body area network

Symbols

E_{tx}	Energy consumed for transmission	Joules
E_{rx}	Energy consumed for reception	Joules
E_{res}	Residual energy of a sensor node	Joules
E_h	Harvested energy	Joules
E_{init}	Initial energy of a sensor node	Joules
E_{total}	Total energy consumed	Joules
$Q(i, j)$	Q-value for routing from node i to j	Unitless
r_{ij}	Reward for selecting next hop j from i	Unitless
α	Learning rate in Q-learning	Dimensionless
γ	Discount factor in Q-learning	Dimensionless
$P_h(t, i)$	Probability of energy harvesting for node i at time t	Dimensionless
w_1, w_2, w_3	Weights for reward function components	Dimensionless
d_{ij}	Distance between nodes i and j	Meters
N	Number of sensor nodes	Unitless
S_i	State of node i	Unitless
A_i	Action of node i (next hop selection)	Unitless
T	Time period or simulation duration	Seconds
P_{tx}	Transmission power	Watts
P_{rx}	Reception power	Watts
R_{pdr}	Packet delivery ratio	Dimensionless
L_{net}	Network lifetime	Seconds
A_{det}	Anomaly detection score	Unitless

Chapter 1

Introduction

1.1 Background

WSN (Wireless Sensor Network) form the backbone of numerous upcoming applications, i.e. smart city, smart homes, healthcare monitoring, environmental monitoring, disaster management, etc. These networks contain a multitude of sensor nodes distributed over an area of interest to wirelessly gather and disseminate a wide array of information. Such a deployment, as illustrated in Figure 1.1 [1, 2]. WSNs are especially prominent in applications where human involvement is difficult or impractical, for instance, monitoring wilderness or dangerous factory settings .

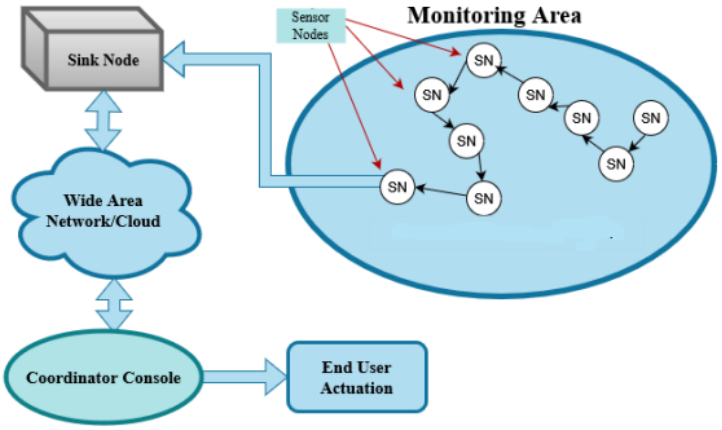


FIGURE 1.1: Sensor node deployment in an area of interest.

A typical sensor node (SN) is composed of sensing, processing, communication and power supply modules. These nodes transmit the collected data to a Base Station (BS) or a sink node, either directly or through multi-hop communication [3]. In most real-world applications, sensor nodes (SN) rely on a restricted energy supply, often in the form of non rechargeable batteries [4, 5]. Consequently, they operate under strict energy constraints, making efficient energy management essential. To ensure the effective execution of fundamental tasks such as sensing, processing, and communication, energy consumption must be carefully optimized [4–6]. Structure of wireless sensor node (WSn) is give in Figure.1.2.

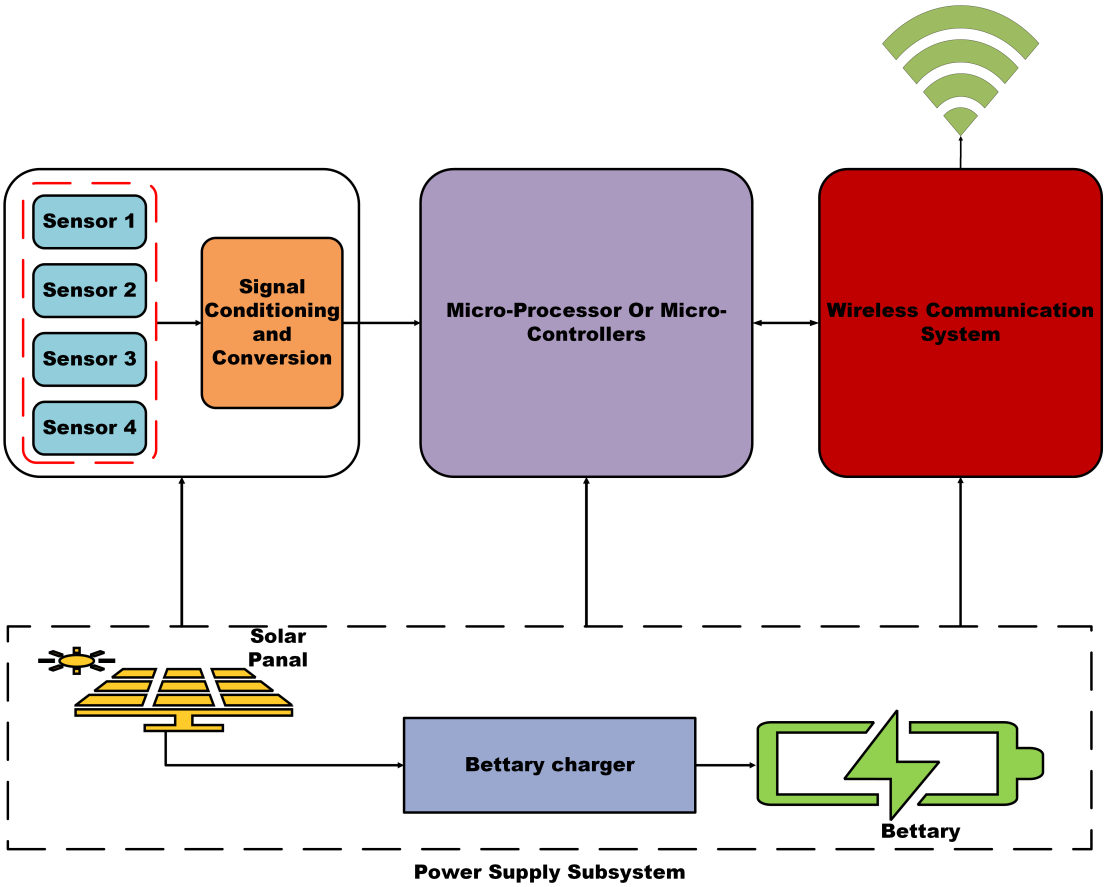


FIGURE 1.2: Wireless Sensor Node Structure.

While WSNs offer immense utility, the limited energy sources of sensor nodes (SN) are a major bottleneck. In many deployments, replacing batteries is impractical due to cost, accessibility, or environmental factors. To address this issue, energy harvesting (EH) techniques have emerged as a promising solution. EH enable sensor nodes (SN) to produce energy from ambient energy sources such as sun light, Heat, or radio frequency (RF) signals. By integrating energy harvesting

mechanisms with efficient data routing protocols, the longevity and performance of WSN systems can be significantly improved. So energy efficient routing protocols are designed to prolong the operational lifetime of the network [7].

Energy harvesting technologies have gained traction as a means to mitigate energy constraints in WSNs. For instance, solar power is ideal for outdoor deployments with ample sunlight, while thermal energy converts heat differences into usable power. Vibrational energy, on the other hand, can be harvested from mechanical vibrations in industrial environments. Although EH extends the operational lifetime of sensor nodes, incorporating harvested energy into routing protocols remains a challenge. An efficient routing protocol must dynamically adapt to fluctuating energy availability while balancing the communication load across nodes [8].

Wireless Sensor Networks (WSNs) are used over many domains to facilitate real time monitoring and decision making. This includes healthcare monitoring, precision agriculture, industrial automation, environmental monitoring, and smart cities. In the domain of healthcare supervision, wearable sensors play a pivotal role in tracking patient vitals, allowing patients to be monitored from a distance while also facilitating the continuous monitoring of heart conditions leading to early detection of such cases [9]. Precision agriculture benefits from WSNs by monitoring soil moisture, temperature, and crop health, optimizing farming efficiency, while WSN-based irrigation systems enhance water usage [10]. Industrial automation relies on WSNs for monitoring machinery health, predicting failures to minimize downtime, and providing real-time alerts for hazardous conditions to improve worker safety [11]. Environmental monitoring applications include detecting and tracking natural disasters such as floods and forest fires, as well as monitoring air and water quality in urban areas. In smart cities, WSNs enable traffic monitoring, smart parking systems, and efficient energy management [12]. These applications underline the need for robust, energy efficient WSN protocols that ensure reliable data transmission over extended periods.

Machine Learning (ML), Deep Learning (DL), and Deep Reinforcement Learning (DRL) enhance Wireless Sensor Network (WSN) performance by optimizing energy efficiency, routing, and data management.

ML and DRL, subsets of Artificial Intelligence (AI), enable data-driven decision-making. ML includes supervised, unsupervised, and DRL methods. DRL, using Artificial Neural Networks (ANNs), extracts patterns from data, benefiting applications like image recognition and predictive analytics.

For WSNs, DRL trains networks using historical data to optimize channel allocation and routing. However, model training demands high computational power, limiting feasibility in some cases.

WSNs integrate modern technologies for efficient sensing and computation. ML and DRL enhance WSNs in several ways:

i. Energy-efficient routing: ML predicts energy consumption, optimizing data transmission paths.

ii. Object detection & tracking: DRL aids in monitoring vehicles and intruders for security.

iii. Anomaly detection: ML identifies sensor data anomalies, enabling early interventions.

iv. Predictive maintenance: ML forecasts node failures, reducing downtime.

v. Resource allocation: ML assigns tasks based on node capabilities, maximizing efficiency.

These techniques improve WSN reliability and efficiency. Future research should enhance accuracy while maintaining coverage, connectivity, and computational efficiency [13].

WSN design depends on coverage, connectivity, and power allocation. Transmission power affects coverage, but reducing it extends node lifespan [14]. Probabilistic modeling optimizes energy use and coverage, balancing connectivity and power savings.

Routing is critical for reliable data transfer in WSNs. Challenges include:

i. Addressing: Large sensor deployments make global addressing impractical.

ii. Data streaming: Continuous data transmission conflicts with standard networks.

iii. Redundancy: Multiple sensors generate similar data, increasing energy and bandwidth use [15].

1.2 Problem Statement

Routing in WSNs involves determining the optimal path for transmitting data from sensor nodes to the sink. Energy-efficient routing is critical because communication consumes a significant portion of a node's energy. In multi-hop routing, intermediate nodes relay packets on behalf of others, reducing the transmission distance for individual nodes. However, this approach introduces several challenges:

1.2.1 Dynamic Adaptability

WSNs operate in dynamic environments where node failures, energy depletion, or environmental interference frequently occur. Most existing protocols are static, failing to adapt to these changes effectively. This leads to dropped packets, increased energy consumption, and reduced network reliability [16].

1.2.2 Energy-Efficient Routing

The energy consumed for transmitting a packet over a distance d can be modeled as follows:

$$E_{\text{tx}} = k \cdot E_{\text{elec}} + k \cdot \epsilon \cdot d^n \tag{1.1}$$

where k is the packet size in bits, E_{elec} represents the energy consumed per bit for processing, ϵ is the amplifier energy constant, d is the transmission distance, and n is the path-loss exponent, typically $n = 2$ for free space but varying between 3 and 4 in urban environments. This model highlights the exponential impact of distance on energy consumption, emphasizing the importance of multi-hop communication

in energy-constrained wireless sensor networks (WSNs). Instead of direct long-range transmissions, multi-hop routing reduces the required transmission power per hop, thereby prolonging network lifetime. Since energy harvesting in WSNs is limited and unpredictable, reducing transmission distance through multi-hop communication helps optimize energy consumption and enhances overall network performance [16]. To minimizing the transmitting energy E_{tx} a balance is required between direct and multi-hop communication intermediate nodes in multi-hop paths deplete their energy quickly and die out that create break in path called hot-spot that fragment the network and reduce the network lifetime [17].

1.2.3 Integration of Energy Harvesting

While energy harvesting offers additional energy resources, existing protocols often fail to utilize harvested energy effectively. Routing decisions should prioritize nodes with surplus energy while avoiding nodes nearing depletion [18].

1.2.4 Load Balancing

Without proper load balancing, intermediate nodes in heavily used paths deplete their energy faster. Uneven energy consumption across the network reduces the overall network lifetime, particularly in high-traffic scenarios [19].

1.3 Research Objectives

This research aims to develop an Energy-Efficient Multi-Hop Routing Protocol using RL for energy-harvesting WSNs. The key objectives are:

i. To design a reinforcement learning-based routing protocol that dynamically adapts to network conditions.

ii. To integrate energy harvesting into routing decisions, maximizing the utilization of harvested energy.

iii. To evaluate the protocol against existing solutions in terms of network lifetime and energy consumption.

1.4 Contributions

The primary contributions of this research are:

i. A novel reinforcement learning-based routing protocol that optimizes energy efficiency in multi-hop WSNs.

ii. Integration of energy harvesting models to dynamically adapt to available energy resources.

iii. Comprehensive simulation and performance evaluation demonstrating significant improvements over state-of-the-art protocols.

1.5 Thesis Organization

This thesis investigates energy-efficient routing in Wireless Sensor Networks with energy harvesting.

i. Chapter 1: Introduction outlines the energy challenges in WSNs, defines the research problem, and presents EH-RL-AEBRP's objectives.

ii. Chapter 2: Background Study covers WSNs, IoT, Q-learning, and energy harvesting, establishing foundational concepts.

iii. Chapter 3: Literature Review evaluates prior routing protocols (e.g., RL-EBRP), highlighting gaps addressed by EH-RL-AEBRP.

iv. Chapter 4: Proposed Model describes RL-AEBRP and EH-RL-AEBRP Q-learning framework, energy models, and MATLAB simulation setup.

v. Chapter 5: Results and Performance Evaluation presents simulation results (network lifetime, packet delivery ratio) compared to benchmarks and concludes with RL-AEBRP and EH-RL-AEBRP contributions and limitations.

Chapter 2

Background Study

2.1 Overview

This chapter gives an overview of the fundamental problems of Wireless Sensor Networks (WSNs), specifically energy harvesting WSNs, and applying Reinforcement Learning (RL) to this field. The chapter first introduces the fundamental features of WSNs, with special focus on the potential and issues of energy harvesting methods. The chapter then provides a concise overview of Reinforcement Learning (RL), detailing its basic concepts, categories, and well-known algorithms applicable to adaptive and dynamic systems.

Finally, the chapter summarizes the pros and cons of leveraging RL application in energy harvesting WSNs, as well as its role in extending a life-time of the network and maximizing its performance in energy-limited scenario.

2.2 Wireless Sensor Networks

Wireless sensor networks (WSNs) are one of the most promising and evolving technologies that are used in various real-time applications in diverse domains such as military, automation, vehicle tracking, environmental, wildlife tracking, agriculture, etc [7].WSNs have several merits including; low cost, small in size, easy of deployment, self-organization, and low maintenance cost [20]. They include

small sensor nodes, deployed randomly over a specific domain for measurement and data collection. The essential hardware of a sensor node consists of a sensing unit, transceiver, processing unit, battery, and memory unit.

WSNs can include Mobile Wireless Sensor Networks (MWSNs), which have broad applications in various fields. In a mobile wireless sensor network, the sensor nodes can move within the network. The rapid growth of mobile technology and the Internet has made MWSNs a popular research area in WSNs.

2.2.1 Potential Benefits of WSNs

One of the primary benefits of WSNs is their flexibility and adaptability. Sensor nodes can be deployed in remote or hard-to-reach areas, allowing for the monitoring of locations that are otherwise inaccessible [21] this is due to the self-organizing and ad-hoc nature of these networks. Moreover, WSNs are more economical than wired sensor systems, and thus more affordable for a wide range of applications [22].

Another major benefit of WSNs is that they can provide continuous data and consistent monitoring. Sensor nodes can be programmed to gather data periodically and send it to a central processing unit for immediate analysis and action. In addition, growing advancements in energy-harvesting technologies enable sensor nodes to extract energy from their environment, be it from solar, thermal or kinetic sources, thus, providing an extra layer of sustainability and longevity to WSN deployments at the cost of frequent battery replacements during the WSN lifetime [23].

Besides this, wireless sensor networks are also utilized in many areas, equally healthcare, agronomy, smart cities, and the industry.

2.2.2 Possible Applications of WSNs

The versatility of WSNs has led to their adoption in a wide range of applications, each with its unique set of requirements and challenges some application are as follows.

2.2.2.1 Military Applications

WSNs have found applications in the military domain, where they can be used for battlefield surveillance, reconnaissance, and targeting systems. Sensor nodes can be deployed in remote locations to monitor troop movements, equipment, and vehicle activity, providing valuable intelligence to military personnel [24, 25].

2.2.2.2 Environmental Monitoring

WSNs have been successfully deployed for environmental monitoring applications, such as forest fire detection, water quality monitoring, and wildlife tracking. In such cases, sensor nodes are deployed in the environment to monitor specific parameters, including temperature, humidity, soil moisture and pH levels. WSNs can also be embedded with IoT technology to facilitate smart city functionality by supervising environmental conditions [26]. as ensures interfacing with other urban instrumentation and expands the capacities for real-time analysis of air pollution, noise, and traffic conditions.One specific example of environmental monitoring using WSNs is deploying sensor nodes on agricultural fields to monitor soil moisture and humidity levels. Farmers can optimize irrigation schedules, reduce water consumption, and improve crop yields by collecting and analyzing this kind of data.

2.2.2.3 Healthcare and Assisted Living

Adoption of WSNs has also extended to the medical and healthcare sector.For example, sensor nodes can monitor patient indicators, including heart rate and blood pressure, blood oxygen levels, enabling early identification of health problems and the capability to conduct remote patient monitoring. Wireless sensor networks (WSNs) can also be used to track the location and movement of patients, seniors, or healthcare workers, enhancing safety and efficiency in healthcare settings [27].WSNs are also used in the field of assisted living, where WSNs continuously monitor the activity of the elderly or disabled individuals and notify caregivers about any behavioral changes or potential emergencies. This can allow these individuals to remain independent at home for a longer period while providing even greater safety and support [28].

2.2.2.4 Transportation and Logistics

Data for transportation and logistics Vehicle tracking, traffic, and supply chain are some of the applications where WSNs have been used.For example, sensor nodes can be deployed in vehicles, or alongside transport routes to gather information on traffic flow, road conditions, and vehicle telemetry. By computing this information, administrators will be able to route delivery trucks more efficiently, help keep congestion at bay, and ensure that transportation and logistics operations run smoothly [29].One example of WSN application in this domain has sensor nodes that monitor the condition of heavy machinery such as construction or mining equipment. This ability to monitor variables such as vibration, temperature, and oil quality in real time empowers operators to introduce predictive maintenance, mitigate the risk of sudden breakdowns, increase operational efficiency, and reduce costs [30].

2.2.3 Wireless Sensor Networks Classifications

Due to the diversity of application used for wireless sensor networks, there are different types of wireless sensor networks classified based on their architecture.

2.2.3.1 Structured WSNs

They are characterized by a fixed placement of sensor nodes, usually in a regular or clustered structure as depicted on Figure 2.1. This creates more effective routing of data and improved management and maintenance of the network [31].

2.2.3.2 Unstructured WSNs

which are depicted in Figure 2.2, unstructured networks have a more random node distribution and are usually deployed randomly [32].

This can serve as an advantage in case the configuration has to be done in the deployment area and accessibility is limited like in the mountains or extreme weather [25].

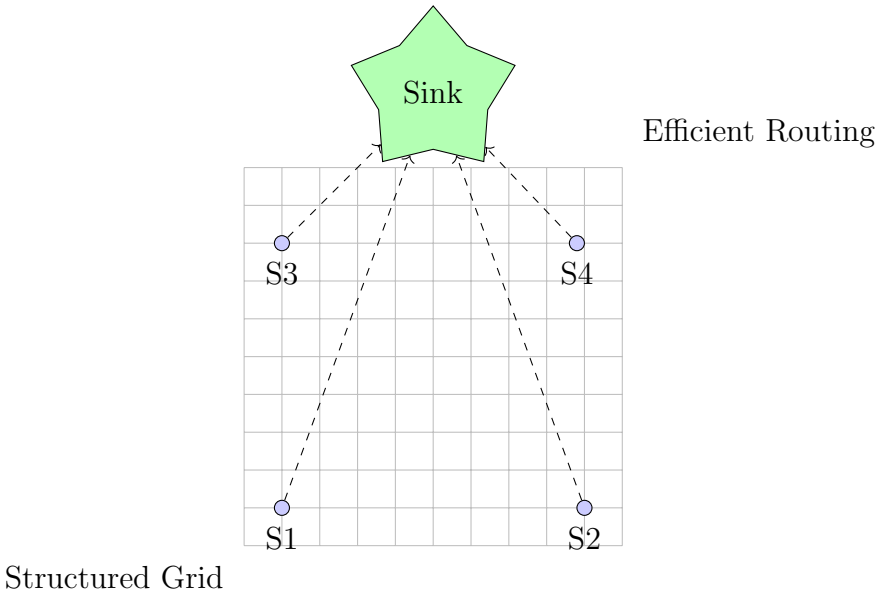


FIGURE 2.1: Structured WSN Layout

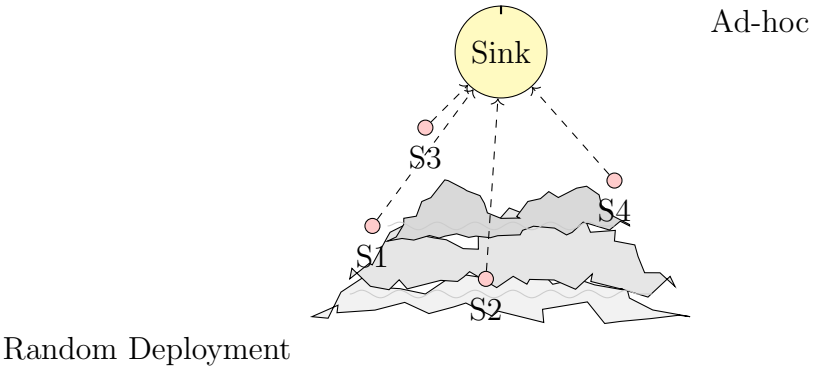


FIGURE 2.2: Unstructured Wireless Sensor Network Deployment

2.2.3.3 Static Wireless Sensor Network

A key class of WSN is the static wireless sensor network in which sensor nodes are spread out over a predetermined location.

These networks are commonly applied in environmental monitoring, military surveillance, and industrial sensing [33, 34]. This means the routing protocols in static WSNs must ensure the efficiency of transmitting data from the sensor node to the sink to gather information and process it further [35].

Routing protocols at this level of network must take into account issues including energy efficiency and reliability of the data since the nodes are required to use limited battery resources [36, 37].

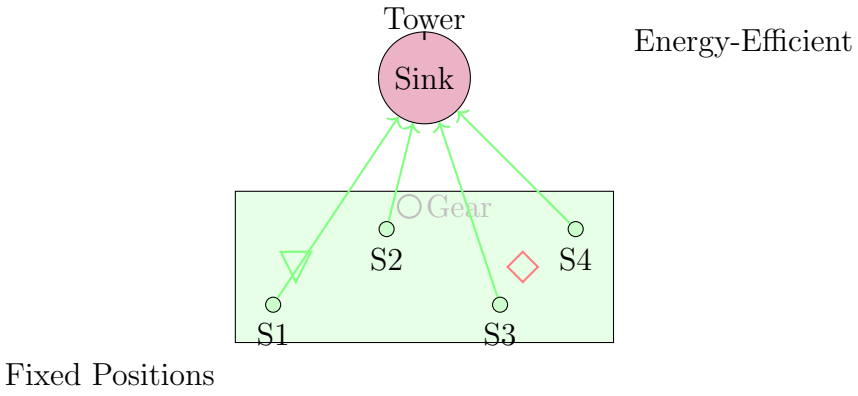


FIGURE 2.3: Static Wireless Sensor Network with Fixed Nodes

2.2.3.4 Mobile Wireless Sensor Networks

in MWSN, sense nodes are fixed to mobile objects that are mobile inside the networks. This mobility improves the coverage and adaptability of the network, making it appropriate for mobile applications, such as either wildlife monitoring or disaster recovery scenario [38]. Mobile sinks are nodes that go around collecting the data from stationary sensors, and their integration can enhance data collection efficiency significantly by minimizing data transmission energy [37], [33, 35].

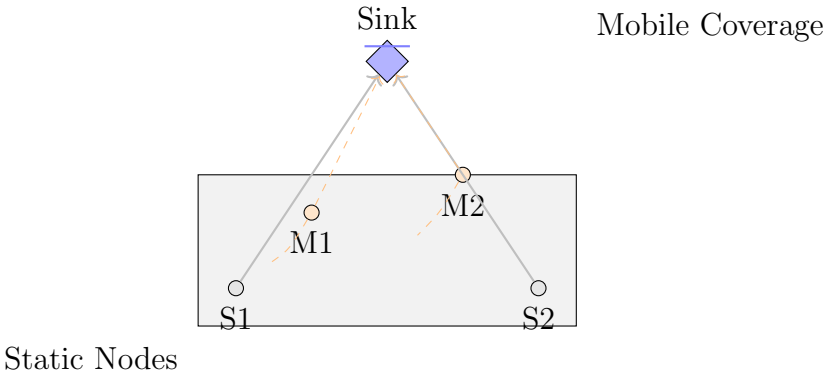


FIGURE 2.4: Mobile Wireless Sensor Network with Dynamic Nodes

2.2.3.5 Hybrid WSNs

HWSNs consist of static and moving nodes, which use the feature of both types of configurations. The above yields a potential application in smart agriculture and animals tracking, where a stable backbone of static sensors interrogates, to maximize the data acquisition and monitoring [38].

By combining the best features of static and mobile networks, the hybrid approach provides flexibility and efficiency for data collection.

Hybrid Wireless Sensor Networks combine both static and mobile nodes, leveraging the strengths of each configuration. This architecture is beneficial for applications like smart agriculture and animal tracking, where a stable backbone of static sensors interacts with mobile nodes to optimize data collection and monitoring.

A hybrid method overcomes the challenges associated with static or mobile-only networks, offering improved flexibility and efficiency in data collection.

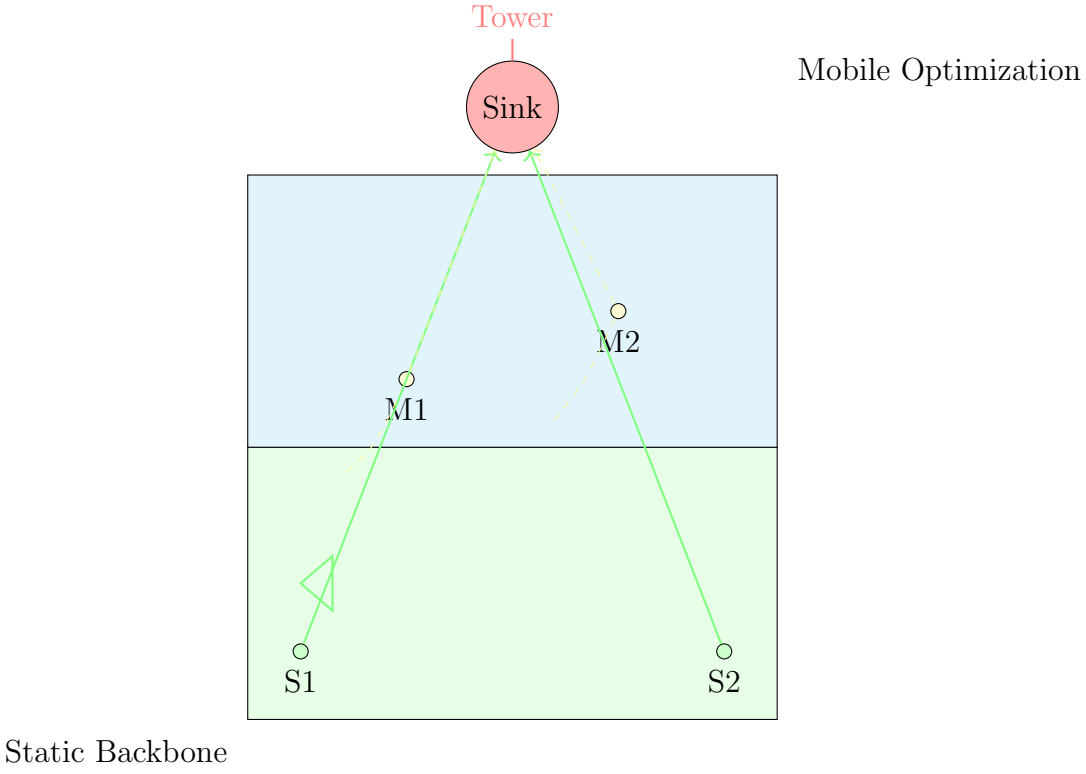


FIGURE 2.5: Hybrid Wireless Sensor Network Architecture

2.2.3.6 Wireless Body Area Networks

WBANs are a type of WSN focused on medical applications, which involve sensors being mounted on or adjacent to the human body to track vital signs and health statistics.

These networks have low power consumption and high qualities to deliver real-time data in reading of tele-medicine and personal health monitoring [39].

Considering the unique challenges associated with human body motion such as mobility etc, the design of WBANs should also comply with continuous monitoring requirements but not at the cost of user's comfort.

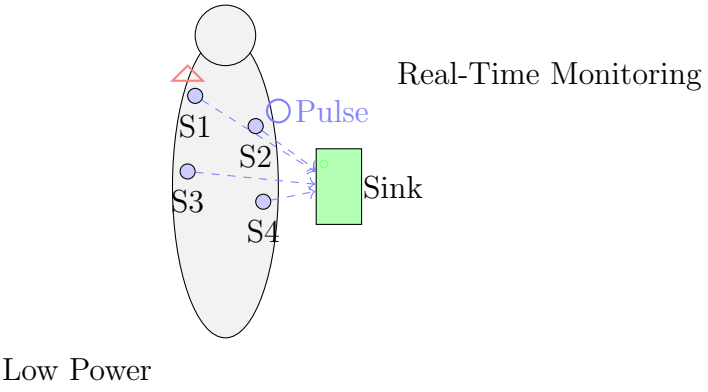


FIGURE 2.6: Wireless Body Area Network on Human Body

2.2.3.7 Multimedia Wireless Sensor Networks

WSN that is defined as MWSN often consists of a network of sensors that can capture and convey multimedia data like video, audio, and images.

They are ideal for use in applications needing higher levels of detailed and continuous analysis, like surveillance, traffic observation, and environmental monitoring [40].

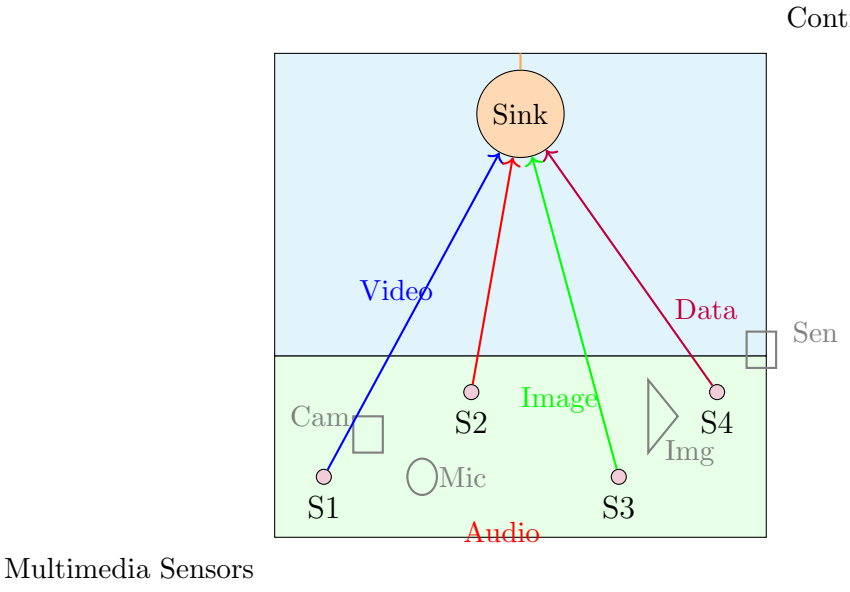


FIGURE 2.7: Multimedia Wireless Sensor Network Data Flow

2.2.3.8 Wireless Sensor Network Underwater and Underground

Highly specialized WSNs exist for difficult-to-access environments like underwater or underground. UW-WSN are used for applications like ocean exploration, pollution monitoring, and disaster prevention [41], while UG-WSN are employed in applications like mining, soil monitoring, and infrastructure monitoring.

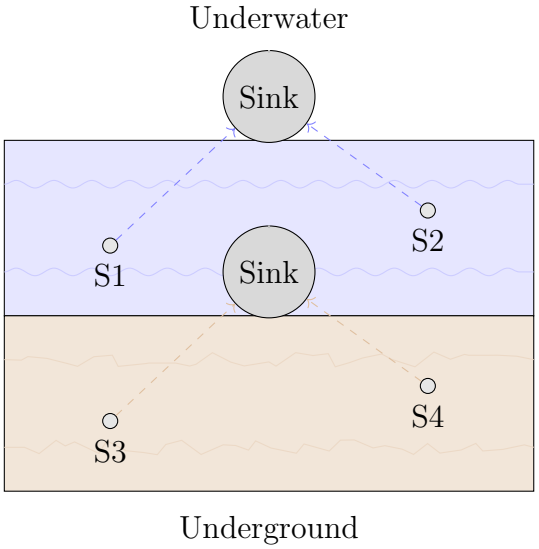


FIGURE 2.8: Underwater and Underground WSN Deployment

2.2.3.9 Aerial WSNs:

Aerial Wireless Sensor Networks, which utilize drones or other aerial platforms to deploy sensor nodes, have also been explored for applications such as precision agriculture, disaster response, and environmental monitoring.

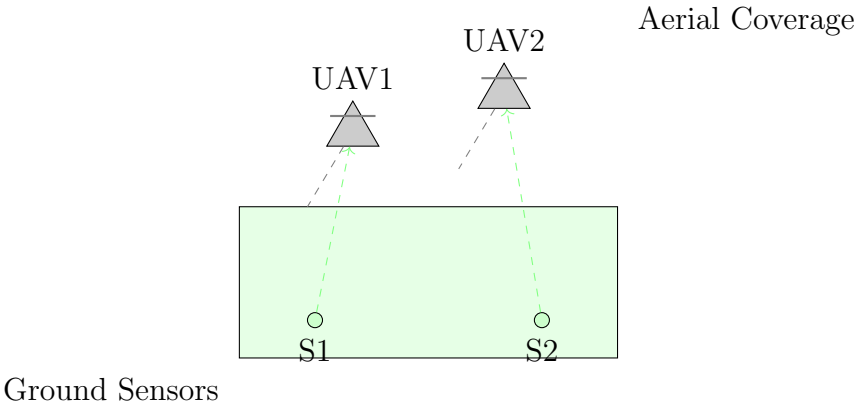


FIGURE 2.9: Aerial Wireless Sensor Network with UAVs

2.3 Static Wireless Sensor Network

Static Wireless Sensor Networks find use across many deployment settings but face distinct problems and barriers when placed in land, underground, and underwater conditions. Sensor networks have different types depending on their location type which includes terrestrial nodes overground and underground devices.

Sensor nodes in static WSNs stay in place after installation because their fixed deployment proves useful for environmental monitoring industrial control and military watch duties [42].

2.3.1 Network Topology and Routing Protocols

The working of static WSNs is heavily affected by their network topology, which determines how data packets are routed from sensor nodes to the sink. Common topologies used in static WSNs include flat/unstructured, chain, tree, and cluster-based configurations.

Each topology has its advantages and limitations:

i. Flat/Unstructured Topology: All nodes have equal roles, making it simple to deploy but less scalable for large networks [43].

ii. Chain Topology: Nodes are connected in a linear sequence, which is energy-efficient but vulnerable to node failures [44].

iii. Tree Topology: Nodes are organized hierarchically, enabling efficient data aggregation but requiring careful management of parent-child relationships.

iv. Cluster-Based Topology: Nodes are grouped into clusters, with cluster heads responsible for data aggregation and transmission.

    This approach offers better scalability and energy efficiency while incurring overhead for cluster formation and maintenance.

Hybrid topologies mixing different configurations (e.g, star, tree, mesh, etc.) are commonly adopted for large scale static WSNs.

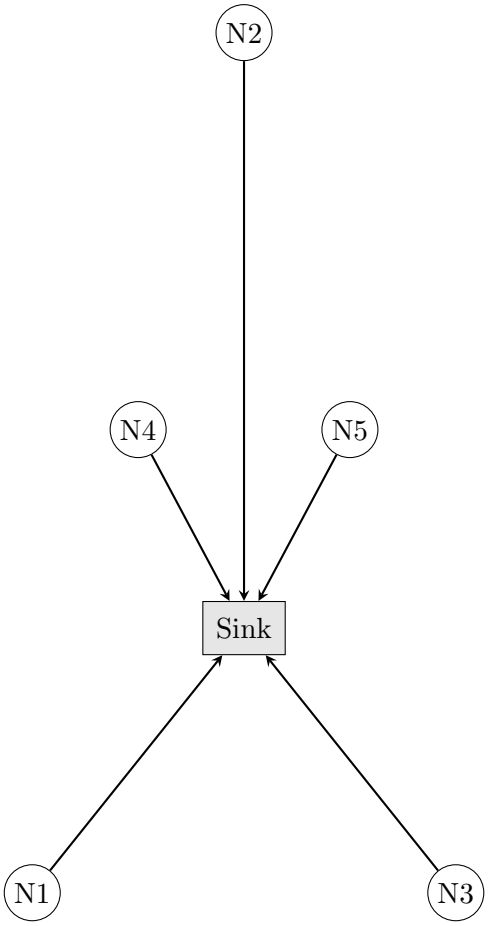


FIGURE 2.10: Flat/Unstructured Topology: All nodes connect directly to the sink.

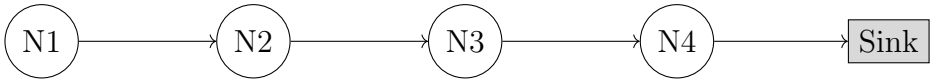


FIGURE 2.11: Chain Topology: Nodes are connected in a linear sequence to the sink.

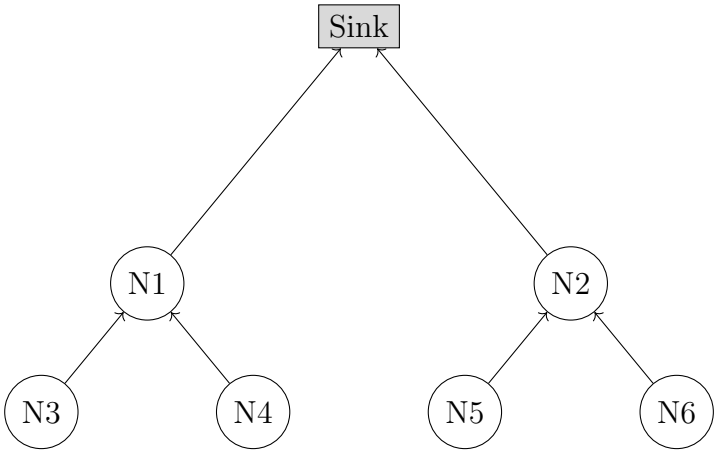


FIGURE 2.12: Tree Topology: Nodes are organized hierarchically with parent-child relationships.

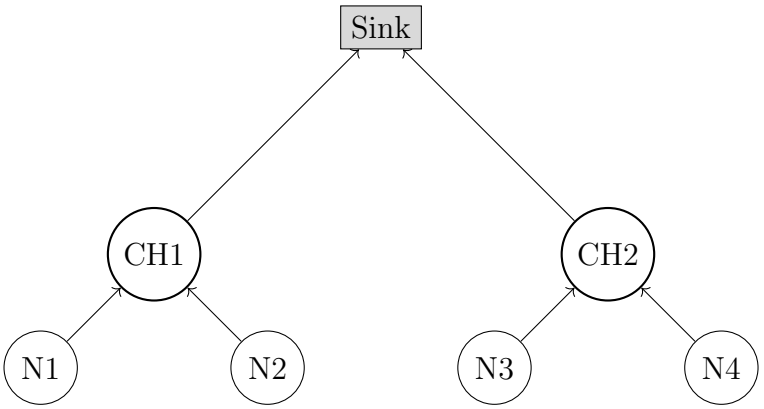


FIGURE 2.13: Cluster-Based Topology: Nodes are clustered with cluster heads.

As a result, hybrid topologies improve the performance of the network and help balance energy consumption while ensuring communication reliability. These topologies and resource constraints of SN must be taken into account for routing protocols for static WSNs so that the data we send reaches its intended destination efficiently.

2.3.2 Challenges in Static WSNs

Routing the key in static WSNs, so designing an effective routing protocol is crucial for reliable data delivery. Key challenges include:

i. Energy Efficiency: In static WSNs, each node is typically battery-powered, so energy conservation becomes a critical issue. keeping the network and routing protocols working and maintaining as low energy consumption as possible.

ii. Data Aggregation: They are the techniques that are required to aggregate the data efficiently such that the redundant transmissions can be reduced and which can prolong the network lifetime.

iii. Network Scalability: The routing protocol needs to scale efficiently with the increasing number of nodes without degrading performance.

iv. Limited resources: Static WSNs are often used under strict resource constraints, including limited memory, computing, and bandwidth. These constraints need to be considered in routing protocols for reliable operation.

2.3.3 Unique Characteristics of Static WSNs

No node mobility issues are faced by static WSNs, unlike mobile WSNs. Nonetheless, the fixed nature of these types of content also has curated complexities, including:

i. Deployment of nodes: Static nodes need to be deployed in a manner that improves coverage and connectivity .

ii. Fault Tolerance: In a static WSN, failure of individual nodes can result in coverage holes and network partitioning; thus, a robust fault-tolerance mechanism is mandatory.

iii. Network Lifetime: static nature of the network provides that the failure of nodes due to battery depletion and/or hardware issues might significantly affect the overall network lifetime.

iv. Network Maintenance: Static nodes, once deployed, cannot readily be repositioned to fix coverage holes, or other node failures.

v. Environmental Factors: Static WSNs in extreme environments (e.g., underground or underwater) need to face the signal attenuation, interference and physical damages.

To summarize, static wireless sensor networks(WSNs) consist of sensors that are immobile, making it well-suited for applications that require long-term data collection and monitoring.

Yet, these challenges, such as limited data efficiency, scalability, and computational restrictions, pose important issues and need to be addressed for the growth and evolution of.

2.3.4 Energy Harvesting Static Wireless Sensor Networks

Wireless Sensor Networks (WSNs) are deployed in various environments, including terrestrial, underground, and underwater settings, each presenting unique challenges and constraints.

Energy Harvesting Wireless Sensor Networks (EH-WSNs) are a specialized category of WSNs that utilize energy harvesting technologies to power sensor nodes.

Unlike traditional WSNs, which rely on finite battery resources, EH-WSNs harvest energy from ambient sources such as solar, thermal, vibration, or radio frequency (RF) signals, enabling sustainable and long-term operation.

Energy harvesting mechanisms play a crucial role in the design and deployment of EH-WSNs.

2.3.4.1 Energy Harvesting Sources and Mechanisms

Common energy harvesting sources for EH-WSNs include [23]:

i. Thermal Energy:Thermoelectric generators can harvest energy from temperature differences, such as those found in industrial settings or human body heat.

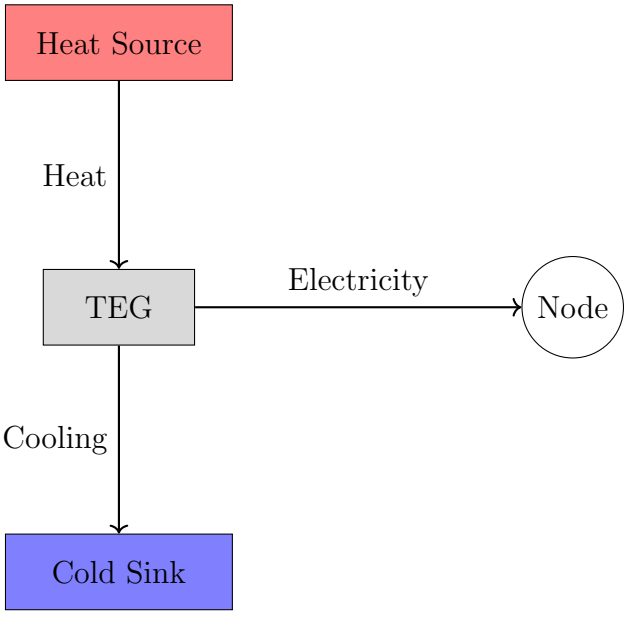


FIGURE 2.14: Thermal Energy Harvesting: Thermoelectric generators (TEG) convert temperature differences into electrical energy.

ii. Vibration Energy: Piezoelectric, electromagnetic, or electrostatic transducers can convert mechanical vibrations into electrical energy, suitable for EH-WSNs installed on machinery or structures.

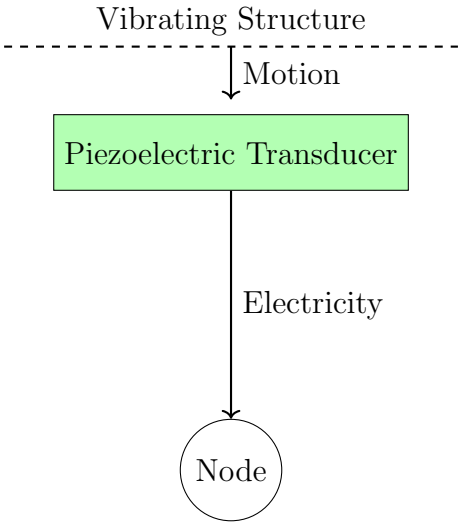


FIGURE 2.15: Vibration Energy Harvesting: Piezoelectric transducers convert mechanical vibrations into electrical energy.

iii. Radio Frequency Energy: Ambient RF signals, such as those from wireless communication networks, can be harvested using RF energy harvesters and used to power sensor nodes [23, 45].

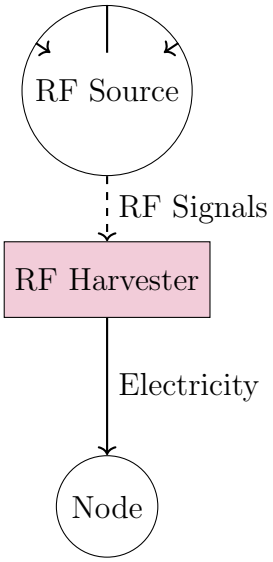


FIGURE 2.16: Radio Frequency Energy Harvesting: RF harvesters convert ambient RF signals into electrical energy.

The choice of energy harvesting source depends on the specific application, environmental conditions, and energy requirements of the EH-WSN.

Energy harvesting mechanisms in EH-WSNs typically involve the following components:

i. Energy Harvesting Unit: Responsible for converting the ambient energy source into electrical energy, such as photovoltaic cells, thermoelectric generators, or piezoelectric transducers.

ii. Energy Storage Unit: Stores the harvested energy for future use, typically in the form of rechargeable batteries or supercapacitors.

iii. Power Management Unit: Manages the flow of energy from the harvesting unit to the storage unit and the sensor node, ensuring efficient power utilization.

The process of solar energy harvesting is shown in Figure 2.17 .

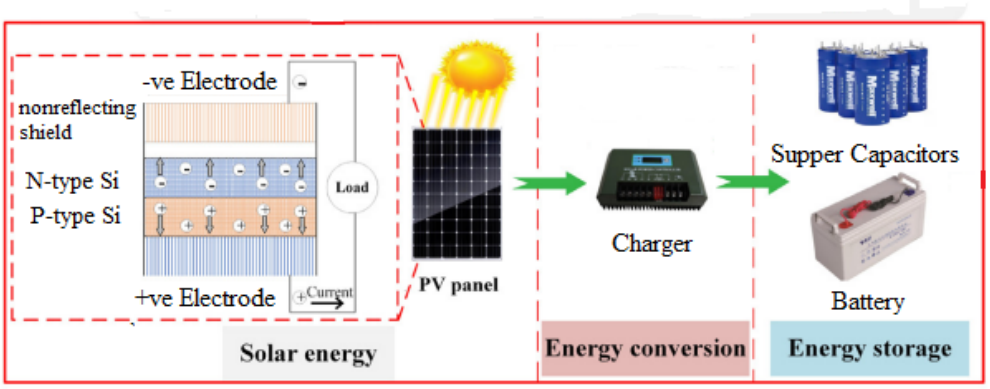


FIGURE 2.17: Solar Energy Harvesting Process.

Thus, the design of the energy harvesting mechanisms should be conducted with regards to energy conversion efficiency, power density, form factor, and cost to optimize the overall performance of the EH-WSNs. [23, 46, 47].

2.3.5 Routing in EH-WSN

A key challenge in EH-WSN is designing efficient routing protocols that can adapt to the dynamic energy availability of the sensor nodes. Routing protocols in EH-WSNs should consider the following factors:

i. Residual Energy: The residual energy in the SNs energy storage should be a primary factor in routing decisions to ensure balanced energy consumption and network lifetime.

ii. Energy Harvesting: The current EH rate of every node should be considered to prioritize nodes with higher energy harvesting capabilities.

iii. Energy Consumption: The energy required for data transmission and other node activities should be minimized to avoid depleting the node's energy reserves.

Many routing protocols have been proposed for EH-WSNs, such as energy-efficient clustering algorithms [48] and reinforcement learning-based approaches [45] to optimize the selection of cluster heads and communication routes.

2.4 Artificial Intelligence

Applying the concepts of Artificial Intelligence (AI) and Machine Learning (ML), particularly **Reinforcement Learning (RL)**, to **static Energy Harvesting Wireless Sensor Networks (EH-WSNs)** presents a unique set of opportunities and challenges. Unlike Mobile WSNs, static EH-WSNs consist of stationary nodes that rely on harvested energy (e.g., solar, thermal, or RF energy) to operate. The primary focus in such networks is on **energy efficiency**, **energy-aware routing**, and **maximizing network lifetime**.

Below is a detailed discussion of how AI and RL can be applied to static EH-WSNs:

One of the key issues addressed in this domain is the autonomous configuration and adaptation of sensor nodes to their environment [49]. Existing solutions either rely on manual configuration or use static heuristics, which fail to adapt to changes in environmental conditions that affect energy harvesting. To address this, researchers have proposed the use of Reinforcement Learning, where the sensor nodes can learn and adapt their operational parameters like sensing rate, communication intervals, and computation tasks to maximize energy utilization and application utility [49] [50] .

Another critical aspect is the design of routing protocols for energy-efficient data delivery. Conventional routing algorithms do not account for the dynamic nature of energy availability in EH-WSNs. Reinforcement Learning-based approaches have

been developed that enable nodes to intelligently select cluster heads and routing paths, balancing energy consumption and network lifetime. For example, one study proposed a RL-based energy-aware clustering algorithm, where neighboring nodes select an optimal cluster head by observing factors like energy consumption and distance to the base station [51]. A two-layer RL network was also framed by anotherwork to optimize the sum rate and energy prediction error simultaneously in IoT networks [45].EH-WSNs will be a promising solution for IOT and other data traffic applications with higher energy efficiency and longer network lifetime leading to better performance by integrating the AI and RL techniques.

2.4.1 Contribution of AI and RL in Static EH-WSNs

i. Adaptive Decision-Making: The learning capability of Reinforcement Learning allows the sensor nodes to autonomously update their operational parameters, which include duty-cycling, sensing rate and communication intervals, to optimally respond to the dynamic environmental conditions as well as energy harvesting patterns without human intervention or the need for a static heuristic [49] [51].

ii. Energy aware routing: RL-based routing protocols may select energy-efficient paths and cluster heads, striking a trade-off between energy consumption and maximizing the network lifetime [52] [45].

    To this end, RL based clustering algorithms have been designed for EH-WSNs based on the interaction of adjacent nodes to select the most suitable cluster head by taking into account their energy levels and distances from the base station, thus maximizing their network lifetime and energy efficiency.

iii. Smart Resource Management: Reinforcement Learning methods can help in managing the available energy in such a way that it is consumed in the most efficient way that meets the application's Quality of Service requirement by employing techniques to optimize resource allocation in terms of energy, computation, and communication for EH-WSNs [53].

iv. Adaptive Energy Harvesting: On top of this, RL can also be applied to allow the sensor node to adapt their energy harvesting and storage strategy

dynamically given the predicted energy availability and application requirements, improving the overall energy efficiency and reliability of the EH-WSN [54].

In conclusion, integrating Artificial Intelligence, particularly Reinforcement Learning, has significant potential to address the unique challenges faced by static Energy Harvesting Wireless Sensor Networks.

Leading to improved energy efficiency, prolonged network lifetime, and enhanced application performance.

2.5 Machine Learning

Machine learning is a field within Artificial Intelligence that enables computers to learn patterns from data and make predictions or decisions without explicit programming [55]. Machine learning algorithms involve feeding computers with large amounts of data, rather than following rigid instructions. By analyzing the data, machine learning algorithms can identify patterns and make predictions. But it is all automated, which is more efficient, reliable, and cost-effective than people writing code to do the same, because algorithms learn how to do the job better over time without any need for ongoing human involvement. Recently, the use of ML has proliferated, all the more with the availability of large data sets with the Big Data era [56].

Machine learning has found applications in various area's, including computer vision, NLP, and robotics, among others [57, 58]. Various techniques can be employed in machine learning, including supervised, unsupervised, and reinforcement learning approaches. Supervised learning: it is a type of learning in which the algorithm is trained on labeled data, i.e. the desired output is known. This allows the algorithm to predict on new data that it has not seen before[59]. Unsupervised learning is concerned with discovering patterns and structures in data without labels or a predefined output [60]. Machine learning algorithms become more effective as they are exposed to larger datasets, making them well-suited for complex and diverse data [61]. Reinforcement learning is a machine learning approach that

involves an agent learning by interacting with an environment. The agent takes actions and receives rewards or penalties, which guide its learning process [62, 63]. In addition to supervised, unsupervised, and reinforcement learning, there are also hybrid approaches that combine these techniques. Machine learning have become increasingly popular in classification, regression and density estimation problems, to name a few cutting-edge research has occurred in fields such as engineering, computer science, bioinformatics, computer vision and graphics processing. Wireless Sensor Networks (WSNs) and Internet of Things (IoT) have been the subject of a significant number of studies to apply machine learning techniques for their fields [12]. These applications have enhanced the energy harvesting, data processing, fault tolerance, security and network's self-organizing within WSNs, removing the necessity for manual reprogramming. These progresses resulted in development of more efficient, robust and intelligent wireless sensor networks which can now be adapted to various applications. As wireless sensor networks are often battery-powered, they use duty cycling to conserve energy. This means that the nodes are in a idle state and only active when some event occur. This approach enables the prolongation of the network's operational duration [64, 65].

Various WSN applications necessitate the achievement of distinct performance objectives, including low latency, good energy efficiency, and energy balance. These metrics often involve inherent conflicts; for example, emphasizing energy conservation typically leads to nodes spending more time in dormant mode, which can elevate response times. Such delays can be especially harmful in applications that are sensitive to latency [66, 67]. Thus, striking an optimal equilibrium among these performance metrics is crucial for satisfying the distinct requirements of varied applications. Nevertheless, achieving this balance represents a formidable challenge in duty-cycled wireless sensor networks.where the dynamic activity patterns of nodes and energy limitations further complicate the design and management of the network.

2.5.1 Methods of Machine Learning

A review of literature shows that many researchers have employed machine learning algorithms to solve issues related to WSNs [68].

The taxonomy of ML techniques is shown in figure 3.3.

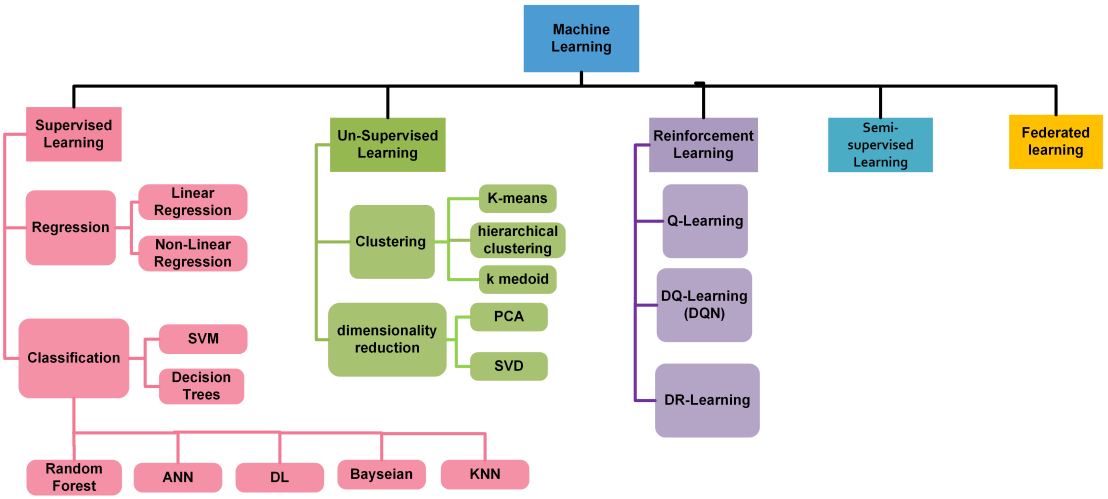


FIGURE 2.18: Taxonomy of machine learning algorithms.

2.5.2 Supervised Learning

One of the key areas of machine learning is supervised learning. Algorithms are trained using labeled datasets to find class of data or estimate output.

The task is to infer a function $g : a \rightarrow b$ that describes the best possible approximation of the output given the input. The supervised learning algorithms have an important role in developing a model that learns the mapping between the input features and the target outputs by capturing the relations and dependencies between this sets.

The model learns the relationship of prior datasets for predicting a new data. The main purpose of supervised learning is to create an accurate and precise prediction fro unseen data based on the model, through the use of the learned function [69], [70].

Supervised learning can help to solve number of problems in WSNs, which are given below:

i. Node Localization: Improving Positioning Precision [71].

ii. Coverage: Improving the spatial coverage area of distribution of sensor nodes.

iii. Routing: Optimizing the paths of data packets [72].

iv. Data Aggregation: Aggregating and consolidating data efficiently.

v. Congestion Control: Monitoring and controlling network traffic to avoid congestion.

vi. Energy Harvesting: Ensure energy efficient and sustaining energy harvesting.

supervised learning approaches have been used for solving these challenges in WSNs [73]. The usage of supervised learning algorithms, including decision trees, NN, and SVM, has provided good results in WSN studies [73]. Indeed, most of these algorithms are able to learn from labeled data and transfer knowledge to unseen scenarios, making them increasingly popular for improving the performance and reliability of wireless sensor networks.

2.5.3 Unsupervised Learning

Unsupervised learning refers to a type of machine learning where the model learns patterns and relationships within the input data without predefined output labels. Unlike supervised learning, where the model is trained using labeled data, unsupervised learning explores the underlying structure of the dataset without prior knowledge of the expected outcomes. This approach is commonly used for tasks such as anomaly detection, clustering, and dimensionality reduction.

In the context of wireless sensor networks (WSNs), unsupervised learning plays a crucial role in addressing key challenges such as connectivity, routing, and data aggregation [73–75].

One of the most frequently used clustering techniques in WSNs is the **k-means clustering algorithm**, particularly within the LEACH (Low-Energy Adaptive Clustering Hierarchy) protocol. The equation for centroid update in k-means clustering is shown in equation 2.1.

$$\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j \tag{2.1}$$

with μ_i being the new centroid of the i^{th} cluster, and C_i the set of points belonging to that cluster.

PCA is another common technique exposure through unsupervised learning that is used, primarily for dimensionality reduction / feature extraction. PCA compresses the high dimension information to low dimensional plane while retaining the highest variance of the data.

This process removes duplicated features by helping those more significant features of the dataset to become evident and makes the data easier, especially for better visualization and computation.

PCA does dimensionality reduction by projecting the data into its principal components The covariance matrix needed in PCA is defined through equation 2.2.

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \tag{2.2}$$

The eigen functions, or principal components, are found by solving the eigenvalue equation 2.3.

$$Sv = \lambda v \tag{2.3}$$

Here v are the eigenvectors and λ are the eigenvalues. These eigenvalues represent the variance explained by each principal component and can be used to identify the most meaningful dimensions of the data. By leveraging unsupervised learning techniques such as clustering and dimensionality reduction, WSNs can enhance their efficiency in data processing, energy conservation, and network optimization.

2.5.4 Reinforcement Learning

RL is based on the ability of an agent to learn to behave optimally through interaction with some sort of environment, in a reward-based manner as opposed to a human-defined behaviour. This process is similar to how a beginner learns a skill from practice and feedback, adjusting by the nuances of challenges over

time. Reinforcement learning is particularly valuable in aspects of Wireless Sensor Networks (WSNs) collections of distributed sensors to measure variables such as temperature or movement. These networks suffer from lasting constraints (especially limited energy) that must be managed intelligently in order to maintain performance. In Energy Harvesting WSNs (EH-WSNs), the complexity increases due to the variability in the power supply available as nodes harvest energy from random ambient sources like solar energy or radio frequency energy. The strength of RL is in its ability to handle such uncertainty and optimize decisions on-the-fly. This part explores RL's fundamentals, mathematical framework, and its utility in WSNs and EH-WSNs, using recent literature to demonstrate its effectiveness.

Q-learning, a widely adopted algorithm, refines the Q-function iteratively:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[R_{t+1} + \gamma(\max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t))] \tag{2.4}$$

Here, α governs the learning rate, enabling adaptation to evolving conditions—a critical feature for WSN applications. Figure 2.19 It depicts an agent (a sensor node) within an environment (the WSN). The agent observes its current state s_t, such as battery level or data backlog, and selects an action a_t, like transmitting or idling. The environment sends two arrows back: one with reward r_{t+1} (e.g., Positive for efficient transmission; Negative for overuse) and another with the next state s_{t+1}. The loop shown represents the agent continuously improving its policy over iterations, indicating the iterative approach to learning in RL the process is illustrated in figure 2.19.

2.6 Applications of Machine Learning in WSNs

Here we show the application of machine learning (ML) techniques and algorithms in case of wireless sensor networks (WSNs). In summary, we have discussed how ML can be helpful in solving these issues in WSN. It investigates the appropriateness or requirements of different ML techniques for several WSN tasks.

Machine learning (ML) has surfaced as a game-changer for wireless sensor networks (WSNs) by providing compelling solutions for fundamental challenges such as

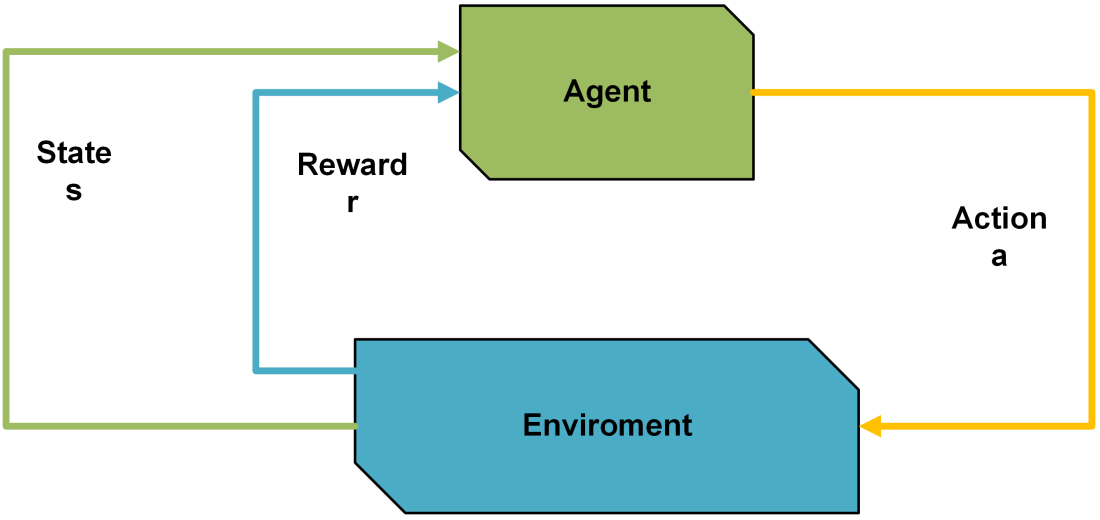


FIGURE 2.19: Reinforcement Learning.

energy constraints, data processing and dynamism adaptability. Let's take a look at how ML techniques are leveraged to tackle various WSN challenges, backed by the recent research data.

2.6.1 Energy Management

Due to the limited power supply with the sensor nodes, energy management is an essential aspect of WSNs.

Reinforcement Learning (RL) algorithms like the Q-learning are used to allow the nodes to learn a sleep schedule and/or a transmission schedule dynamically that minimizes their energy consumption during the Packet Transmission.

In a review [76], demonstrate how.deep learning (DL) techniques, consisting of RL, minimize the computational complexity while maximizing energy consumption in WSNs, at the expense of increasing the training time.

The clustering methods (e.g., the k-means) are significant as they push the nodes into energy-efficient clusters, reducing communication.expenses. As depicted in Figure 2.20, there are three sensor nodes with their respective battery indicators which report back energy metrics to the Q-learning agent to make.the optimizations that can potentially redistribute energy resources and ultimately enforce sleep mode.

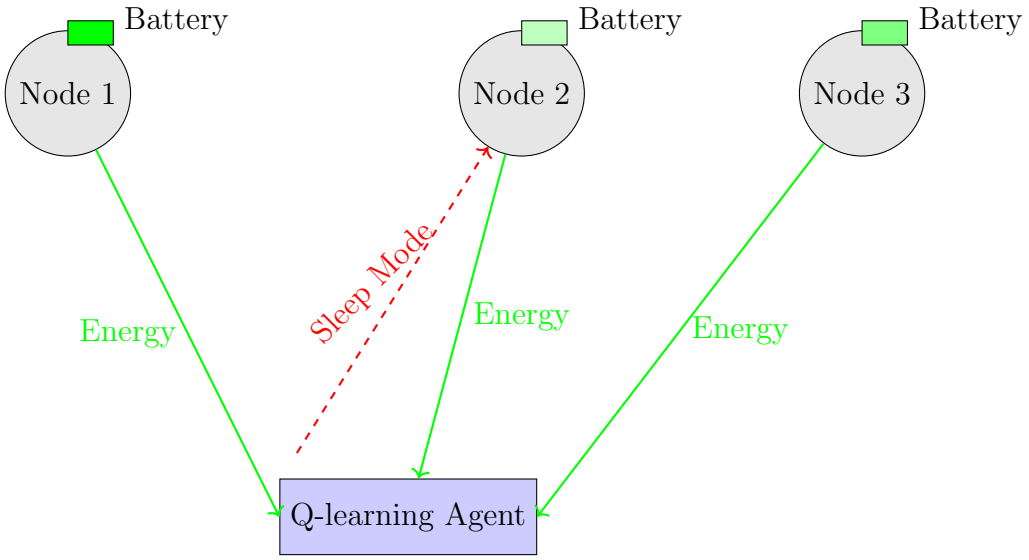


FIGURE 2.20: Energy management using RL in WSNs, showing optimized energy distribution and sleep mode.

2.6.2 Data Aggregation and Compression

Data aggregation and compression tackle the issue of redundant data in WSNs, a common problem due to spatial and temporal correlations. Techniques like principal component analysis (PCA) and autoencoders compress data effectively.

A 2024 study by Chen et al. proposes a convolutional neural network (CNN) and attention-based joint source-channel coding approach for semantic communications in WSNs, achieving efficient data transmission with reduced bandwidth needs [77].

This process is depicted in Figure 2.21, where two sensor nodes send raw data (wavy blue lines) to a CNN block, which compresses it (narrower wavy line) before forwarding it to a base station, visually emphasizing the reduction in data size.

2.6.3 Anomaly Detection

Anomaly detection ensures WSN reliability by identifying faulty readings or security threats. Supervised learning methods like support vector machines (SVM) and decision trees excel when trained on labeled data. For instance, a 2024 paper

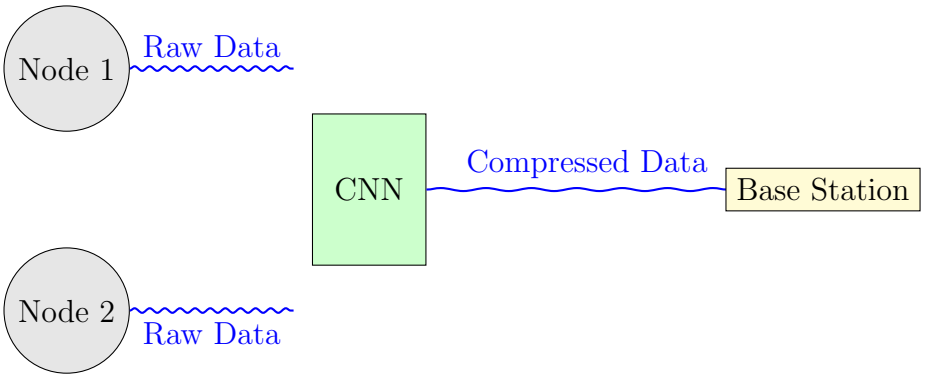


FIGURE 2.21: Data aggregation and compression using CNN, illustrating data reduction in WSNs.

by Al-Qurishi et al. demonstrates Random Forest achieving a 98% accuracy in detecting cyberattacks in IoT-based WSNs, outperforming SVM (90%) and k-NN (95%) [78]. Unsupervised techniques like DBSCAN are valuable when labeled data is scarce. Figure 2.22 shows this process, with a cluster of blue "Normal Data" points separated by a dashed decision boundary (Random Forest Classifier) from a red "Anomaly" point, highlighting the classification mechanism.

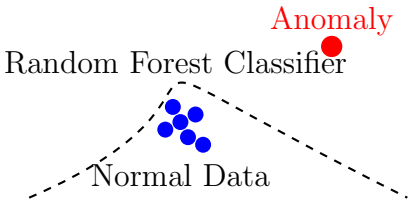


FIGURE 2.22: Anomaly detection using Random Forest, showing normal cluster and anomaly separation.

2.6.4 Routing and Network Optimization

Routing and network optimization adapt to changing conditions like node failures or interference. Genetic algorithms and neural networks optimize paths dynamically. A 2023 study by Singh et al. explores ML-based routing protocols, showing how neural networks predict traffic patterns to balance network load [79]. Figure 5.12 depicts this, showing a network with multiple possible paths (gray dashed lines) and an optimized blue path selected by a neural network, connecting a source to a destination node efficiently.

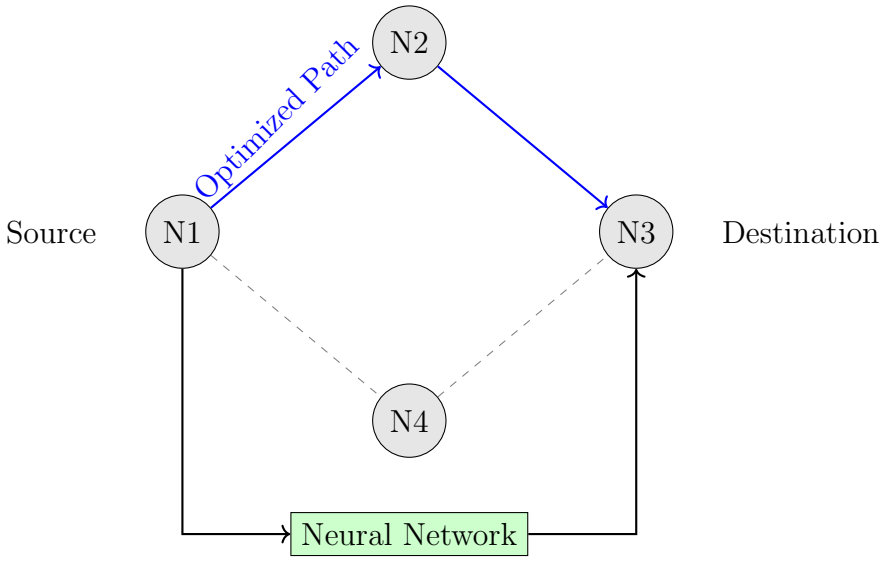


FIGURE 2.23: Routing optimization using neural networks, showing an optimized path in WSNs.

2.6.5 Suitability of ML Techniques

The suitability of ML techniques varies by application. Lightweight algorithms like decision trees or naive Bayes are ideal for resource-limited WSNs [76].

High-stakes applications, such as healthcare monitoring, may leverage deep learning for precision, despite higher energy costs [77]. Real-time needs favor fast methods, while predictive tasks might use recurrent neural networks (RNNs).

In summary, ML empowers WSNs to overcome energy, data, and adaptability challenges, tailoring solutions to specific needs, as supported by recent research [76–80].

2.7 Benefits and Drawbacks of ML in WSN

Integrating ML into WSNs can supercharge their capabilities, but it also comes with some challenges. Let's break it down.

2.7.1 Benefits of ML in WSNs

i. Smarter Data Processing: ML can analyse sensor data on the fly, identifying patterns or anomalies like detecting a forest fire from temperature spikes

without needing constant human oversight. This reduces latency and makes the network more autonomous [81].

ii. Energy Efficiency: By optimizing how sensors communicate or sleep (e.g., predicting when data collection isn't needed), ML can extend battery life, which is critical since these devices often run on limited power in remote areas [82].

iii. Improved Accuracy: ML algorithms can filter out noise or calibrate sensor readings, making the data more reliable. For instance, they might distinguish between a real event and a glitch in a seismic monitoring system [83].

iv. Adaptability: WSNs often operate in dynamic environments. ML lets them adapt to changes like adjusting to new weather patterns or network failures without requiring manual reprogramming [84].

v. Predictive Capabilities: ML can forecast trends, like predicting equipment failure in industrial WSNs, enabling proactive maintenance and reducing downtime [81].

2.7.2 Drawbacks of ML in WSNs

i. Resource Constraints: Sensor nodes have limited processing power, memory, and energy.

Running complex ML models on them is tough[82].

ii. Complexity and Cost: Developing and deploying ML models for WSNs requires expertise and can increase the overall cost of the system, which might not suit low-budget deployments [83].

iii. Communication Overhead: If ML processing is offloaded to a central server (to save sensor energy), it means more data transmission, which can clog the network and drain power anyway [85].

iv. Security Risks: ML models can be vulnerable to attacks, like data poisoning, where bad inputs trick the system. In a WSN, this could mean false readings compromising an entire application, like a smart grid [84].

v. Training Challenges: ML needs data to learn, but WSNs might not always have enough high-quality, labeled data, especially in real-time or remote setups. Plus, retraining models on-the-go is tricky with limited resources [86].

Chapter 3

Literature Review

A brief introduction to Reinforcement Learning is provided in this chapter, along with its significance in the proposed work. This chapter start with the introduction, and conjuncture of routing protocols for Energy Harvesting WSN's, and then describes different types of routing protocols. It focuses especially on energy-efficient routing protocols. Then, brief discussions of RL research in the WSN domain followed by an overview of some existing RL-based routing protocols in the literature. Finally, the chapter presents details on the performance metrics used to evaluate various routing protocols in the context of EH-WSNs and the specific metrics selected to evaluate the routing protocol.

3.1 Introduction

Reinforcement Learning (RL) is a branch of machine learning that focuses on learning optimal decision-making strategies through interaction with an environment. Unlike supervised learning, which relies on labeled data, RL enables an agent to learn by receiving feedback in the form of rewards or penalties for its actions. The objective of RL is to determine a policy that maximizes long-term cumulative reward through trial-and-error interactions [87]. RL is fundamentally different from artificial neural networks (ANNs); however, neural networks can be employed as function approximators within RL frameworks, particularly in deep reinforcement learning (DRL).

Inspired by behavioral psychology and biological learning processes, RL models decision-making as a sequential process under uncertainty, making it well suited for dynamic and resource-constrained environments such as wireless sensor networks (WSNs) [88]. In such networks, nodes must continuously adapt routing decisions based on residual energy, traffic conditions, and network topology changes.

Recent advances in RL and deep learning have attracted significant attention from WSN researchers due to their ability to autonomously learn efficient routing policies without requiring explicit network models. RL-based approaches have been successfully applied to several WSN challenges, including routing [89], localization, data aggregation, intrusion detection, and energy management. In particular, the integration of RL with energy harvesting wireless sensor networks (EH-WSNs) has emerged as a promising solution to address the stochastic nature of harvested energy and the limitations of traditional static routing protocols.

3.2 Routing Protocols For WSNs

Routing protocols are often classified based on diverse features, including network topology and application requirements, indicating that a single routing approach may not be universally suitable or efficient for all types of WSN applications [72].

3.3 Network Architecture Protocols

The network architecture based approach has diverse path selection methodologies, including flat network path selection, which is well suited for large scale networks with many sensor nodes where assigning unique identifiers to each node is impractical [90].

As a result, content based path selection methodologies are employed, where all nodes in the network are regarded as equal and execute the same functions. Some widely recognized flat-based routing protocols are Flooding [91], Gossiping, and Sensor Protocols for Information via Negotiation (SPIN). [92]. The network architecture-based protocols can be devided into three following types.

3.3.1 Flat Routing Protocols

Flat routing protocols operate by treating all nodes in the network uniformly without assigning distinct roles or hierarchies. Two widely recognized examples are Flooding and Sensor Protocol for Information via Negotiation (SPIN).

3.3.1.1 Flooding Routing Protocol

Flooding is a straightforward mechanism where a node transmits data packets to all its neighboring nodes. This process continues until the packet reaches its destination or a maximum transmission limit is reached. Briefly, the work [91] proposed an improved flooding protocol, called NC-Flooding, for Wireless Sensor Networks (WSNs), which incorporates network coding to solve energy efficiency and reduce redundant transmissions. They innovate the protocol by employing mechanisms like active XOR coding, redundancy packet suppression, and local priority broadcasting to make the process of data dissemination efficient. Some advantages of NC-Flooding are energy saving, enhanced network throughput and increased lifetime of wireless sensor nodes in comparison to simple flooding strategies. Such enhancements have made it very applicable for power-limited WSNs. Nonetheless, the protocol comes with drawbacks, such as the overhead in computation and storage, due to network coding mechanisms used, and may not be suitable for sensor nodes that are resource-constrained. In addition, the performance of the method in highly dynamic or sparse networks is uncertain. Nonetheless, NC-Flooding holds significant potential for the design of energy-efficient WSNs and merits further exploration to reduce the gap between practical systems operating in diverse environments.

3.3.1.2 Sensor Protocol for Information via Negotiation

SPIN uses a data negotiation approach to optimize energy usage. before transmission Nodes share metadata to identify relevant data, ensuring that only necessary information is exchanged. The authors of SPIN [93] this improves the efficiency of data propagation through exploiting metadata negotiation to process information subsequently reducing energy wastage from redundant transmissions [94]

the authors highlight its data-centric routing approach, which minimizes network overhead by transmitting only to interested nodes. SPIN offers advantages such as energy efficiency, adaptability to dynamic conditions, and reduced overhead, but it also has notable disadvantages, including vulnerability to security threats like replay and modification attacks, limited scalability in large networks, and reduced robustness in attack-prone environments [95]. Despite these limitations, SPIN serves as a foundational protocol with significant potential for improvements in scalability and security.

3.3.2 Hierarchical Routing Protocols

Hierarchical routing schemes serve to reduce communication overhead as well as improve energy efficiency by using a clustering-based approach. Their clustering structure allows for easier data transmission and prolongs the lifetime of the network. Some well-known hierarchical protocols include LEACH, EM-LEACH, Mob M-LEACH, EEHC and HEED.

3.3.2.1 Low-Energy Adaptive Clustering Hierarchy

Heinzelman et al. [16] introduced the Low-Energy Adaptive Clustering Hierarchy (LEACH)—an energy-efficient protocol for wireless sensor networks (WSNs) based on altering cluster-head status among nodes using adaptive clustering. The protocol further incorporates local data fusion at cluster-heads to reduce the volume of data discouraging transmission to the base station. The protocol is staged over rounds, where each round contains a Setup Phase, in which a random selection of cluster heads (CHs) are generated, and a Steady Phase, where data is sent and aggregated. LEACH's most remarkable features are energy-efficient, which is realized through the data aggregation at the CH level, and load-balancing method, which rotates the role of CH among the nodes to equalize the energy consumption. This methodology decreases direct transmissions to the base station so much that the network lifetime is increased.

Using TDMA scheduling reduces the amount of communication, increasing scalability for larger networks.

3.3.2.2 LEACH Protocol Energy Model

The LEACH (Low-Energy Adaptive Clustering Hierarchy) protocol optimizes energy usage in wireless sensor networks (WSNs) through dynamic clustering and adaptive role rotation. The probability of a node n becoming a cluster head in a given round is determined by the equation (3.1). Where p is the desired percentage of cluster heads, r is the current round, and G is the set of eligible nodes. This ensures equitable distribution of cluster head roles across rounds, balancing energy consumption among nodes.

$$P(n) = \begin{cases} \frac{p}{1-p\cdot(r \mod \frac{1}{p})} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

Energy dissipation for data transmission is modeled as given in equations (3.2) and (3.3), where k is the data packet length, E_{elec} is the energy dissipation per bit, and ϵ_{fs} and ϵ_{mp} are coefficients for free-space and multipath propagation models. The threshold distance d_0 distinguishes between the two propagation models based on the energy required.

$$E_{\text{tx}}(k, d) = E_{\text{elec}} \cdot k + \epsilon_{\text{fs}} \cdot k \cdot d^2 \quad \text{for distance } d < d_0 \tag{3.2}$$

$$E_{\text{tx}}(k, d) = E_{\text{elec}} \cdot k + \epsilon_{\text{mp}} \cdot k \cdot d^4 \quad \text{for distance } d \geq d_0 \tag{3.3}$$

The threshold distance is expressed in equation (3.4) below:

$$d_0 = \sqrt{\frac{\epsilon_{\text{fs}}}{\epsilon_{\text{mp}}}} \tag{3.4}$$

The energy required to receive data is given by equation (3.5) below, where E_{elec} remains constant for each bit:

$$E_{\text{rx}}(k) = E_{\text{elec}} \cdot k \tag{3.5}$$

A single communication round involves a total energy dissipation modeled as in equation (3.6), where m is the number of cluster heads, N is the total number of nodes, E_{CH} is the energy consumed by a cluster head, and E_{nonCH} is the energy consumed by non-cluster head nodes:

$$E_{\text{round}} = m \cdot E_{\text{CH}} + (N - m) \cdot E_{\text{nonCH}} \tag{3.6}$$

The energy consumed by a cluster head is further detailed in equation (3.7), where $d_{\text{to\_BS}}$ is the distance to the base station, N_{cluster} is the number of nodes in a cluster, and E_{DA} is the energy needed for data aggregation per bit:

$$E_{\text{CH}} = E_{\text{tx}}(k, d_{\text{to\_BS}}) + E_{\text{rx}}(k \cdot N_{\text{cluster}}) + E_{\text{DA}} \cdot k \cdot N_{\text{cluster}} \tag{3.7}$$

All equations from (3.1) to (3.7) model the energy dynamics of LEACH, enabling efficient clustering and communication to extend the network operational lifetime.

3.3.2.3 Enhanced Multi-Hop LEACH

The authors in [96] propose EM-LEACH, an enhanced LEACH-based protocol incorporating multi-hop routing, adaptive cluster-head (CH) selection based on residual residual energy, and dynamic round time computation to address LEACH's limitations, such as single-hop communication and random CH selection. Simulations demonstrate improvements in energy efficiency, reliability, and network lifetime compared to LEACH and MR-LEACH. Key advantages include better energy efficiency through residual energy-based CH selection, improved reliability with higher packet delivery ratios, and dynamic adaptation that ensures fair energy distribution. However, the added complexity of multi-hop routing and adaptive mechanisms increases computational overhead, and assumptions like static nodes and homogeneous energy levels may limit real-world applicability. Additionally, the protocol's effectiveness is validated only through simulations, with no hardware testing. Overall, EM-LEACH significantly enhances energy efficiency and reliability, albeit with increased complexity and limited practical validation [96].

3.3.2.4 Model of the EM-LEACH Protocol

The proposed Enhanced Multi-hop LEACH (EM-LEACH) algorithm improves the LEACH protocol by optimizing energy efficiency, extending network lifetime, and increasing data throughput.

Cluster Head (CH) selection combines fairness and energy awareness, using a threshold $P(n)$ defined in Equation (3.8). Let $\beta = \frac{E_{\text{residual}}}{E_{\text{initial}}}$, where p is the desired percentage of cluster heads, r is the round number, and G includes nodes that have not been cluster heads in the last $\frac{1}{p}$ rounds.

$$P(n) = \begin{cases} \frac{\beta \cdot p}{1 - p \times \left(r \mod \frac{1}{p}\right)} + (1 - \beta) \cdot \frac{E_{\text{residual}}}{E_{\text{initial}}}, & \text{if } n \in G \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

Initially, $\beta = 1$, prioritizing fairness (p), while subsequent rounds dynamically adjust β, balancing cluster head selection between fairness and residual energy. EM-LEACH also introduces a variable round length to adapt energy consumption, as given in Equation (3.9), where $\gamma = 1.5$ yields the best results.

The initial round time is fixed to 150% of the LEACH round time, then decreases progressively to a minimum of 50% of the LEACH round time.

$$\text{Round Length} = \text{Initial Round Length} \cdot \left(1 - \frac{1 - \text{Energy Rates}}{\gamma - \text{Energy Rates}}\right) \tag{3.9}$$

Multi-hop routing mitigates energy losses by assigning levels to nodes, creating a tree-based routing structure where packets traverse from higher to lower levels toward the sink, as shown in Figure 3.1.

During setup, nodes broadcast advertisements, form clusters, and organize multi-hop connections. In the steady state, TDMA schedules reduce energy use by enabling nodes to sleep when idle. Cluster Heads (CHs) aggregate and forward data, dynamically updating residual energy for subsequent round computations.

These integrated processes reduce overhead, improve energy efficiency, and prevent network partitioning, with the complete process illustrated in Figure 3.2.

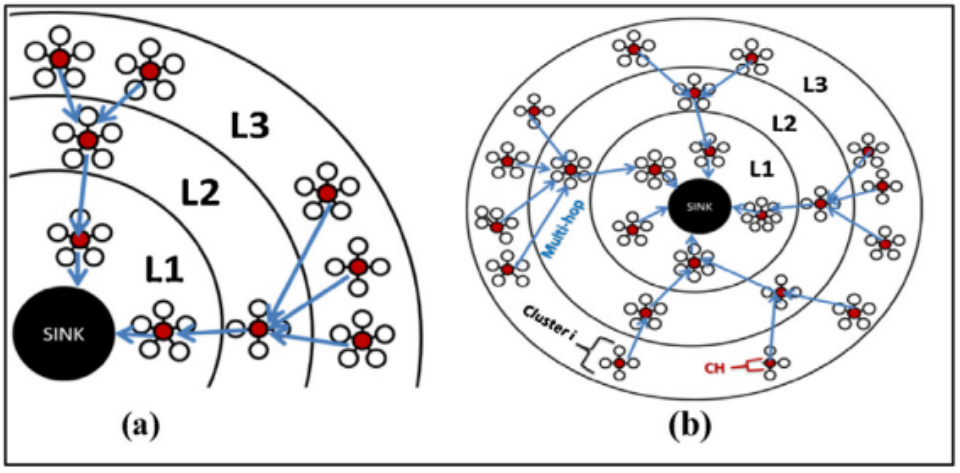


FIGURE 3.1: Tree-based routing structure in EM-LEACH:(a) Sink at the corner of Network. (b) Sink at the center of the network [96].

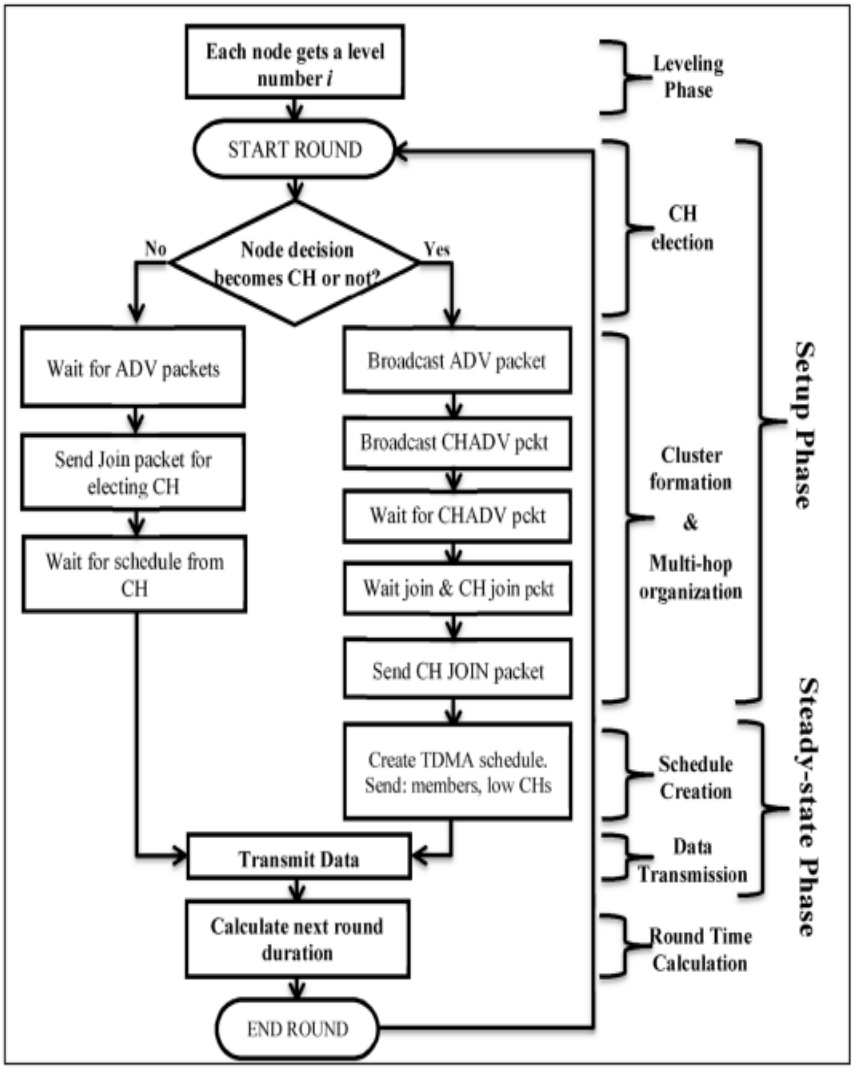


FIGURE 3.2: Complete process of the EM-LEACH protocol [96].

3.3.2.5 Mobility-Induced Multi-Hop LEACH Protocol

The authors in [97] classified nodes as normal, advanced, and super nodes based on energy levels and used residual energy and mobility parameters for dynamic CH selection, the proposed method outperformed the existing protocols.

3.3.2.6 Energy-Efficient Hierarchical Clustering

EEHC improves upon LEACH by introducing a hierarchical structure where sub-cluster heads aggregate data before forwarding it to the main cluster head [8]. This structure reduces energy consumption but increases latency, making it unsuitable for delay-sensitive applications.

3.3.2.7 Hybrid Energy-Efficient Distributed Clustering

HEED selects cluster heads based on residual energy and intra-cluster communication cost. Unlike LEACH, HEED adapts to dynamic network conditions, achieving better energy balancing. However, its iterative clustering process incurs higher computational overhead [98].

3.3.3 Location-Based Routing Protocols

Location-based routing protocols utilize the geographic positions of nodes to make routing decisions, aiming to optimize energy consumption and improve efficiency. These protocols are particularly useful in scenarios where node location information is readily available. Two notable examples are Geographic Routing and Energy-Aware Routing.

3.3.3.1 Geographic Routing

Geographic routing relies on the physical location of nodes to determine the optimal path for data transmission. By leveraging node coordinates, the protocol aims to minimize the path length and reduce energy consumption. Enhances routing

efficiency by selecting shorter and more direct paths. Requires accurate location information, which might involve additional hardware or computational costs.

3.3.3.2 Energy-Aware Routing

Energy-aware routing integrates geographic information with the energy status of nodes to distribute the communication load evenly. This prevents nodes with low energy levels from being overutilized [99] prolongs the network lifetime by balancing energy usage across nodes. Increases protocol complexity due to the need for continuous monitoring of node energy levels.

3.3.4 Reinforcement Learning-Based Routing Protocols

Recent studies have demonstrated that reinforcement learning (RL) offers a powerful framework for designing adaptive and energy-efficient routing protocols in wireless sensor networks, particularly in energy harvesting environments. Traditional routing protocols often rely on static heuristics and fixed energy assumptions, which limit their effectiveness under dynamic energy availability. RL-based routing protocols address this limitation by enabling sensor nodes to autonomously learn optimal routing decisions based on residual energy, network topology, and traffic conditions. Early RL-based schemes primarily focused on minimizing energy consumption or hop count using Q-learning, achieving notable improvements in network lifetime compared to conventional protocols. However, many of these approaches assumed non-harvesting energy sources and lacked long-term sustainability considerations. More recent research has extended RL-based routing to energy harvesting wireless sensor networks (EH-WSNs), incorporating stochastic energy arrival models and adaptive reward functions. Multi-agent reinforcement learning (MARL) approaches further enhance scalability and robustness by allowing distributed cooperation among nodes. Despite these advancements, existing protocols often overlook critical aspects such as energy neutrality, node death and revival dynamics, and explainable convergence behavior. These limitations motivate the development of advanced EH-aware RL-based routing protocols that jointly consider energy harvesting dynamics, distributed learning, and long-term

network sustainability. Routing protocols in Energy Harvesting Wireless Sensor Networks (EH-WSNs) must address the unique challenges of intermittent energy availability and resource constraints.

Reinforcement Learning (RL) has emerged as a powerful approach to design adaptive, energy-efficient routing strategies,This section reviews key studies on RL-based routing protocols for EH-WSNs, focusing on energy efficiency, adaptive power management, multi-objective optimization, clustering.

3.3.4.1 Energy-Efficient Routing Protocols

Energy efficiency is critical in EH-WSNs to maximize network lifetime under variable energy conditions.

Mutombo *et al.* [100] propose EER-RL (Energy-Efficient Routing based on Reinforcement Learning), a protocol for IoT networks that balances energy dissipation using residual energy and hop count to the sink as reward parameters.

Through a feedback mechanism, EER-RL enables nodes to learn optimal routing paths dynamically. Similarly, Mutombo *et al.* [52] introduce EBR-RL (Energy Balancing Routing based on Reinforcement Learning), which employs a comparable reward function to optimize energy efficiency in WSNs.

Ding *et al.* [101] provide a broader perspective, reviewing machine learning-based routing algorithms and emphasizing RL's role in achieving energy-efficient path selection through local node interactions.

More recently, Abuhamdah *et al.* [102] present DOS-RL (Dynamic Objective Selection RL), an energy-efficient routing protocol for Software-Defined WSNs (SD-WSNs), leveraging Q-learning with dynamic objective selection to adapt to network changes and energy constraints.

3.3.5 Reinforcement Learning Approaches

RL facilitates decentralized routing by enabling nodes to adapt based on environmental feedback. Mutombo *et al.* [100] highlight how nodes in EER-RL share

local information (e.g., energy levels, traffic load) to collaboratively refine routing decisions.

Al-Tous and Barhumi [103] apply the SARSA algorithm to minimize transmission delay in multi-hop EH-WSNs, balancing energy availability and latency.

Barat *et al.* [104] propose a cooperative energy management scheme in EH-WSNs using MARL, where nodes share local knowledge to optimize throughput under varying energy conditions, demonstrating improved coordination over single-agent RL.

3.3.5.1 Q-Learning-Based Routing

Q-learning-based protocols dynamically optimize routing paths by learning the Q-values for state-action pairs.

Nodes select actions (routing decisions) based on the Q-value, which is updated iteratively:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right] \tag{3.10}$$

where:

- s: Current states
- s': Next states
- a: Current actions.
- a': Next actions.
- α: Learning rate.
- γ: Discount factor.
- r: Reward.

This approach is adaptable but struggles with scalability in large networks due to the exponential growth of state-action pairs [17].

3.3.5.2 Multi-Agent Reinforcement Learning

In MARL, nodes act as agents that collaboratively optimize routing decisions. This distributed approach improves scalability and robustness but increases communication overhead for inter-agent coordination [105].

3.4 Wireless Sensor Networks and Their Limitations

Wireless Sensor Networks are comprised of sensor nodes that communicate wirelessly to collect data for various applications. These networks are typically deployed in distributed, often remote environments, making them prone to energy limitations. Sensor nodes are powered by small batteries, which have a limited lifespan, and replacing or recharging these batteries is often impractical due to the nodes' locations.

Sah and Amgoth (2020) provide an overview of the challenges related to energy management in WSNs, discussing several techniques that aim to extend the lifetime of sensor nodes. They highlight the importance of energy-efficient communication protocols that minimize power consumption during data transmission and ensure reliable operations across the network [106].

3.5 Energy Harvesting in Wireless Sensor Networks

Energy harvesting is a promising solution that allows sensor nodes to capture and store energy from ambient sources, such as solar, thermal, wind, or vibrational energy. By harvesting energy, WSNs can operate autonomously without frequent battery replacements, making them more sustainable for long-term deployments.

Among these techniques, solar energy harvesting is one of the most widely used due to its availability, cost-effectiveness, and low maintenance. Dvir et al. (2024)

explore the potential of energy harvesting in WSNs, noting that it plays a crucial role in mitigating energy depletion. Their study focuses on a cooperative reinforcement learning approach that helps distribute energy requirements across sensor nodes, thereby improving the overall energy efficiency and reliability of the network [105]. This approach is particularly beneficial for WSNs that rely on solar energy harvesting, as it ensures that nodes balance their energy use to extend operational life.

3.6 Solar Energy Harvesting

Solar energy harvesting involves capturing sunlight using solar panels and converting it into electrical energy to power sensor nodes. This method is particularly suited for outdoor WSNs where sunlight is abundant. However, the effectiveness of solar energy harvesting is influenced by factors such as weather conditions, geographical location, and the orientation of the solar panels.

Hao et al. (2022) propose an energy-efficient routing algorithm based on a greedy strategy designed for energy harvesting WSNs. This algorithm optimizes routing decisions based on the energy harvested through solar energy, ensuring that data is transmitted with minimal energy consumption [107]. Their approach is especially useful in scenarios where solar energy is the primary source of power for the nodes.

3.7 Multi-Hop Routing Algorithms in Wireless Sensor Networks

Due to the limited communication range of individual sensor nodes, data in WSNs often needs to be routed through multiple hops to reach the base station or sink node. Multi-hop routing protocols are used to efficiently route data across the network, ensuring that energy consumption is minimized and data is successfully transmitted.

Zeb et al. (2024) focus on energy-aware multi-hop routing protocols that consider the energy levels of the nodes when selecting the optimal paths for data

transmission. Their research emphasizes the importance of balancing the energy consumption across nodes, which helps extend the network's operational lifetime. These protocols are particularly advantageous in large-scale WSNs where direct communication from source nodes to the base station is not always feasible.

3.8 Multi-Hop Routing Algorithms Using Machine Learning

Machine learning (ML) techniques have been increasingly applied to improve multi-hop routing in WSNs. By allowing nodes to make decisions based on real-time data, ML algorithms can dynamically adjust routing decisions to optimize energy consumption and enhance network performance.

Bouzid et al. (2020) present a reinforcement learning-based (RL) routing protocol that adapts to both the energy states of the nodes and the topology of the network [108]. This protocol uses RL to determine the best routing paths, minimizing energy consumption while ensuring efficient data delivery. The ability of RL to learn optimal routing decisions in real-time makes it a promising solution for energy-aware multi-hop routing in WSNs.

3.9 Reinforcement Learning in Multi-Hop Routing

Reinforcement Learning (RL) has proven to be an effective tool for optimizing multi-hop routing in energy-harvesting WSNs. By using RL, sensor nodes can make informed decisions based on the network's current state and their own energy levels, which helps in selecting the most energy-efficient routing paths.

Zhao et al. (2020) apply Q-learning, a type of RL, to optimize the routing paths in WSNs with energy harvesting capabilities and dual alternative batteries [17].

This approach ensures that the network can maintain a steady power supply by distributing energy use between the two batteries.

The use of RL allows the network to adapt its routing strategy dynamically to avoid premature energy depletion.

Zhang et al. (2021) further extend this concept by incorporating Deep Reinforcement Learning (DRL) into multi-hop routing [109].

DRL combines the strengths of deep learning and RL, allowing nodes to process high-dimensional data and make more complex routing decisions.

Their approach enables the network to learn optimal routing strategies in real-time, even in dynamic and large-scale environments, improving the overall efficiency and robustness of the network.

3.10 Conclusion

Energy harvesting, particularly solar energy harvesting, offers a sustainable solution to the energy limitations of WSNs. However, efficient energy management is crucial to ensure the long-term operation of these networks. Multi-hop routing protocols are essential for ensuring that data is transmitted across the network in an energy-efficient manner. The integration of machine learning, especially reinforcement learning, has significantly improved the performance of multi-hop routing by enabling nodes to adapt their routing decisions based on real-time network conditions.

The work of Sah and Amgoth (2020), Dvir et al. (2024), and others illustrates the importance of combining energy harvesting with intelligent routing algorithms. As WSNs become more complex, the integration of machine learning techniques, particularly reinforcement learning, will continue to enhance the adaptability, efficiency, and sustainability of these networks.

3.10.1 Key Insights from the Literature

i. Energy-Centric Design: Studies emphasize that trade-offs in EH-WSNs revolve around energy availability, with techniques aiming to maintain neutrality while optimizing other metrics.

TABLE 3.1: Comparison of RL-Based Routing Protocols and Strategies for EH-WSNs

Study	Objective	RL Method	Focus	EH	Key Findings	Limitations
Mutombo *et al.* [100]	Energy-efficient routing	Q-learning	Routing (EER-RL)	Yes	Balances energy, extends lifetime via optimal paths	High comp. overhead; no real EH validation
Mutombo *et al.* [52]	Energy balancing	Q-learning	Routing (EBR-RL)	No	Maximizes efficiency with residual energy, hops	Limited scalability; uniform EH assumed
Al-Tous & Barhumi [103]	Delay minimization	SARSA	Routing	Yes	Reduces delay in multi-hop EH-WSNs	Delay-focused; simulation-only
Shresthamali *et al.* [110]	Adaptive power mgmt	SARSA(λ)	Power mgmt	Yes	Adapts power to EH, improves sustainability	Single-node; no routing integration
Shresthamali *et al.* [111]	Multi-obj. scheduling	RL (unspec.)	Res. scheduling	Partial	Balances energy, tasks in IoT	Limited EH focus; vague RL details
Kaur & Aulakh [51]	Energy-efficient clustering	RL (unspec.)	Clustering	No	Enhances lifetime via clustering	Not EH-specific; vague RL details
Ding *et al.* [101]	ML-routing overview	Various RL	Survey	Partial	RL effective for WSN routing	Broad scope; lacks EH depth
Godfrey *et al.* [102]	Energy-efficient routing	Q-learning (DOS)	Routing (DOS-RL)	Yes	Adapts to changes with dynamic obj.	Complex trade-offs; simulation-only
Barat *et al.* [104]	Coop. energy mgmt	Q-learning (MARL)	Energy mgmt	Yes	Improves throughput via coop. learning	High comm. overhead; small-scale

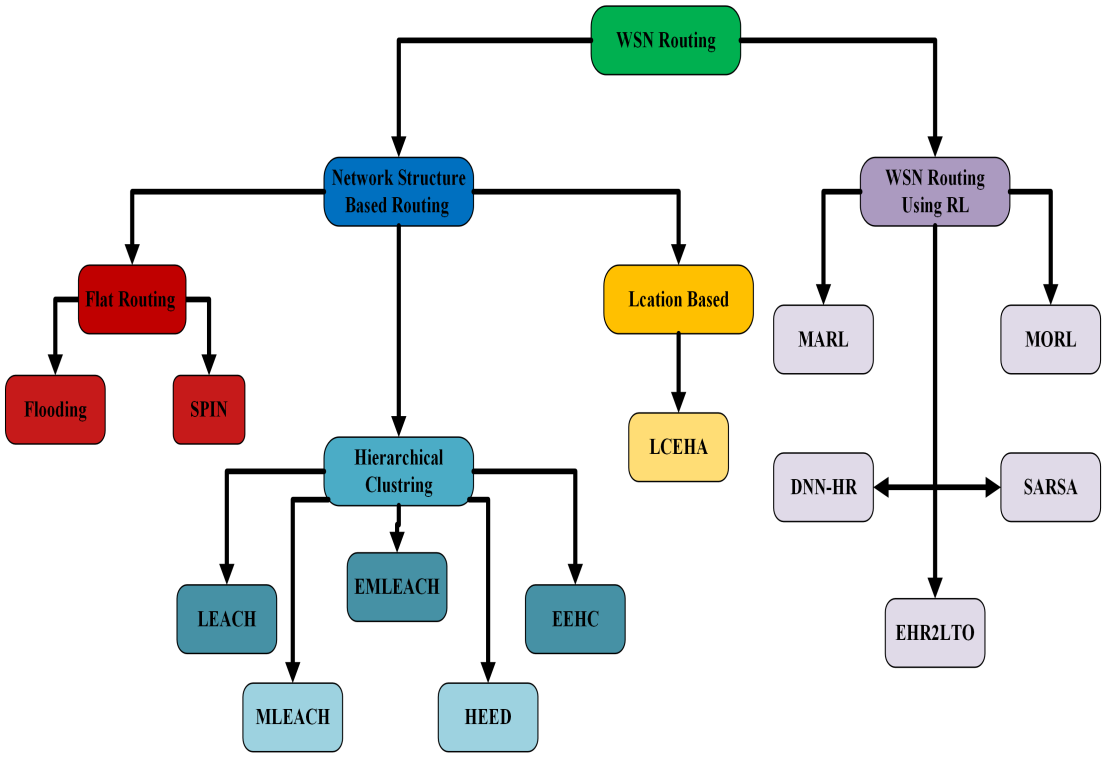


FIGURE 3.3: Taxonomy of WSN Routing Protocols

ii. Mathematical Modeling: Models provide a quantitative framework, with time-dependent metrics (e.g., $E_{harv}(t)$) often integrated over intervals, and some (e.g., scalability) requiring simulation.

iii. Practical Application: These findings inform protocol design, hardware selection, and deployment strategies for applications like environmental monitoring.

TABLE 3.2: Comparison of Metrics, Trade-offs, Techniques, and Mathematical Models in EH-WSNs

Metric	**Trade-offs**	**Techniques to Balance**	**Mathematical Model**
Energy Harvesting Efficiency	High efficiency increases hardware cost/portability issues	Adaptive circuits (e.g., MPPT); optimize harvester size	$\eta = \frac{P_{harv}}{P_{amb}}$
Energy Neutrality	May lower throughput or increase latency	Dynamic duty cycling; energy-aware scheduling	$E_{harv}(t) \geq E_{cons}(t)$
Network Lifetime	Perpetual operation may reduce data rate/coverage	Energy storage buffers; redundant nodes	$T_{life} = \frac{E_{total}}{\sum P_{avg}}$
Latency	Low latency increases energy use, risking neutrality	Prioritize critical packets; energy-aware multi-hop routing	$L = \sum_{i=1}^{h}(T_{proc,i} + T_{trans,i})$
Throughput	High throughput drains energy, reducing lifetime	Adaptive data aggregation; compression	$S = \frac{N_{succ} \cdot B}{T}$
Packet Delivery Ratio	High PDR increases energy via retransmissions	Error correction codes (e.g., FEC); efficient retransmission	$PDR = \frac{N_{deliv}}{N_{sent}}$
Energy Consumption/Packet	Low energy may reduce quality or increase latency	Optimize power control; low-power protocols (e.g., BLE)	$E_{pkt} = E_{sense} + E_{proc} + E_{tx}$
Coverage	Full coverage increases energy; sparse coverage risks event misses	Dynamic node activation; energy-aware clustering	$C = \frac{A_{cov}}{A_{total}}$
Reliability	High reliability raises energy/cost via redundancy	Fault-tolerant routing; periodic health checks	$R = 1 - P_{fail}$
Scalability	Large scale causes congestion/interference	Distributed protocols (e.g., LEACH); adaptive channel allocation	$S_c = f(N, P_{int})$
Data Accuracy	High accuracy increases energy via frequent sampling	Adaptive sampling; predictive algorithms	$A = 1 - \frac{\lvert D_{meas} - D_{true} \rvert}{D_{true}}$
Energy Storage Efficiency	High capacity increases cost/size; low capacity risks outages	Hybrid storage; energy-aware management	$\eta_{stor} = \frac{E_{stor}}{E_{harv}}$
Robustness to Variability	Robustness may reduce throughput in low-energy periods	Predictive models; opportunistic transmission	$R_v = \frac{\sigma_{perf}}{\sigma_{env}}$

3.10.2 Summary of Key Literature Findings

i. Most works such as in table 3.1 focus on hierarchical clustering for reducing energy consumption but assume static nodes and non-harvesting energy sources.

ii. Recent approaches integrating Reinforcement Learning (RL) (e.g., [17, 112]) improve dynamic adaptability but lack realistic energy harvesting.

iii. Dynamic reward weight adjustment can improve the performance.

iv. EH-aware routing protocols (e.g., [103]) address partial aspects of EH modeling but often neglect reinforcement driven long term optimization.

v. Few works combine distributed RL, dynamic clustering, and energy harvesting based stochastic modeling simultaneously, leaving a gap for multi agent energy harvesting WSN protocols.

3.10.3 Identified Research Gaps

To clearly illustrate the existing gaps, Table 3.1 summarizes major research findings and their corresponding limitations.

3.10.4 Narrative Gap Discussion

From the synthesized findings, it is evident that:

i. Existing clustering and routing mechanisms perform well in static or battery-powered WSNs but fail under fluctuating energy conditions caused by harvesting dynamics.

ii. Energy harvesting has been modeled simplistically or deterministically in most prior works, ignoring stochastic EH generation and battery constraints.

iii. Centralized learning-based models are not scalable or fault-tolerant in distributed, large-scale EH-WSNs.

iv. Few studies integrate dynamic clustering, multi-agent RL and energy harvesting aware optimization together.

3.10.5 Motivation for the Proposed Model

To address these limitations, this thesis proposes an enhanced Reinforcement Learning-based Energy Efficient Routing framework that:

i. Integrates **stochastic EH modeling** and realistic Energy Harvesting.

ii. Adopts **distributed RL (D-SARSA)** for scalable and cooperative learning.

iii. Introduces **dynamic weight adjustment** to adapt to varying energy and traffic conditions.

iv. Evaluates **multiple performance metrics** including energy consumption, energy balance, and network lifetime.

These identified gaps form the foundation and motivation for the development of the proposed Energy Harvesting Reinforcement Learning-based Routing Protocol (EH-RL-AEBRP) discussed in the next chapter.

Chapter 4

Proposed Model

4.1 Introduction

Wireless Sensor Networks (WSNs) consist of a large number of low-power sensor nodes collaboratively monitoring physical or environmental phenomena. A fundamental limitation of WSNs is the restricted energy availability of sensor nodes, which directly constrains network lifetime and reliability.

Energy harvesting (EH) has emerged as a promising solution by enabling sensor nodes to replenish energy from ambient sources such as solar, vibration, or thermal gradients. However, the stochastic and intermittent nature of harvested energy introduces significant challenges in routing protocol design.

Conventional routing protocols and many reinforcement learning (RL)-based approaches assume static battery depletion and permanent node death.

Such assumptions are unsuitable for EH-enabled WSNs, where nodes may temporarily exhaust their energy and later revive once sufficient energy is harvested.

Furthermore, static reward formulations fail to adapt routing behaviour to dynamically changing energy availability, network connectivity, and traffic load.

This chapter proposes an Adaptive Energy Balancing in Wireless Sensor Networks via Q-Learning: An Integrated Energy Harvesting and Adaptive Routing Protocol (EH-RL-AEBRP).

The protocol integrates realistic energy harvesting, distributed Q-learning, adaptive reward weight adjustment, continuous traffic generation, and node death-revival dynamics.

The proposed **EH-RL-AEBRP** protocol significantly enhances the baseline **RL-EBRP** by introducing:

1. A realistic seasonal-diurnal energy harvesting model enabling node revival,
2. An adaptive reward weight adjustment mechanism based on real-time network energy statistics ($\bar{E}(t)$, $\sigma_E^2(t)$, $\phi_A(t)$, $\phi_H(t)$),
3. Traffic-aware routing decisions that reduce unnecessary transmissions,
4. A unified framework that performs effectively in both harvesting-enabled and conventional battery-powered WSNs.

Unlike the baseline's fixed weights and continuous full-load traffic, EH-RL-AEBRP dynamically balances energy consumption, hop efficiency, and transmission distance, resulting in better energy distribution, longer network lifetime, and higher packet delivery performance under dynamic and intermittent energy availability.

The design strictly follows the implemented simulation framework and enables fair evaluation under multiple energy scenarios.

4.1.1 Protocol Objectives

The EH-RL-AEBRP protocol aims to achieve the following objectives:

i. Maximize Network Lifetime: Distribute energy consumption evenly across nodes to extend the operational duration of the WSN.

ii. Minimize Energy Depletion: Utilize adaptive multi-hop routing to reduce energy expenditure per node.

iii. Leverage Energy Harvesting: Incorporate energy harvesting to sustain node operation and enhance network longevity.

iv. Ensure Robust Path Selection: Maintain reliable routing paths under dynamic network conditions, such as node failures or energy fluctuations.

4.2 System Architecture

The Wireless Sensor Network (WSN) for the EH-RL-AEBRP protocol comprises N sensor nodes, each uniquely indexed $i \in \{1, 2, \ldots, N\}$, positioned at coordinates (x_i, y_i), and initialized with energy E_0 J. Each node has a communication range of R_c, enabling it to connect with neighbors within a Euclidean distance, provided both nodes have positive energy. A central sink node, located at $(x_{\text{sink}}, y_{\text{sink}})$, serves as the data collection point with unlimited energy. The nodes form a dynamic multi-hop network topology, where routing paths adapt to node energy levels and network conditions. Additionally, each node can realistically harvest energy based on a modified diurnal-seasonal harvesting model, with rate $H_{\max}$ J per round.

To ensure loop-free routing and monotonic progress toward the sink, a Breadth-First Search (BFS) algorithm computes the minimum hop count from every alive node to the sink. This hop information is dynamically updated whenever node death or revival occurs.

4.2.1 Key Components

The EH-RL-AEBRP protocol integrates several core components to achieve energy-efficient routing. Central to the protocol is a Q-learning mechanism, where each node i maintains a Q-table for neighbors j, representing the expected long-term reward of forwarding data to node j.

The Q-values are updated using the Bellman equation, with learning rate α, discount factor $\gamma \in [0, 1)$, and a reward R. The energy model accounts for transmission, reception, and data aggregation costs, with transmission energy following a free-space propagation model. Energy harvesting is modeled with a dynamic rate. Routing is achieved by selecting the next-hop neighbor, employing an ε-greedy policy to balance exploration and exploitation, thereby ensuring energy-efficient and reliable paths to the sink.

4.3 Mathematical Model

The mathematical model formalizes the Wireless Sensor Network (WSN), energy consumption, Q-learning, Reward function, Adaptive reward weight adjustment, and energy harvesting components of the Energy Harvesting Reinforcement Learning Adaptive Energy Balancing Routing Protocol (EH-RL-AEBRP).

4.3.1 Network Model and Assumptions

The WSN is modeled as a dynamic graph $G(t) = (V, E(t))$, where the vertex set V includes N sensor nodes and a sink node.

The edge set $E(t)$ comprises edges between nodes i and j within the communication range R, provided node i is alive.

Edges are removed as nodes deplete their energy, reflecting the network's dynamic nature.

The Euclidean distance between nodes i and j is given by Eq. (4.1):

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{4.1}$$

The distance from node i to the sink, positioned at $(x_{\text{sink}}, y_{\text{sink}})$, is given by Eq. (4.2):

$$d_{i,\text{sink}} = \sqrt{(x_i - x_{\text{sink}})^2 + (y_i - y_{\text{sink}})^2}. \tag{4.2}$$

The hop count from node i to the sink, assuming a uniform communication range, is approximated as in Eq. (4.3):

$$h_i \approx \left\lceil \frac{d_{i,\text{sink}}}{R} \right\rceil. \tag{4.3}$$

Each node i is characterized by its energy level $E_i(t)$, operational condition $c_i(t)$ (where $c_i(t) = 0$ if $E_i(t) \leq E_{\text{min}}$, indicating a dead node, and $c_i(t) = 1$ otherwise, indicating an alive node), and neighbor set given in Eq. (4.4):

$$\mathcal{N}_i(t) = \{j \in V \setminus \{i\} \mid d_{ij} \leq R \wedge c_j(t) = 1\}. \tag{4.4}$$

The neighbor set $\mathcal{N}_i(t)$ supports multi-hop routing and updates dynamically as nodes' energy levels change. The model assumes bidirectional links and a fixed sink at coordinates $(x_{\text{sink}}, y_{\text{sink}})$, consistent with the network's deployment area (e.g., 100×100 m^2). This structure is visually represented in Figure 4.1, which illustrates the sink's central position, node connectivity within R, and the isolation of dead nodes, with edges directed towards the sink to reflect typical data flow in the WSN.

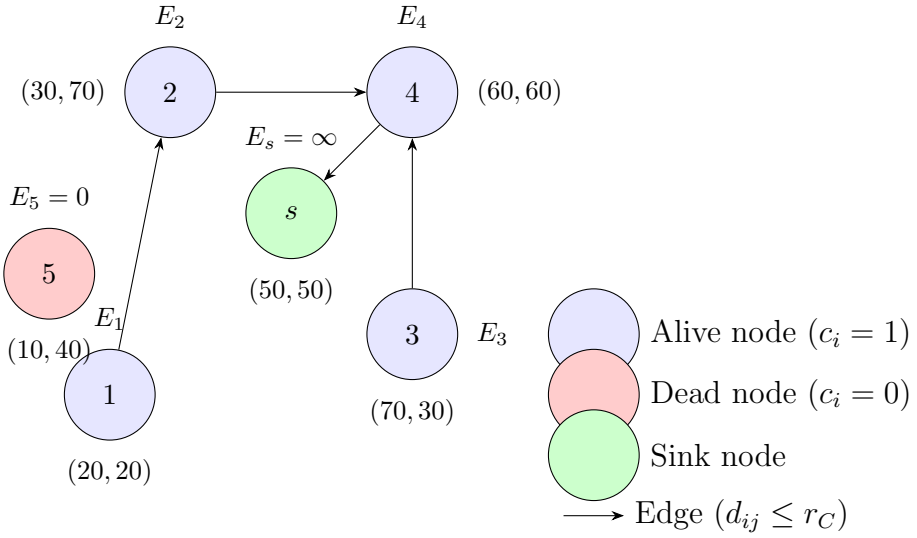


FIGURE 4.1: Network model of the WSN, depicted as a directed graph $G = (V, E)$. The sink node s is at $(x/2, y/2) = (50, 50)$, with sensor nodes at coordinates (x_i, y_i). Edges connect nodes within communication range r_C and are directed towards the sink to represent data flow, with dead nodes (e.g., node 5) shown in red and isolated due to $E_i = 0$.

The proposed EH-RL-AEBRP protocol is developed under the assumption that all sensor nodes possess global localization capability, enabling position awareness for routing and distance estimation. An acknowledgement (ACK) mechanism is available to ensure reliable packet delivery and transmission confirmation.

4.3.2 Energy Consumption Model

The energy consumption model adopts the first-order radio model [16], accounting for transmission, reception, and data aggregation, assuming no packet loss and a fixed packet size k bits. The energy required to transmit k bits over distance d is:

$$E_{\text{tx}}(k, d) = E_{\text{elec}} \cdot k + \epsilon_{\text{fs}} \cdot k \cdot d^2 \tag{4.5}$$

where E_{elec} represents the electronic circuitry energy, E_{DA} denotes the data aggregation energy, and E_{amp} is the transmit amplifier coefficient. The energy for receiving k bits is:

$$E_{\text{rx}}(k) = E_{\text{elec}} \cdot k \tag{4.6}$$

The energy for aggregating k bits, performed by intermediate nodes in multi-hop paths, is:

$$E_{\text{da}}(k) = E_{\text{da}} \cdot k \tag{4.7}$$

When node i transmits to node j, the energy updates are:

$$E_i \leftarrow E_i - E_{\text{tx}}(k, d_{ij}) \tag{4.8}$$

$$E_j \leftarrow E_j - E_{\text{rx}}(k) \tag{4.9}$$

If node j aggregates data before forwarding, its energy is further reduced:

$$E_j \leftarrow E_j - E_{\text{da}}(k) \tag{4.10}$$

For node i transmitting to node j with aggregation at j, the total energy consumption combines Equations (4.5), (4.6), and (4.7).

For direct transmission to the sink, only Equation (4.8) applies, with distance $d = d_{is}$ from Equation (4.2).

4.3.3 Energy Harvesting Model

Each node i can harvest energy with a seasonal and modified diurnal solar harvesting model, subject to a maximum energy harvest rate $H_{\max}$.

The energy level $E_i(t)$ is updated after a successful harvest as in Eq. (4.11):

$$E_i(t+1) = \min\left(E_{\max}, E_i(t) - E_{\text{consumed}}(t) + E_h(t)\right), \tag{4.11}$$

where $E_h(t)$ denotes the amount of energy harvested in round t, and $E_{\text{consumed}}(t)$ is the energy consumed by the node in that round.

4.3.3.1 Seasonal and Diurnal Solar Harvesting Model

A physically realistic solar energy harvesting model is employed, accurately reproducing both diurnal (day/night) and seasonal insolation variations observed at latitude ϕ. The instantaneous harvested energy per round is computed as given in Eq. (4.12):

$$E_h(t) = H_{\max} \cdot f_{\text{season}}(t) \cdot f_{\text{diurnal}}(t) \cdot (1 + \eta(t)), \tag{4.12}$$

where $H_{\max}$ is the maximum harvestable energy at noon in summer, $f_{\text{season}}(t)$ models the seasonal variations in maximum harvested energy output throughout the year, $f_{\text{diurnal}}(t)$ is the modified diurnal (day/night) profile, and $\eta(t)$ introduces $\pm 10\%$ random variability to model cloud cover and atmospheric fluctuations (with $|\eta(t)| \leq 0.1$).

4.3.3.2 Seasonal Component

The seasonal component $f_{\text{season}}(t)$ models the annual variation in solar insolation due to Earth's tilt and orbital eccentricity, which cause predictable changes in daily solar radiation received at any latitude.

For the simulation at Islamabad (Pakistan), this component is critical because Islamabad experiences a pronounced subtropical climate with hot summers and mild winters, resulting in a moderate variation in peak solar irradiance between solstices—less extreme than at higher latitudes but still significant for energy-neutral operation.

The mathematical formulation is a Fourier series fitted to NASA POWER meteorological data (1991–2020 hourly solar radiation averages) as given in Eq. (4.13): and The curve fitting results are shown in Figure 4.2.

$$\begin{aligned} f_{\text{season}}(\text{doy}) = 0.6467 &- 0.2455 \cos\left(\frac{2\pi \cdot \text{doy}}{365}\right) + 0.0519 \sin\left(\frac{2\pi \cdot \text{doy}}{365}\right) \\ &- 0.0469 \cos\left(\frac{4\pi \cdot \text{doy}}{365}\right) - 0.0544 \sin\left(\frac{4\pi \cdot \text{doy}}{365}\right) \\ &+ 0.0031 \cos\left(\frac{6\pi \cdot \text{doy}}{365}\right) + 0.0344 \sin\left(\frac{6\pi \cdot \text{doy}}{365}\right) \end{aligned} \tag{4.13}$$

where doy is the day of year (1 to 365). aligns the peak with the northern hemisphere summer solstice (June 21, doy ≈ 172). The amplitude of 0.325 is half the seasonal swing, producing a minimum of 0.35 at winter solstice (December 21, doy ≈ 355), with an offset of 0.675 ensuring the annual average is approximately 1 for long-term equilibrium.

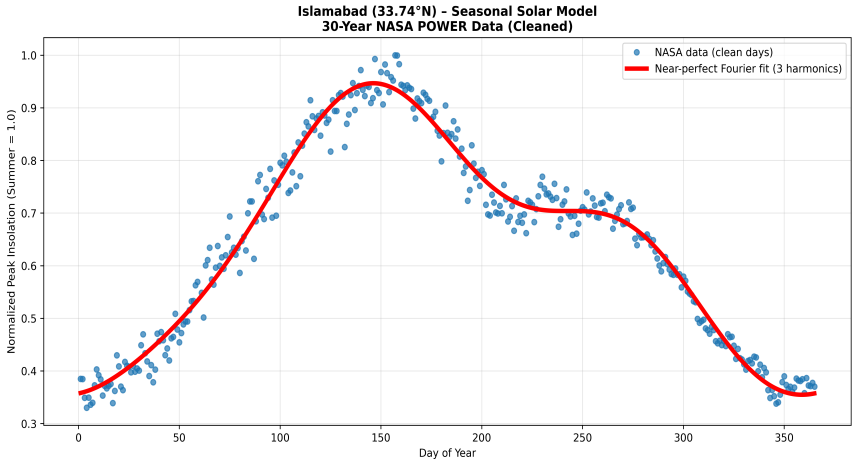


FIGURE 4.2: Fitted seasonal multiplier $f_{\text{season}}(\text{doy})$ using Fourier series for Islamabad (33.74°N), based on NASA POWER solar radiation data (1991–2020).

4.3.3.3 Modified Diurnal Component

The diurnal component accurately replicates the daily solar cycle and enforces strictly zero energy harvesting during nighttime a critical physical constraint absent in most literature. The instantaneous harvesting power follows a half-sine profile during daylight hours as given in Eq. (4.14):

$$f_{\text{diurnal}}(t) = \max\left(0, \sin\left(\frac{\pi(t \mod 1440)}{720}\right)\right), \tag{4.14}$$

where t is the current simulation round, 1440 rounds correspond to one full day (60-second time step), $t \mod 1440$ extracts the time-of-day in rounds, and the argument sweeps from 0 to π over 720 rounds (12 daylight hours). This formulation yields exactly zero harvesting from approximately 18:00 to 06:00 local solar time, a smooth sinusoidal rise and fall, peaking at solar noon (12:00).

It provides perfect alignment with the physical elevation of the sun above the horizon for a horizontally mounted solar panel.

Combined with the seasonal component , the complete harvesting model is obtained by multiplying the two components as previously shown in Eq. (4.12).

The half-sine diurnal model is analytically derived from the solar elevation angle for a flat-plate collector and is widely adopted in high-fidelity energy-harvesting literature [8]. Unlike averaged 24-hour models that incorrectly assume nighttime harvesting, our implementation enforces zero energy intake for 12 hours daily, significantly increasing the realism and difficulty of achieving energy-neutral operation, thereby strengthening the contribution of the proposed model.

Figure 4.3 shows the diurnal energy variation over a 10 days period, clearly illustrating the strict nighttime cutoff and smooth daytime profile.

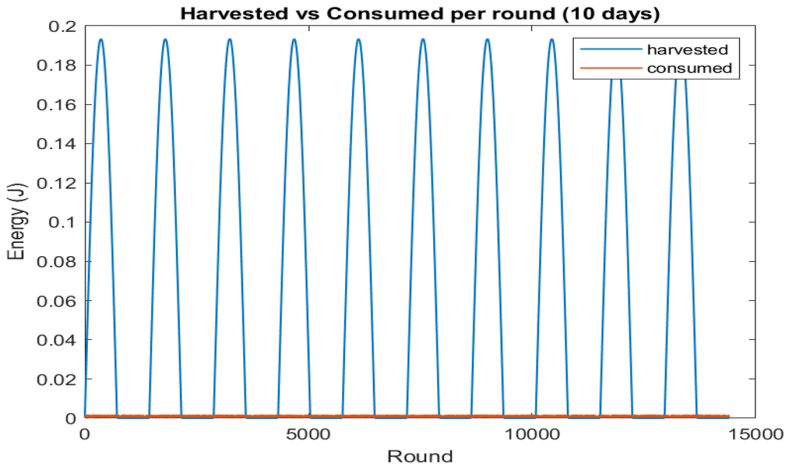


FIGURE 4.3: Diurnal component of the solar energy harvesting model over ten full days. The half-sine profile enforces zero harvesting during nighttime hours while providing a realistic peak at solar noon.

4.3.4 Node Death and Revival Mechanism

Unlike conventional WSN models, EH-RL-AEBRP explicitly supports node revival. A node i is declared dead when

$$E_i(t) \leq E_{\min}. \tag{4.15}$$

A dead node may revive if both conditions are satisfied:

$$E_i(t) \geq E_{\text{revival}} \quad \text{and} \quad (t - t_{\text{death}}) \geq T_{\text{cooldown}}. \tag{4.16}$$

Upon revival, the node re-enters routing operations, BFS hop counts are recomputed, and Q-values are heuristically reinitialized to ensure learning stability as shown in figure 4.4 .

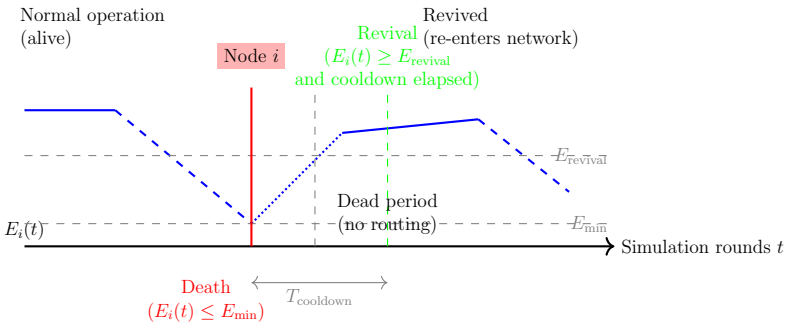


FIGURE 4.4: Illustration of the node death and revival mechanism in EH-RL-AEBRP. A node dies when its energy falls below E_{min}, remains inactive during a cooldown period T_{cooldown}, and revives once its harvested energy reaches E_{revival}. Upon revival, BFS hop counts and Q-values are updated.

4.3.5 Traffic Model

The network operates under a continuous full-load traffic model. At each simulation round, every alive node generates exactly one data packet destined for the sink:

$$\forall i \in \mathcal{A}(t), \quad i \text{ generates and transmits one packet per round,} \tag{4.17}$$

where $\mathcal{A}(t) = \{i \mid c_i(t) = 1\}$ denotes the set of alive nodes at round t.

This traffic model imposes sustained stress on the routing protocol and enables rigorous evaluation of energy sustainability and routing robustness.

4.4 Reinforcement Learning Routing Framework

4.4.1 State, Action, and Learning Structure

Each sensor node functions as an autonomous RL agent. The state is implicitly represented by the current node index i (i.e., the agent is always in "state" i when

making a forwarding decision at node i). The action a corresponds to selecting one of the alive neighboring nodes $j \in \mathcal{N}_i(t)$ as the next hop.

A global Q-table $Q(i, j)$ stores the expected long-term utility (discounted cumulative reward) of forwarding packets from node i to node j.

The Q-values are updated using the standard Q-learning Bellman equation:

$$Q(i,j) \leftarrow Q(i,j) + \alpha \left[R(i,j,t) + \gamma \max_{k \in \mathcal{N}_j(t)} Q(j,k) - Q(i,j) \right], \tag{4.18}$$

where $\alpha \in (0, 1]$ is the learning rate, $\gamma \in [0, 1)$ is the discount factor, $R(i, j, t)$ is the immediate reward received after taking action j from node i at round t, and $\max_{k \in \mathcal{N}_j(t)} Q(j, k)$ represents the estimated future utility from the best possible action at the next hop j. Routing decisions follow an ε-greedy policy to balance exploration and exploitation:

- With probability $1 - \varepsilon$: select the greedy next hop

$$j^* = \arg \max_{j \in \mathcal{N}_i(t)} Q(i,j).$$

- With probability ε: select a next hop uniformly at random from the alive neighbors $\mathcal{N}_i(t)$.

This distributed per-node Q-learning approach shown in figure 4.5 allows each node to independently learn energy-aware and robust forwarding policies without requiring global network state information, while the ε-greedy mechanism ensures sufficient exploration of alternative paths, especially important in dynamic EH-enabled networks with node death and revival.

4.4.2 Reward Function Design and Rationale

The reward function is the core mechanism through which the proposed EH-RL-AEBRP protocol translates network objectives into learnable feedback for routing decisions. In energy-harvesting wireless sensor networks, routing decisions must simultaneously satisfy multiple, often conflicting, objectives: conserving residual

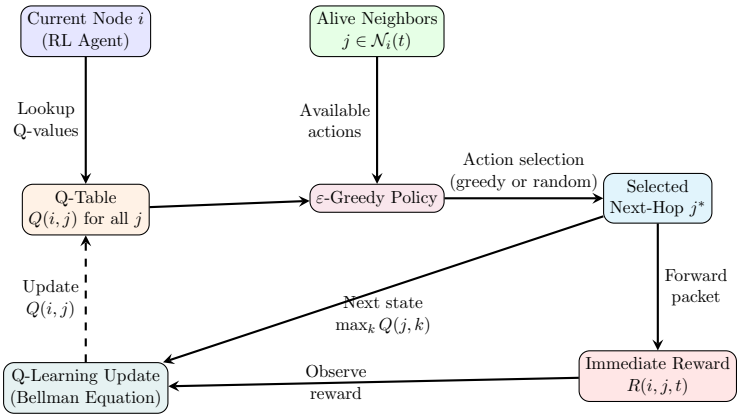


FIGURE 4.5: Reinforcement Learning framework in EH-RL-AEBRP. Each sensor node acts as an independent RL agent.

energy, ensuring progress toward the sink, maintaining short routing paths, and exploiting favorable energy harvesting conditions. A poorly designed reward function may lead to rapid node depletion, routing loops, or unstable learning behavior. To address these challenges, the proposed protocol employs a multi-objective composite reward function that integrates energy awareness, hop efficiency, and link quality, with dynamically adjusted weights that reflect the current network condition.

4.4.2.1 General Reward Structure

For a packet forwarded from node i to node j at round t, the immediate reward is defined as:

$$R(i,j,t) = w_E(t)\, R_E(j,t) - w_H(t)\, R_H(j,t) + w_D(t)\, R_D(i,j), \tag{4.19}$$

where:

- $R_E(j,t)$ is the energy-based reward component,
- $R_H(j,t)$ is the hop-count–based reward component,
- $R_D(i,j)$ is the distance-based reward component,
- $w_E(t)$, $w_H(t)$, $w_D(t)$ are adaptive, non-negative weights satisfying:

$$w_E(t) + w_H(t) + w_D(t) = 1. \tag{4.20}$$

This formulation allows the routing policy to adaptively emphasize different objectives depending on network state and energy availability as shown in figure 4.6.

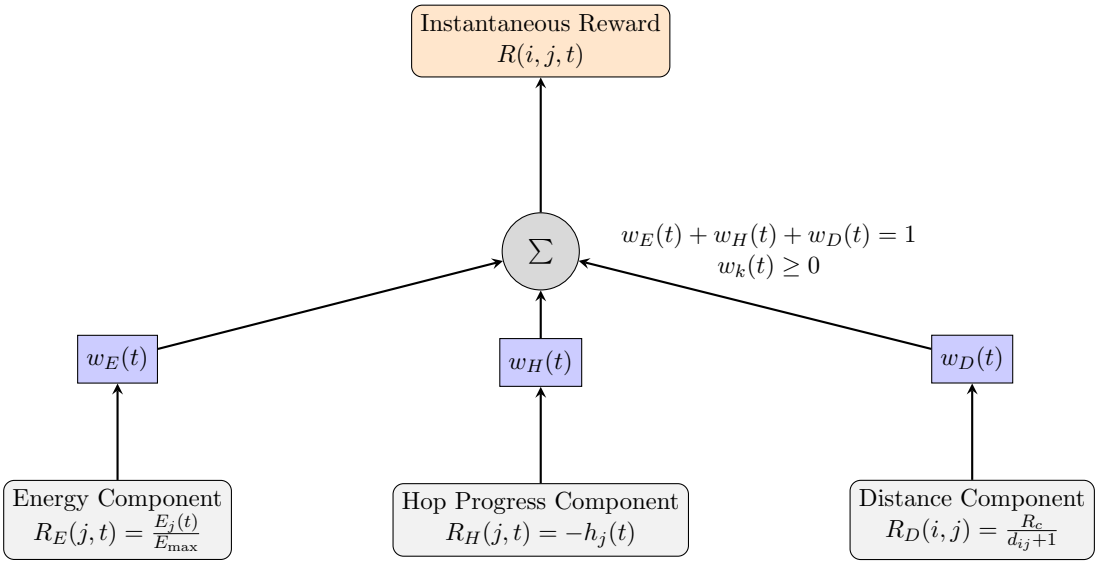


FIGURE 4.6: Structure of the multi-objective reward function in EH-RL-AEBRP. Three normalized components (energy-aware, hop progress, and distance-based link quality) are weighted and summed to produce the instantaneous reward signal used for Q-learning.

4.4.2.2 Energy-Aware Reward Component

The energy-aware component promotes the selection of next-hop nodes with higher residual energy, thereby balancing energy consumption and preventing premature node failures. It is defined as:

$$R_E(j,t) = \frac{E_j(t)}{E_{\max}}, \tag{4.21}$$

where $E_j(t)$ is the residual energy of node j and $E_{\max}$ is the maximum battery capacity. This normalization ensures that:

$$0 \leq R_E(j,t) \leq 1,$$

making the reward scale-independent and numerically stable for learning. Selecting neighbors with higher energy reserves distributes forwarding load more evenly across the network, reduces energy variance, and significantly extends network lifetime particularly critical under continuous traffic conditions.

4.4.2.3 Hop-Count Based Reward Component

To ensure consistent progress toward the sink and prevent routing loops, a hop-count–based penalty is incorporated:

$$R_H(j,t) = \frac{-h_j(t)}{h_{\max}}, \tag{4.22}$$

where $h_j(t)$ denotes the minimum hop count from node j to the sink, obtained via BFS and $h_{\max}$ comes from network area.

This term explicitly discourages forwarding packets to nodes that are farther from the sink in terms of hop distance.Without hop-awareness, RL agents may favor energy-rich nodes that lead to detours or excessively long paths.

Incorporating hop count enforces directional routing while maintaining learning flexibility.

4.4.2.4 Distance-Based Reward Component

The distance-based component captures link quality and transmission efficiency:

$$R_D(i,j) = \frac{R_c}{d_{ij}+1}, \tag{4.23}$$

where d_{ij} is the Euclidean distance between nodes i and j, and R_c is the communication range.The addition of unity in the denominator prevents singularity and bounds the reward.

Shorter links require less transmission energy due to the quadratic path-loss model. This term indirectly minimizes transmission energy and reduces packet loss risk.

4.4.2.5 Complete Reward Expression

Substituting the components, the complete reward function becomes:

$$R(i,j,t) = w_E(t)\frac{E_j(t)}{E_{\max}} - w_H(t)\frac{h_j(t)}{h_{\max}} + w_D(t)\frac{R_c}{d_{ij}+1}. \tag{4.24}$$

4.4.2.6 Interaction with Energy Harvesting

Unlike traditional battery-powered WSNs, EH-WSNs exhibit time-varying energy availability.

The reward function implicitly accounts for this through:

1. The residual energy term $E_j(t)$,
2. The adaptive weights $w_E(t)$, which increase during energy scarcity,
3. Node revival mechanisms that reintroduce previously dead nodes into routing.

Thus, the reward function enables opportunistic routing behavior that exploits periods of abundant harvesting while maintaining survivability during scarcity.

4.4.2.7 Reward Boundedness and Learning Stability

Each reward component is bounded:

$$0 \leq R_E(j,t) \leq 1, \quad -1 \leq R_H(j,t) \leq 0, \quad 0 < R_D(i,j) \leq R_c.$$

Therefore, the total reward is bounded, ensuring numerical stability of Q-learning updates and preventing divergence.

4.4.2.8 Design-Advantages

The proposed reward function offers the following advantages:

- Multi-objective optimization without scalarization bias,
- Dynamic prioritization through adaptive weights,
- Loop-free routing via hop-based penalties,
- Energy sustainability via residual energy awareness,
- Compatibility with EH dynamics and node revival.

4.4.2.9 Practical Interpretation

From a networking perspective, the reward function can be interpreted as follows:

A forwarding decision is preferred if it selects a neighbor that is energetically healthy, closer to the sink in terms of hops, and reachable via a short, energy-efficient link while dynamically adapting these priorities to the current network state.

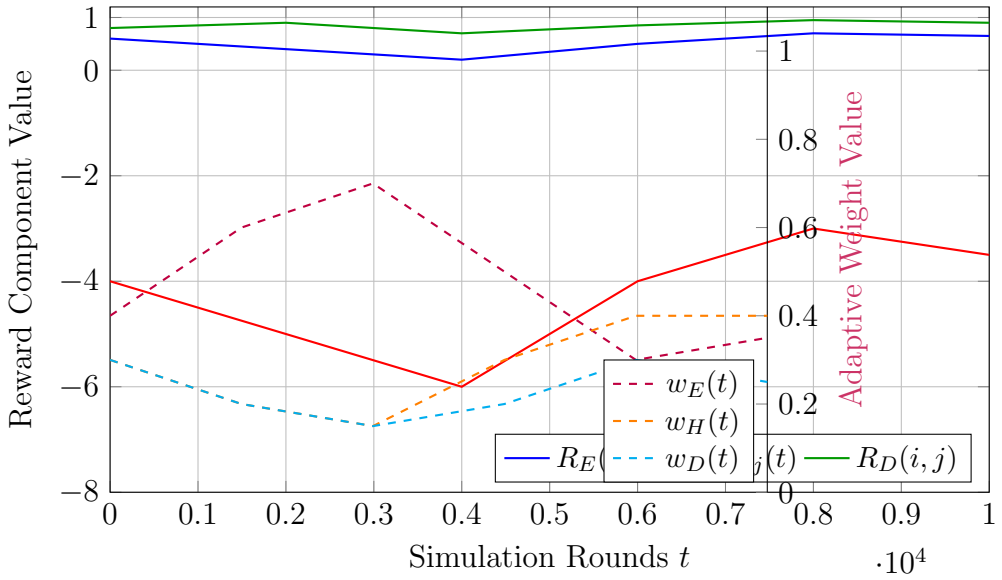


FIGURE 4.7: Contribution of individual reward components and evolution of adaptive weights over simulation time in a representative scenario. The primary (left) y-axis shows the reward components $R_E(j,t)$, $-h_j(t)$, and $R_D(i,j)$; the secondary (right) y-axis shows the adaptive weights $w_E(t)$, $w_H(t)$, and $w_D(t)$. The legend is placed at the bottom-right corner for clarity.

Figure 4.7 illustrates the contribution of individual reward components and their adaptive weights over simulation time.

4.4.3 Adaptive Reward Weight Adjustment Mechanism

In energy-harvesting wireless sensor networks, static reward weighting fails to capture long-term variations in network conditions such as fluctuating harvested energy, heterogeneous residual energy distribution, and dynamic node availability.

To address this limitation, the proposed EH-RL-AEBRP protocol employs an adaptive reward weight adjustment mechanism that continuously tunes the relative

importance of energy awareness, routing efficiency, and transmission distance in the reinforcement learning reward function.

This mechanism allows the routing policy to autonomously shift its behavior between energy conservation and latency reduction, thereby enhancing network sustainability and stability under diverse operating conditions.

4.4.3.1 Network-Level Energy Statistics

Let $\mathcal{A}(t) \subseteq \{1, 2, \ldots, N\}$ denote the set of alive nodes at round t.

The average residual energy of the network is computed as:

$$\bar{E}(t) = \frac{1}{|\mathcal{A}(t)|} \sum_{i \in \mathcal{A}(t)} E_i(t). \tag{4.25}$$

To capture energy imbalance across the network, the normalized energy variance is defined as:

$$\sigma_E^2(t) = \frac{1}{|\mathcal{A}(t)|} \sum_{i \in \mathcal{A}(t)} \left(\frac{E_i(t) - \bar{E}(t)}{E_{\max}} \right)^2. \tag{4.26}$$

A high value of $\sigma_E^2(t)$ indicates uneven energy distribution, which increases the risk of premature node depletion along heavily utilized routes.

4.4.3.2 Harvesting-Aware Sustainability Indicator

To incorporate energy harvesting dynamics, the average harvested energy over a sliding window of W rounds is calculated as:

$$\bar{H}(t) = \frac{1}{W} \sum_{\tau=t-W+1}^{t} H(\tau). \tag{4.27}$$

The window size W is selected proportional to the number of rounds per day, allowing the protocol to capture diurnal harvesting patterns without reacting to short-term fluctuations.

The normalized harvesting sustainability factor is then defined as:

$$\phi_H(t) = \frac{\bar{H}(t)}{H_{\max}}. \tag{4.28}$$

A low value of $\phi_H(t)$ signals energy scarcity, prompting the routing policy to prioritize energy preservation.

4.4.3.3 Network Survivability Indicator

To explicitly account for node availability, the fraction of alive nodes is computed as:

$$\phi_A(t) = \frac{|\mathcal{A}(t)|}{N}. \tag{4.29}$$

As $\phi_A(t)$ decreases, the routing policy is encouraged to reduce hop count and avoid excessive route elongation, which could otherwise accelerate network fragmentation.

4.4.3.4 Base Weight Computation

Using the derived network indicators, three unnormalized base weights are computed for the reward components:

i. Energy Awareness Weight: This formulation increases the importance of energy-aware routing when the average energy is low or when energy imbalance is high.

$$\tilde{w}_E(t) = \alpha_E \left(1 - \frac{\bar{E}(t)}{E_{\max}}\right) + \beta_E \sigma_E^2(t). \tag{4.30}$$

ii. Hop Count Weight: This term emphasizes hop-count minimization when a significant portion of the network is unavailable.

$$\tilde{w}_H(t) = \alpha_H \left(1 - \phi_A(t)\right). \tag{4.31}$$

iii. Distance-Based Link Quality Weight: When harvested energy is scarce, the routing policy favors shorter transmission distances to reduce transmission

energy consumption.

$$\tilde{w}_D(t) = \alpha_D \left(1 - \phi_H(t)\right). \tag{4.32}$$

4.4.3.5 Weight Normalization via Softmax

To ensure numerical stability and maintain a convex combination of reward components, the base weights are normalized using the softmax function:

$$w_k(t) = \frac{\exp\left(\tilde{w}_k(t)\right)}{\sum_{l\in\{E,H,D\}} \exp\left(\tilde{w}_l(t)\right)} \quad \forall k \in \{E, H, D\}. \tag{4.33}$$

This normalization guarantees:

$$\sum_k w_k(t) = 1, \quad 0 \leq w_k(t) \leq 1.$$

4.4.3.6 Final Reward Function with Adaptive Weights

The resulting adaptive reward used in the Q-learning update is:

$$R(i, j, t) = w_E(t)\frac{E_j(t)}{E_{\max}} - w_H(t)h_j(t) + w_D(t)\frac{R_c}{d_{ij}+1}. \tag{4.34}$$

This formulation enables the protocol to dynamically balance energy efficiency, routing progress, and transmission reliability.

4.4.3.7 Stability and Convergence Considerations

To prevent oscillatory behavior, the adaptive weights are updated at a slower timescale than Q-value updates.

Furthermore, sliding-window averaging smooths stochastic harvesting variations, ensuring gradual and stable policy adaptation.

Figure 4.8 illustrates the temporal evolution of the adaptive reward weights under different energy scenarios, demonstrating the protocol's ability to autonomously shift routing priorities in response to changing network conditions.

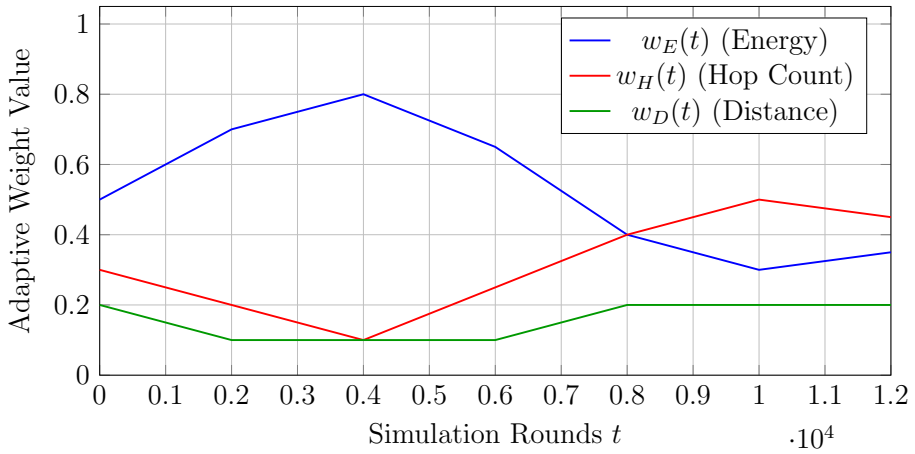


FIGURE 4.8: Temporal evolution of adaptive reward weights under a representative Energy-Scarce scenario followed by recovery. The protocol increases $w_E(t)$ during low energy periods, raises $w_H(t)$ when node deaths occur, and modestly elevates $w_D(t)$ during prolonged harvesting scarcity.

4.4.3.8 Discussion

The proposed adaptive reward weight adjustment mechanism enables EH-RL-AEBRP to:

- Avoid premature energy depletion,
- Mitigate route overuse,
- Maintain connectivity during energy scarcity,
- Exploit abundant harvesting periods to reduce latency.

By integrating network-level energy statistics and harvesting dynamics directly into the reward structure, the protocol achieves superior long-term sustainability compared to static-weight and non-harvesting routing approaches.

4.5 Algorithm and Flowchart

The RL-EBRP algorithm operates in distinct phases: initialization, round loop, transmission, energy harvesting, metrics update, and visualization/plotting.

Each phase is detailed with its own algorithm and flowchart below, followed by an overall flowchart illustrating the complete flow.

4.5.1 Initialization Phase

This phase sets up the network topology, initializes node parameters, computes essential routing information, and prepares efficient data structures for the distributed Q-learning process.

Algorithm 1 Initialization Phase of EH-RL-AEBRP

1: Deploy N sensor nodes randomly in the area, place sink at fixed position (e.g., $(50, 50)$)

2: Initialize each node i: $E_i(0) = E_{\text{init}}$, $c_i(0) = 1$ (alive)
3: Compute Euclidean distances d_{ij} for all node pairs i, j
4: Run BFS from sink to compute minimum hop counts h_i for all nodes
5: Initialize Q-table entries:
$Q(i, j) = \frac{1}{h_j}$ for all $j \in \mathcal{N}_i(0)$ ▷ Heuristic: closer nodes in hops are preferred

$Q(i, j) = 0$ otherwise (or omit non-neighbors)
6: Construct neighbor adjacency structure $\mathcal{N}_i(0)$ (alive nodes within range R_c)
7: Initialize sparse matrices/data structures for efficient storage of neighbors and Q-values
8: Compute initial adaptive reward weights $w_E(0), w_H(0), w_D(0)$ using network statistics

The initialization phase, as summarized in Algorithm 1 and illustrated in Figure 4.9, ensures that all nodes begin with consistent and meaningful prior knowledge (favoring shorter-hop paths)

while maintaining full compatibility with the distributed Q-learning and adaptive reward mechanisms described in Sections 4.4- 4.4.3

4.5.2 Round Loop Phase

This phase constitutes the core simulation loop of EH-RL-AEBRP, advancing the network state round by round until either all nodes are depleted or the predefined maximum number of rounds tot_rnd is reached.

Each round represents a fixed time step (e.g., 60 seconds).

During which traffic generation, routing, transmission, energy consumption, harvesting, node death/revival, and adaptive weight updates occur.

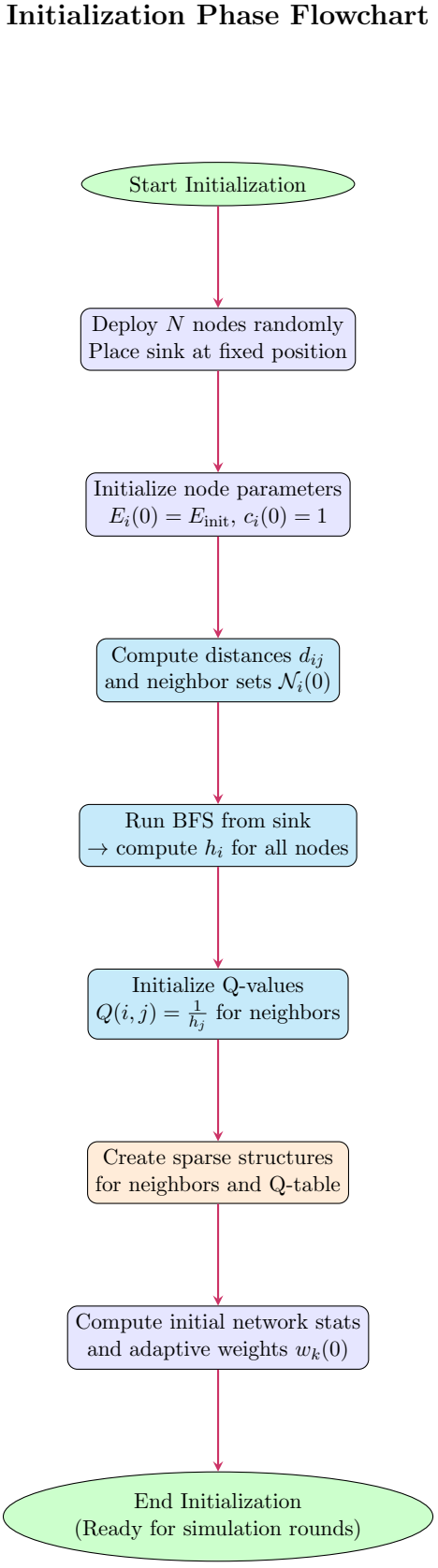


FIGURE 4.9: Flowchart of the Initialization Phase in EH-RL-AEBRP. This phase establishes the network topology, computes hop counts via BFS, heuristically initializes Q-values based on hop distance to the sink, and prepares data structures and initial adaptive weights before the main simulation loop begins.

Round Loop Phase Flowchart

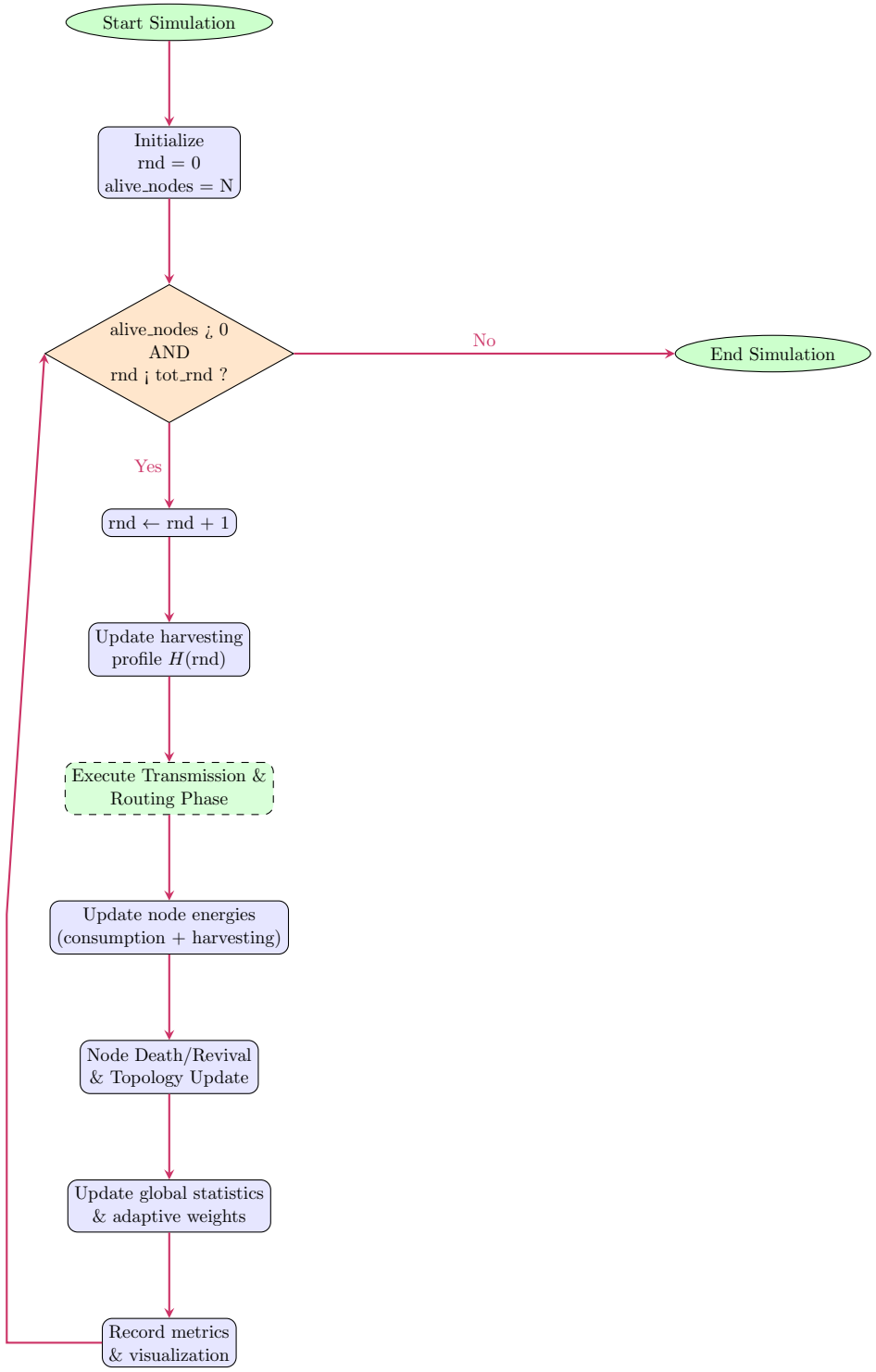


FIGURE 4.10: Flowchart of the Round Loop Phase in EH-RL-AEBRP. The loop continues until network collapse or the maximum number of rounds is reached, incorporating routing, energy updates, node state changes, adaptive weight adjustment, and metric collection in each iteration.

Algorithm 2 Round Loop Phase of EH-RL-AEBRP

Initialize rnd $\leftarrow 0$, alive_nodes $\leftarrow N$
while alive_nodes > 0 **and** rnd $<$ tot_rnd **do**
    rnd $\leftarrow$ rnd $+ 1$
    Update global energy harvesting profile $H(\text{rnd})$ (diurnal + seasonal model)
    Execute **Transmission and Routing Phase** (Algorithm 3)
    Update node energies due to transmission/reception (radio model)
    Execute **Energy Harvesting Phase**: $E_i(\text{rnd}) \leftarrow \min(E_{\max}, E_i(\text{rnd}-1) + H(\text{rnd}))$
    Check and execute **Node Death/Revival** (Eqs. (4.15)–(4.16))
    Update network topology (BFS recomputation if topology changed)
    Update global statistics: $\bar{E}(\text{rnd}), \sigma_E^2(\text{rnd}), \phi_H(\text{rnd}), \phi_A(\text{rnd})$
    Periodically update adaptive reward weights $w_k(\text{rnd})$
    Record performance metrics (alive nodes, packets delivered, etc.)
    Optional: Real-time visualization/plotting
    Update alive_nodes $\leftarrow |\mathcal{A}(\text{rnd})|$
end while

The round loop phase, formalized in Algorithm 2 and visualized in Figure 4.10, orchestrates the time evolution of the network. It integrates all previously described mechanisms distributed Q-learning routing , realistic energy harvesting, node death/revival dynamics, and adaptive reward weighting into a coherent simulation cycle that accurately models the long-term behavior of energy-harvesting WSNs under continuous traffic load.

4.5.3 Transmission Phase

This phase implements the core distributed routing and transmission logic of EH-RL-AEBRP. In each simulation round, every alive node generates one data packet and attempts to forward it toward the sink using the learned Q-values.

The phase incorporates epsilon greedy exploration, reward computation, Q-learning updates, energy consumption, and direct transmission when the sink is within range.

The transmission phase, as presented in Algorithm 3 and illustrated in Figure 4.11, forms the operational core of the protocol. It is invoked once per simulation round within the main loop (Algorithm 2) and realizes the distributed reinforcement learning-based routing behavior described in Sections 4.4.

Algorithm 3 Transmission and Routing Phase (executed each round)

Update neighbor sets $\mathcal{N}_i(t)$ and ensure Q-table reflects only alive neighbors
for each node $i = 1$ to N **in parallel** ▷ Distributed execution **do**
  if $c_i(t) = 0$ **or** $E_i(t) \leq E_{\min}$ **then**
    Skip (node dead)
    continue
  end if
  Generate one data packet at node i
  if $\mathcal{N}_i(t) = \emptyset$ **and** $d_{i,\text{sink}} > R_c$ **then**
    Drop packet ▷ Isolated node
    continue
  end if
  if $d_{i,\text{sink}} \leq R_c$ **then** ▷ Direct transmission possible
    Transmit packet directly to sink
    Deduct transmission energy $E_{\text{tx}}(i, \text{sink})$ from $E_i(t)$
    Increment successful delivery counter
    continue
  end if
  ▷ Select next hop using ε-greedy policy
  if random() ¡ ε **then**
    $j \leftarrow$ uniform random from $\mathcal{N}_i(t)$ ▷ Exploration
  else
    $j \leftarrow \arg\max_{k \in \mathcal{N}_i(t)} Q(i,k)$ ▷ Exploitation
  end if
  Forward packet from i to j
  Deduct $E_{\text{tx}}(i,j)$ from $E_i(t)$ and E_{rx} from $E_j(t)$
  Compute immediate reward $R(i,j,t)$ (Eq. (4.24))
  Update Q-value:
  $Q(i,j) \leftarrow Q(i,j) + \alpha\big[R(i,j,t) + \gamma \max_{k \in \mathcal{N}_j(t)} Q(j,k) - Q(i,j)\big]$
end for

4.5.4 Energy Harvesting and Node Status Update Phase

This phase updates node energies based on the realistic diurnal and seasonal harvesting model, caps battery levels, and manages node death and revival dynamics as discussed in Section 4.3.3.1 and Section 4.3.4.

The energy harvesting and node status update phase, formalized in Algorithm 4 and visualized in Figure 4.12, integrates the realistic solar harvesting model with the node death and revival mechanism.

It is executed after the transmission phase in each simulation round (Algorithm 2), ensuring accurate modelling of temporal energy dynamics and network resilience

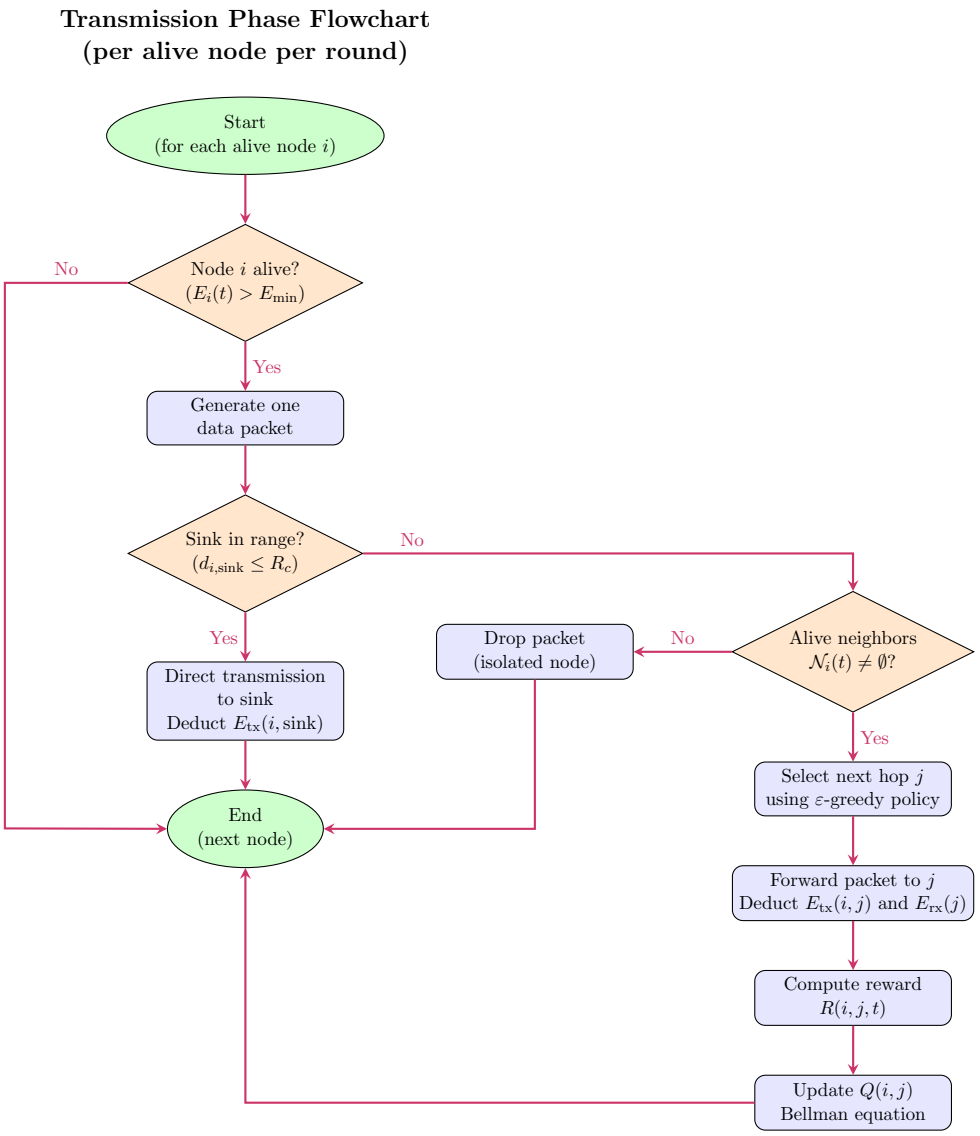


FIGURE 4.11: Flowchart of the Transmission and Routing Phase in EH-RL-AEBRP using right-angle (orthogonal) arrows for clearer routing paths.

in EH-enabled WSNs.

4.5.5 Metrics Update Phase

This phase computes and accumulates key performance metrics at the end of each simulation round to enable quantitative evaluation of network lifetime, energy efficiency, and operational node count.

The metrics update phase, formalized in Algorithm 5 and illustrated in Figure 4.13, is executed at the end of each simulation round within the main loop (Algorithm 2).

Algorithm 4 Energy Harvesting and Node Status Update Phase (executed each round)

for each node $i = 1$ to N **do**
  Compute harvested energy $E_h(i, \text{rnd})$ using diurnal + seasonal model
  if $c_i(\text{rnd} - 1) = 1$ **then** ▷ Node was alive last round
    $E_i(\text{rnd}) \leftarrow \min(E_{\max}, E_i(\text{rnd} - 1) + E_h(i, \text{rnd}))$
  end if
  Check node death:
  if $E_i(\text{rnd}) \leq E_{\min}$ **then**
    $c_i(\text{rnd}) \leftarrow 0$ ▷ Mark as dead
    Record death time $t_{\text{death},i} \leftarrow \text{rnd}$
    Update network topology and statistics
  end if
  Check node revival (if previously dead):
  if $c_i(\text{rnd} - 1) = 0$ **and** $E_i(\text{rnd}) \geq E_{\text{revival}}$ **and** $(\text{rnd} - t_{\text{death},i}) \geq T_{\text{cooldown}}$ **then**
    $c_i(\text{rnd}) \leftarrow 1$ ▷ Revive node
    Recompute BFS hop count h_i
    Heuristically reinitialize Q-values for node i
    Update network topology and statistics
  end if
end for

Algorithm 5 Metrics Update Phase (executed each round)

Compute number of dead nodes this round: $\text{dead\_rnd}(\text{rnd}) \leftarrow \sum_{i=1}^{N}(1 - c_i(\text{rnd}))$
Update cumulative dead-node-rounds: $\text{cum\_dead\_rnd}(\text{rnd}) \leftarrow \text{cum\_dead\_rnd}(\text{rnd} - 1) + \text{dead\_rnd}(\text{rnd})$
Compute total energy consumed this round: $\text{E\_consumed}(\text{rnd}) \leftarrow \sum_{i=1}^{N}(E_{\max} - E_i(\text{rnd}))$
Compute number of operational (alive) nodes: $\text{op\_nodes}(\text{rnd}) \leftarrow \sum_{i=1}^{N} c_i(\text{rnd})$
Record additional metrics as needed (e.g., packets delivered, average hop count, etc.)

It provides essential data for evaluating network lifetime (via cumulative dead-node-rounds), energy utilization, and connectivity over time.

4.6 Protocol Operation

This section describes the step-by-step operation of the EH-RL-AEBRP protocol, integrating the mathematical models, algorithms, and flowcharts presented earlier

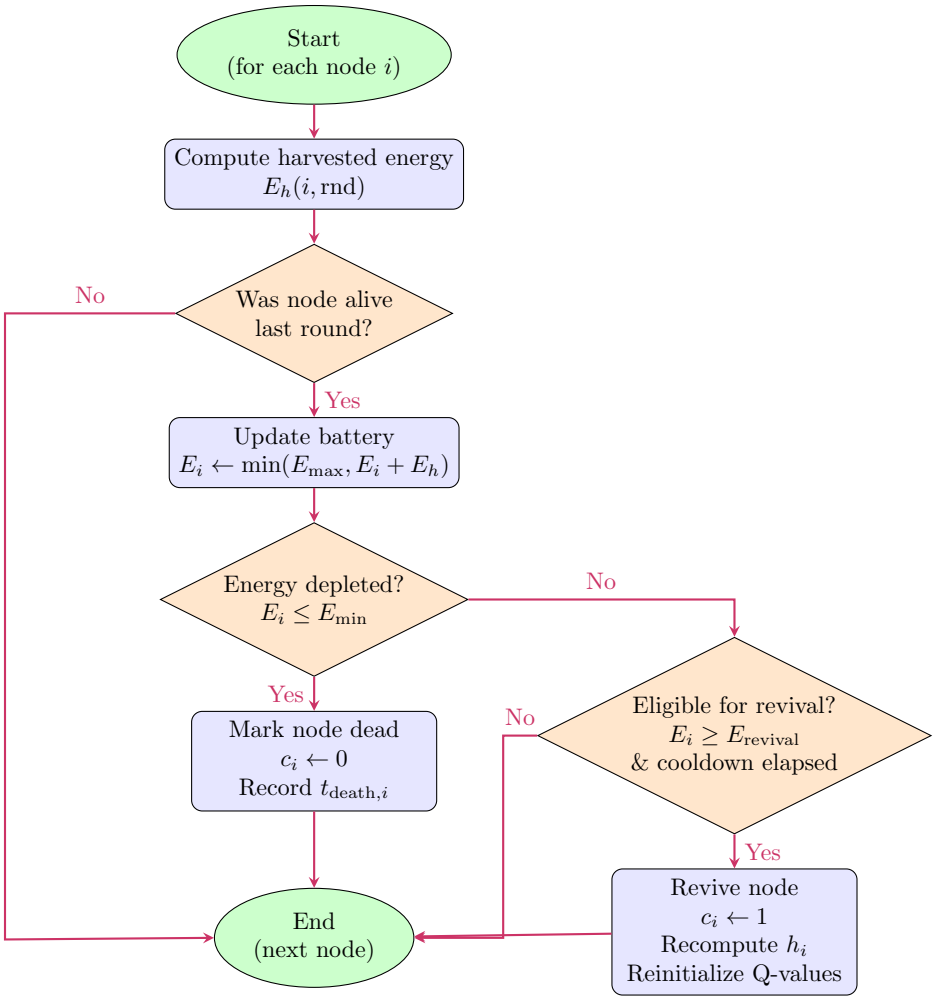


FIGURE 4.12: Flowchart of the Energy Harvesting and Node Status Update Phase in EH-RL-AEBRP.

to illustrate its complete functionality.

4.6.1 Network Setup

The protocol begins with the Initialization Phase (Algorithm 1, Figure 4.9), where the network is modeled as a graph $G = (V, E)$. Each node $i \in V$ is assigned an initial energy $E_i(0) = E_{\text{init}}$, and its position (x_i, y_i) is randomly generated within a 100×100 m^2 area. The sink is positioned at the center $(50, 50)$. Euclidean

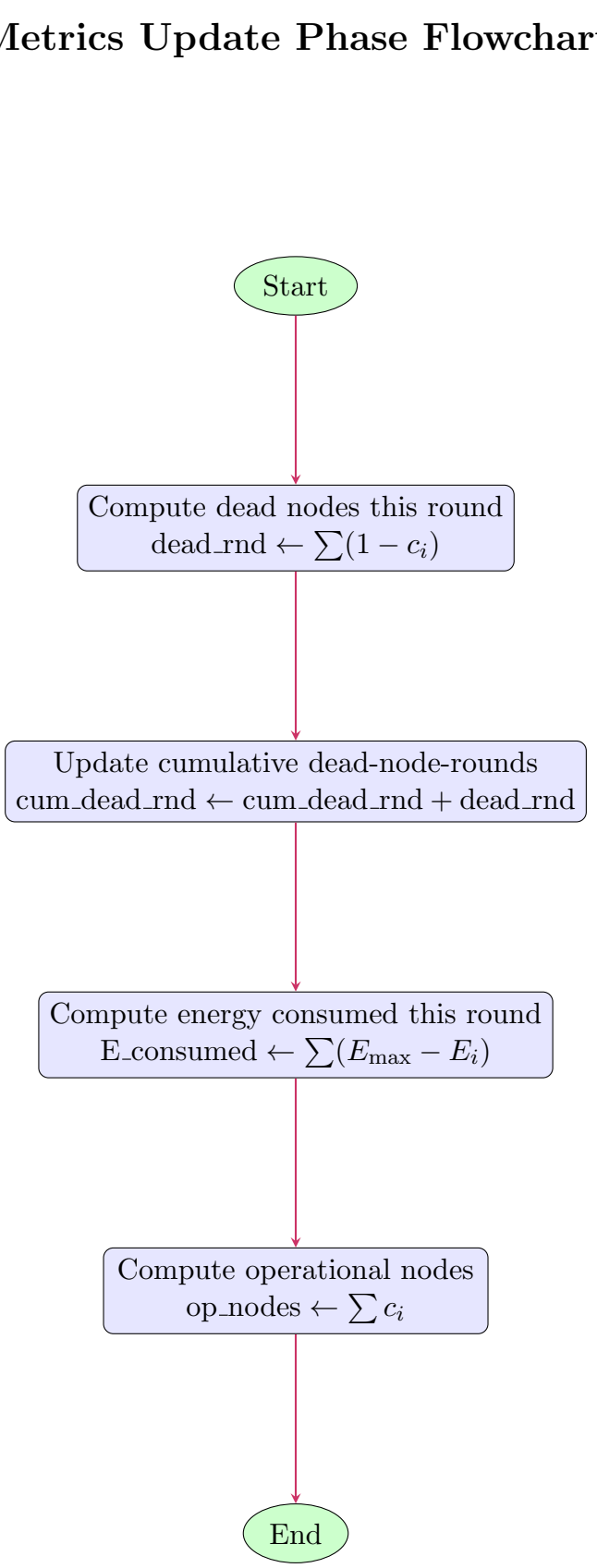


FIGURE 4.13: Flowchart of the Metrics Update Phase in EH-RL-AEBRP. This simple sequential phase runs after energy harvesting and node status updates to record key performance indicators for post-simulation analysis.

distances between nodes are computed using Eq. (4.1), and distances to the sink via Eq. (4.2). Minimum hop counts h_i are determined via BFS from the sink. Initial Q-values are heuristically set as $Q(i,j) = 1/h_j$ for neighbors $j \in \mathcal{N}_i(0)$, favoring shorter-hop paths. Neighbor sets $\mathcal{N}_i(0)$ are constructed according to Eq. (4.4), including only nodes within communication range R_c. Initial adaptive reward weights $w_k(0)$ are computed from network statistics.

4.6.2 Round Loop and Global Updates

The main simulation is governed by the Round Loop Phase (Algorithm 2, Figure 4.10). It executes for up to tot_rnd rounds or until no operational nodes remain. Each round updates the harvesting profile $H(\text{rnd})$ using the diurnal and seasonal model, then sequentially invokes the transmission, energy harvesting, node status update, adaptive weight adjustment, and metrics collection phases.

4.6.3 Transmission and Routing

In the Transmission and Routing Phase (Algorithm 3, Figure 4.11), every alive node generates one data packet. If the sink is within range ($d_{i,\text{sink}} \leq R_c$), direct transmission occurs with corresponding energy deduction. Otherwise, the next hop j is selected using the ε-greedy policy:

With probability ε a random alive neighbor is chosen (exploration), otherwise the neighbor maximizing $Q(i,k)$ is selected (exploitation). The packet is forwarded, transmission and reception energies are deducted, the immediate reward $R(i,j,t)$ is computed (Eq. (4.24)), and the Q-value is updated via the Bellman equation (Eq. (4.18)).

4.6.4 Energy Harvesting and Node Status Update

The Energy Harvesting and Node Status Update Phase (Algorithm 4, Figure 4.12) follows transmission. Harvested energy $E_h(i, \text{rnd})$ is added to each node's battery (capped at $E_{\max}$). Node death is checked: if $E_i(\text{rnd}) \leq E_{\min}$, the node is

marked dead ($c_i = 0$). Previously dead nodes are checked for revival eligibility: if $E_i(\text{rnd}) \geq E_{\text{revival}}$ and sufficient cooldown has elapsed, the node revives ($c_i = 1$), BFS hop counts are recomputed, and Q-values are heuristically reinitialized. Network topology and statistics are updated accordingly.

4.6.5 Metrics Update

The Metrics Update Phase (Algorithm 5, Figure 4.13) records key performance indicators at the end of each round: number of dead and operational nodes, energy consumption, and cumulative metrics for lifetime analysis.

4.6.6 Summary

The EH-RL-AEBRP protocol operates through a structured sequence of phases that integrate realistic energy harvesting, distributed Q-learning routing, adaptive reward weighting, and node death/revival dynamics. The mathematical models ensure accurate energy management and loop-free progress toward the sink, while the distributed execution enables emergent, robust, and energy-sustainable routing behavior under continuous traffic load and varying harvesting conditions.

Chapter 5

Results and Performance Evaluation

This chapter presents a comprehensive performance evaluation of the proposed Adaptive EH-RL-AEBRP protocol. The proposed scheme is compared against the baseline RL-EBRP protocol and evaluated under five distinct operational scenarios:

i. Baseline RL-EBRP without energy harvesting,

ii. Proposed EH-RL-AEBRP without energy harvesting (algorithmic gain only),

iii. Proposed EH-RL-AEBRP under energy-scarce harvesting conditions,

iv. Proposed EH-RL-AEBRP under energy-neutral harvesting conditions,

v. Proposed EH-RL-AEBRP under energy-abundant harvesting conditions.

This multi-scenario analysis enables a fair and insightful assessment of the adaptive reward weights, reinforcement learning enhancements, and the impact of different levels of energy harvesting availability.

5.1 Simulation Setup

All experiments were conducted using MATLAB. The network consists of 100 sensor nodes randomly deployed in a 100×100 m^2 square area, with the sink node fixed at the center $(50, 50)$ m and assumed to have unlimited energy supply. The common simulation parameters are listed in Table 5.1.

5.1.1 Scenario Specific: Energy Harvesting Conditions

The five evaluated scenarios differ primarily in the energy harvesting availability, which is controlled by scaling the average harvested energy per round ($\bar{H}$) relative to the average energy consumption per node per round ($\approx$ 0.008–0.012 J/round under typical routing load). The diurnal-seasonal solar harvesting model (Islamabad, Pakistan location) is used in all harvesting-enabled scenarios, with $H_{\max} =$ 0.015 J/round at peak and $\pm$10% random variability.

All other parameters (node deployment, radio model, RL hyperparameters, traffic pattern, etc.) remain identical across scenarios to ensure fair comparison. All EH-RL-AEBRP variants employ the realistic diurnal and seasonal solar harvesting model with scenario-specific maximum harvesting rates $H_{\max}$. The baseline RL-EBRP uses static reward weights ($p = 0.3, q = 0.7$) and lacks node revival support.

5.1.2 Performance Metrics

The performance of the proposed EH-RL-AEBRP protocol and the baseline RL-EBRP is evaluated using the following key metrics:

i. Network Lifetime: Time (in rounds) until first node death (HND) and last node death (90%ND).

ii. Packet Delivery Ratio (PDR): Fraction of generated packets successfully delivered to the sink.

iii. Energy Balance: Variance of residual energy across alive nodes over time.

iv. Jain's Fairness Index: It is a dimensionless fairness score bounded in $[0, 1]$.

TABLE 5.1: Simulation Parameters

Parameter	**Value**	**Description**
Network area	100×100 m^2	Deployment field
Number of sensor nodes (N)	100	Randomly deployed
Sink location	$(50, 50)$ m	Fixed central position
Initial energy per node (E_0)	0.5 J	Standard battery capacity
Maximum battery capacity ($E_{\max}$)	0.5 J	Upper bound after harvesting
Minimum operational energy ($E_{\min}$)	0.00058 J	Threshold for node death
Revival energy threshold (E_{revival})	$0.5 \times E_0$ J	Minimum energy to revive
Revival cooldown (T_{cooldown})	500 rounds	Minimum dead time before revival
Communication range (R_c)	30 m	Maximum transmission range
Data packet size (k)	4000 bits	Fixed packet length
Energy Consumption Model (First-Order Radio Model)		
Electronics energy (E_{elec})	50 nJ/bit	Circuitry energy
Data aggregation energy (E_{DA})	5 nJ/bit/message	Aggregation cost
Free-space amplifier (ϵ_{fs})	100 pJ/bit/m^2	Short-range transmission
Reinforcement Learning Parameters		
Learning rate (α)	0.1	Q-value update step size
Discount factor (γ)	0.9	Future reward importance
Exploration probability (ε)	0.2 (initial)	ε-greedy policy
ε decay	Linear to 0.05 over 5000 rounds	Annealing schedule
Traffic and Time Model		
Time step per round	60 seconds	Real-time equivalent
Traffic generation	1 packet/round/alive node	Continuous full-load (baseline)
Maximum simulation rounds	20,000	Stopping criterion

TABLE 5.2: Scenario Specific Energy Harvesting Levels

Scenario	**Energy Harvesting**	**Avg. Harvested Energy ($\bar{H}$)**	**Relation to Consumption**
1. Baseline RL-EBRP	Disabled	0 J/round	No replenishment
2. Proposed EH-RL-AEBRP			
(No EH)	Disabled	0 J/round	Algorithmic gain only
3. Energy-Scarce EH	Enabled	0.0005 J/round	$\bar{H} <$ consumption
4. Energy-Neutral EH	Enabled	0.002 J/round	$\bar{H} \approx$ consumption
5. Energy-Abundant EH	Enabled	0.004 J/round	$\bar{H} >$ consumption

5.1.3 Energy Balance

Energy balance is one of the most critical performance indicators in energy-harvesting wireless sensor networks (EH-WSNs), as uneven energy consumption often leads to hotspot formation, premature node failures, and network partitioning especially near the sink in converge cast traffic patterns.

The proposed EH-RL-AEBRP protocol aims to distribute forwarding load more evenly across nodes by dynamically prioritizing energy-rich neighbours through the adaptive reward mechanism. To quantify this property, we use the following two complementary metrics:

5.1.3.1 Normalized Energy Variance

This metric captures the spread of residual energy levels across all alive nodes at each round t:

$$\sigma_E^2(t) = \frac{1}{|\mathcal{A}(t)|} \sum_{i \in \mathcal{A}(t)} \left(\frac{E_i(t) - \bar{E}(t)}{E_{\max}} \right)^2, \tag{5.1}$$

where:

- $\mathcal{A}(t)$ is the set of alive nodes at round t,
- $E_i(t)$ is the residual energy of node i,

- $\bar{E}(t) = \frac{1}{|\mathcal{A}(t)|} \sum_{i \in \mathcal{A}(t)} E_i(t)$ is the average residual energy,
- $E_{\max}$ is the battery capacity (used for normalization).

Lower values of $\sigma_E^2(t)$ indicate better energy balance (more uniform residual energy distribution). In ideal conditions, $\sigma_E^2(t) \to 0$ when all alive nodes have nearly identical energy levels.

5.1.3.2 Jain's Fairness Index

It is a dimensionless fairness score bounded in $[0, 1]$, we also compute Jain's index:

$$J(t) = \frac{\left(\sum_{i \in \mathcal{A}(t)} E_i(t)\right)^2}{|\mathcal{A}(t)| \cdot \sum_{i \in \mathcal{A}(t)} E_i(t)^2}, \tag{5.2}$$

where $J(t) = 1$ represents perfect fairness (all nodes have equal energy) and $J(t) \to 1/|\mathcal{A}(t)|$ in the worst case (one node has all the energy).

5.2 Comparative Analysis of Baseline RL-EBRP and EH-RL-AEBRP (No Energy Harvesting)

This section presents a controlled comparative analysis between the baseline RL-EBRP protocol and the proposed EH-RL-AEBRP operating *without energy harvesting*. By disabling the harvesting module, both protocols rely solely on their initial battery energy, ensuring a fair and unbiased evaluation of routing intelligence and energy management strategies.

Although energy harvesting is disabled, EH-RL-AEBRP retains its adaptive reward mechanism, which dynamically balances residual energy, hop count, and transmission distance.

In contrast, RL-EBRP employs fixed or implicitly weighted reward components, making it less responsive to network dynamics.

5.2.1 Network Lifetime Comparison

Network lifetime is evaluated using the *Half Node Death (HND)* and *Last Node Death (LND)* metrics. The HND metric is particularly relevant in reinforcement learning based routing, as it reflects the protocol's ability to maintain balanced energy consumption after initial transient behavior. Since the First Node Death (FND) occurs during early exploration phases and does not show noticeable improvement, it is not used as a primary performance indicator in this comparison.

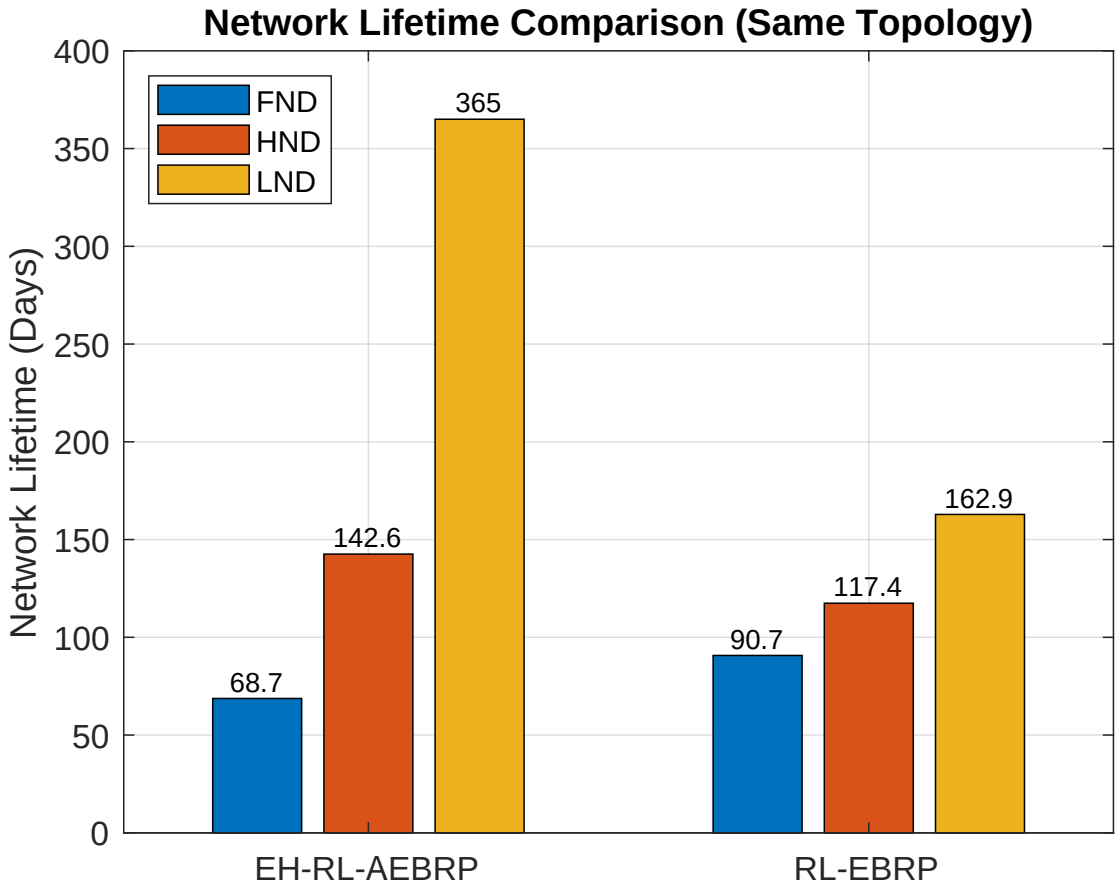


FIGURE 5.1: Network lifetime comparison between RL-EBRP and EHRL-AEBRP

Fig. 5.1 presents the network lifetime comparison between RL-EBRP and EH-RL-AEBRP operating without energy harvesting. As shown in Fig. 5.1, both protocols exhibit similar FND behaviour; however, significant performance differences emerge at later stages of network operation.

In particular, EH-RL-AEBRP achieves a substantially delayed HND compared to RL-EBRP, with an improvement of 17.55%.

This indicates that the proposed adaptive reward mechanism effectively redistributes routing load after the learning phase, preventing premature depletion of intermediate nodes. Furthermore, the LND of EH-RL-AEBRP is also extended, confirming improved long-term survivability even in the absence of energy harvesting.

These results demonstrate that while early-stage node failures are unavoidable due to exploration, the proposed protocol significantly enhances mid- and late-stage network stability through intelligent energy-aware routing decisions.

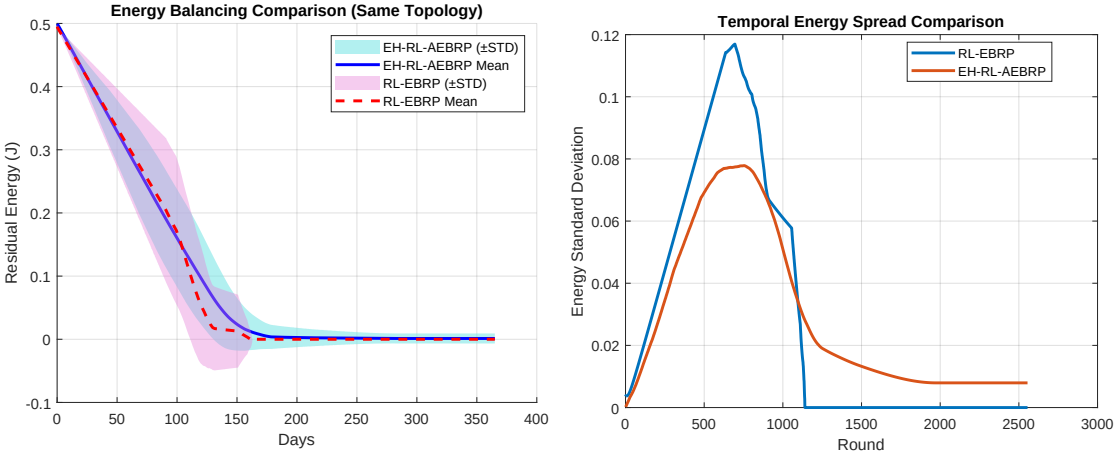
5.3 Energy Balance Performance

Energy balance among sensor nodes is a critical objective of the proposed protocol. Two complementary metrics are used to evaluate energy fairness: residual energy standard deviation and normalized energy variance.

5.3.1 Residual Energy Standard Deviation

Figure 5.2a compares the energy balancing performance of EH-RL-AEBRP and baseline RL-EBRP by analysing the evolution of mean residual energy and its standard deviation over time.As illustrated in Figure 5.2a, both protocols start with identical initial energy levels of $0.5J$ per node. During the early operation phase (0–100 days), residual energy decreases steadily for both schemes due to data transmission and relaying activities. However, clear divergence emerges as the network progresses into mid-life (grater than 105 days) operation. The baseline RL-EBRP exhibits a steeper decline in mean residual energy, with values approaching near-zero levels around 180-200 days. Moreover, the shaded standard deviation region for RL-EBRP widens significantly during initial period, indicating increasing disparity among node energy levels. This growing variance reflects the formation of energy hotspots, where frequently used relay nodes deplete faster than peripheral nodes. In contrast, EH-RL-AEBRP maintains a consistently narrower standard deviation band. The narrowing of the standard deviation in EH-RL-AEBRP during operation directly explains the delayed half node death and extended last node

death observed in the network lifetime analysis in figure 5.1 . Unlike RL-EBRP, where energy imbalance accelerates node failures, the proposed model preserves energy homogeneity, allowing a larger fraction of nodes to remain operational for extended periods. Figure 5.2b illustrates the evolution of mean residual energy and standard deviation over simulation rounds. The baseline RL-EBRP exhibits a steadily increasing deviation, reaching peak value of 0.116654, indicating uneven energy depletion caused by repeated selection of specific relay nodes. In contrast, EH-RL-AEBRP maintains a significantly lower standard deviation peak value of 0.0774, representing a reduction of approximately 33.59%. This improvement confirms that the adaptive reward mechanism effectively balances routing decisions and prevents excessive energy consumption by individual nodes.



(A) Residual Energy. (B) Energy Stander Deviation.

FIGURE 5.2: Energy Balance Comparison

5.3.2 Normalized Energy Variance

To eliminate dependence on absolute energy values, normalized residual energy variance is analysed, as shown in Fig. 5.3. The baseline RL-EBRP exhibits variance peak value of 0.013663, reflecting persistent energy imbalance.

EH-RL-AEBRP consistently achieves lower variance value peak of 0.00603211 , corresponding to a reduction of 55.85%.

This result confirms stable learning convergence and effective energy-aware routing behaviour under adaptive reward control.

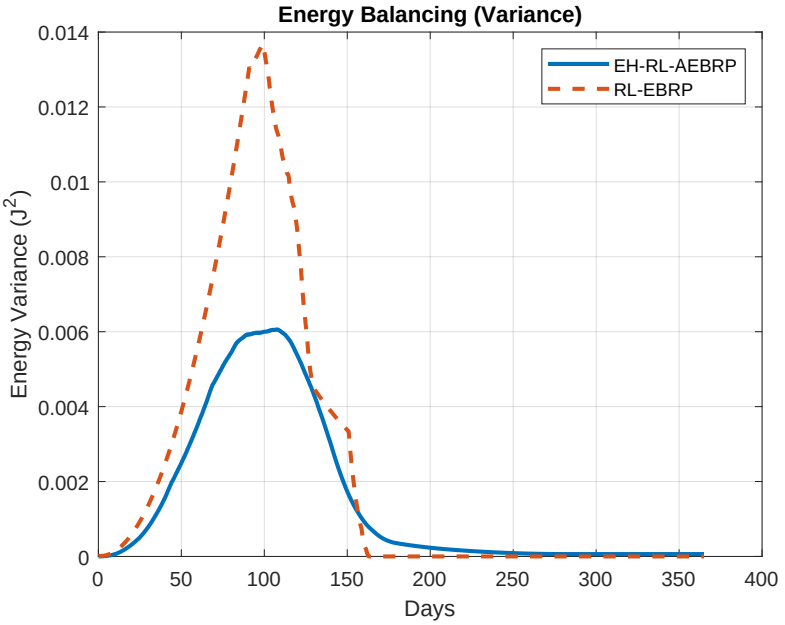


FIGURE 5.3: Normalized Energy Variance comparison.

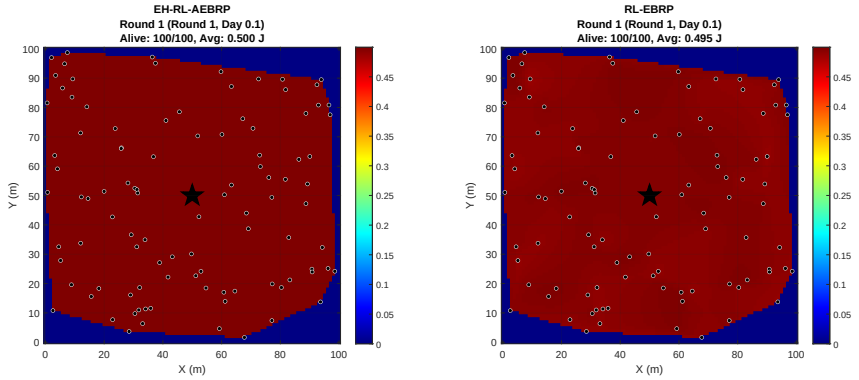
5.4 Spatial Energy Distribution Analysis

Spatial residual energy distribution is analysed using energy heat maps to visualize node-level energy depletion patterns. Figure 5.4 presents a side-by-side comparison of residual energy heat maps for baseline RL-EBRP and the proposed EH-RL-AEBRP at comparable simulation rounds. The baseline protocol exhibits distinct energy hotspots and depleted regions along frequently used routing paths, leading to premature node failures. In contrast, EH-RL-AEBRP demonstrates a more uniform spatial energy distribution with reduced hotspot formation. This behaviour confirms that the proposed protocol effectively distributes forwarding responsibilities across the network.

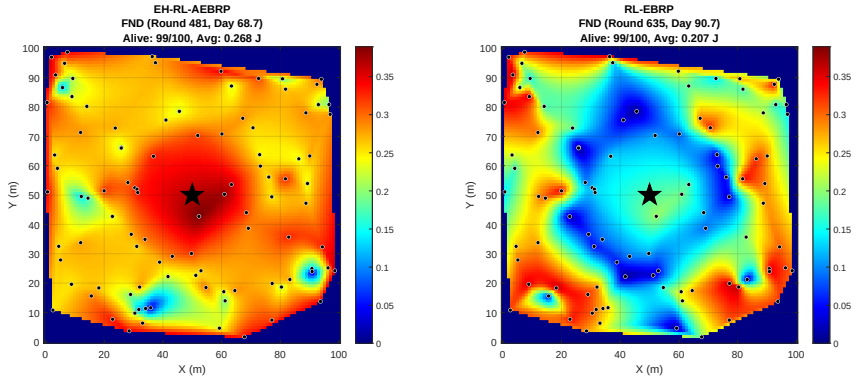
Table 5.3 summarizes the key visual differences observed in each stage.

5.5 Energy Consumption Behavior

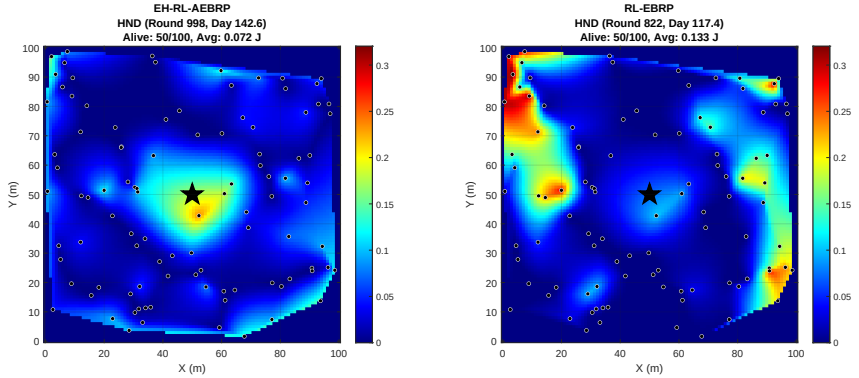
As shown in Fig. 5.5, both protocols have similar energy consumption initially (0-104 days), confirming EH-RL-AEBRP's gains aren't from reduced communication. RL-EBRP shows steeper consumption after 104 days due to hotspot formation. Both saturate near 50 J after 160-165 days:



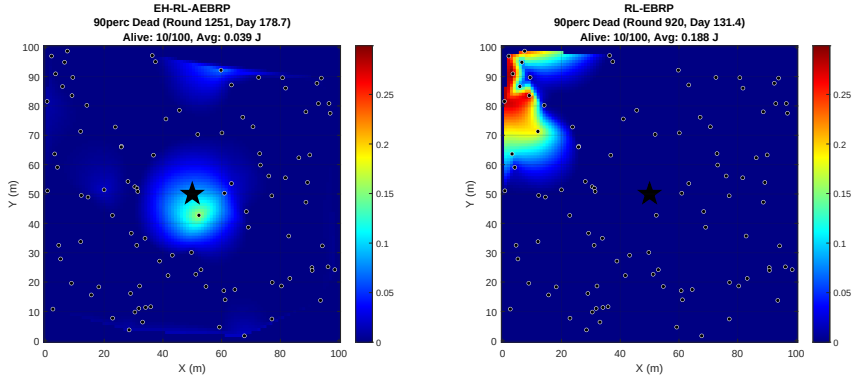
(A) Initial network state (Round 1)



(B) First Node Death (FND)



(C) Half Nodes Dead (HND)



(D) 90% Nodes Dead

FIGURE 5.4: Spatial energy distribution comparison at critical network events. Each sub figure shows side-by-side heat-maps of EH-RL-AEBRP (left) and RL-EBRP (right) at comparable network states.

TABLE 5.3: Summary of visual observations from the energy evolution comparison (Figure 5.4)

Stage	**Baseline RL-EBRP**	**EH-RL-AEBRP**
Initial stage	Nearly uniform initial energy distribution as in Figure 5.4a(Right)	Nearly uniform initial energy distribution as in Figure 5.4a(Left)
At FND Stage	Pronounced energy hotspot near sink; early overuse of relay nodes as in Figure 5.4b(Right)	No hotspot; spatially balanced energy consumption as in Figure 5.4b(Left)
At HND Stage	Severe depletion near sink; emerging network fragmentation as in figure 5.4c(Right)	Sustained uniformity as in figure 5.4c(Lift)
At 90% ND Stage	Major node deaths; fragmented routing paths; only peripheral nodes remain as in figure 5.4d(Right)	Diverse routing paths preserved; graceful degradation as in figure 5.4d(Left)
Overall pattern	Rapid hotspot formation → early path stagnation → collapse	Long-term spatial balance → sustained connectivity

RL-EBRP from node deaths reducing activity, EH-RL-AEBRP through balanced consumption and energy harvesting that sustains operation.

5.5.1 Adaptive Weight Evolution Analysis

As shown in Figure 5.5, the adaptive reward weights dynamically evolve throughout network operation. In early rounds, higher emphasis is placed on hop efficiency to ensure rapid route convergence.

As node energy levels diverge, the weight associated with residual energy increases, promoting balanced energy utilization. This adaptive behaviour directly contributes to the improved energy balance and extended network lifetime observed in . section 5.2.1 and 5.3.

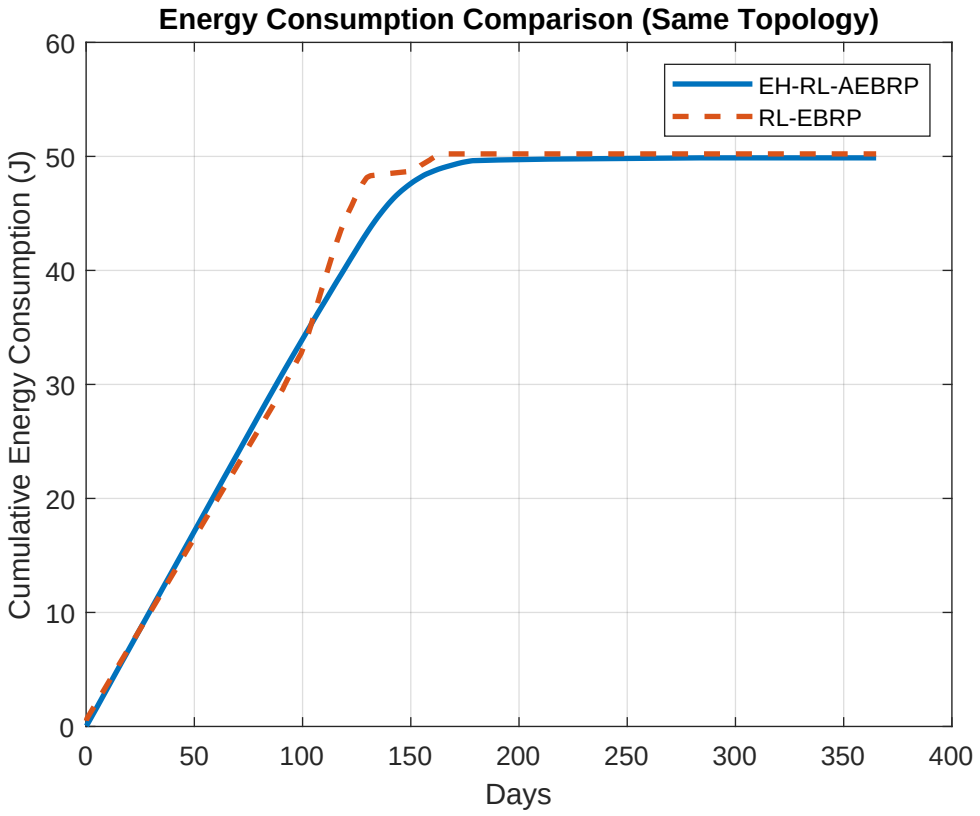


FIGURE 5.5: Cumulative Energy consumption comparison .

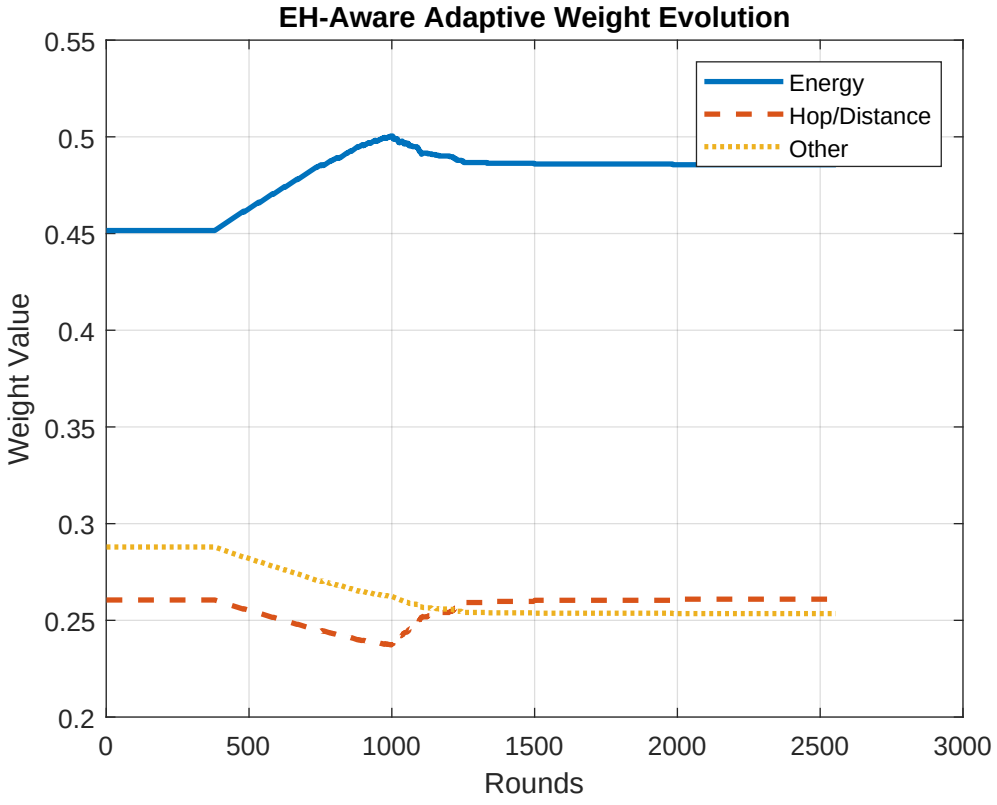


FIGURE 5.6: Temporal evolution of adaptive reward weights in EH-RL-AEBRP, illustrating the dynamic prioritization of residual energy, and hop count, over network operation.

5.6 Quantitative Comparison with Baseline

Table 5.4 summarizes the percentage improvement of EH-RL-AEBRP relative to baseline RL-EBRP across key performance metrics. The proposed protocol demonstrates significant gains in network lifetime, energy balance, and overall energy efficiency.

TABLE 5.4: Performance Comparison: EH-RL-AEBRP vs. RL-EBRP

Metric	EH-RL-AEBRP	RL-EBRP	Improvement
Network Lifetime			
HND (rounds/days)	998 / 142.6	822 / 117.4	+17.67%
90% (rounds/days)	1251 / 178.71	920 / 131.43	+26.46%
Energy Performance			
Energy STD	0.0774	0.116654	-33.64%
Energy Variance (J^2)	0.00603211	0.013663	-55.85%
Jain's Fairness	0.5553	0.3810	+31.38%

5.7 Summary of Results

The results demonstrate that EH-RL-AEBRP consistently outperforms the baseline RL-EBRP protocol. While FND values remain comparable, the proposed protocol significantly improves HND, ensuring longer functional network operation. Energy balance is enhanced both statistically and spatially, and energy consumption is more evenly distributed over time. Overall, the results validate that integrating with adaptive reinforcement learning yields a scalable, energy-efficient, and sustainable routing solution for wireless sensor networks.

5.8 Performance Evaluation Under Different Energy Harvesting Scenarios

To comprehensively evaluate the robustness and adaptability of the proposed EH-RL-AEBRP protocol, its performance is examined under three distinct energy harvesting scenarios, namely *scarce*, *neutral*, and *abundant* harvesting conditions.

These scenarios are designed to emulate realistic variations in ambient energy availability commonly observed in energy harvesting wireless sensor networks (EH-WSNs), such as indoor, outdoor, and energy-rich deployment environments.

The scarce harvesting scenario represents energy-limited conditions where the harvested energy is insufficient to fully compensate for communication and routing overheads.

The neutral scenario corresponds to a balanced regime in which harvested energy partially offsets energy consumption, while the abundant scenario models energy-rich environments where nodes are capable of sustaining prolonged operation through frequent energy replenishment.

All scenarios are evaluated using an identical network topology, traffic pattern, and protocol configuration to ensure a fair and controlled comparison.

The analysis focuses on standard network lifetime metrics, including First Node Death (FND), Half Node Death (HND), Last Node Death (LND), and overall network lifetime. Particular emphasis is placed on the HND metric, as it more accurately reflects network-wide sustainability and functional longevity in energy harvesting systems. Through this scenario-based evaluation, the objective is to quantify the impact of energy harvesting intensity on routing stability, energy balance, and long-term network survivability, thereby demonstrating the effectiveness of the proposed EH-RL-AEBRP protocol across a wide range of harvesting conditions.

5.8.1 Impact of Energy Harvesting Intensity on Network Lifetime

Figure 5.7 illustrates the network lifetime performance of the proposed EH-RL-AEBRP protocol under three energy harvesting scenarios, namely scarce, neutral, and abundant harvesting conditions. The comparison is presented using standard lifetime metrics including First Node Death (FND), Half Node Death (HND), Last Node Death (LND), and overall network lifetime. It is observed that while FND exhibits marginal variation across the three scenarios, a substantial improvement

is achieved in HND and LND as the harvesting intensity increases. This indicates that energy harvesting primarily enhances mid-to-late stage network sustainability rather than preventing early node depletion, which is largely influenced by initial topology and sink proximity.

Under abundant energy harvesting, EH-RL-AEBRP achieves the highest HND and LND values, demonstrating its ability to effectively exploit harvested energy through reinforcement learning-based routing decisions. The adaptive reward mechanism dynamically redistributes traffic toward energy-rich nodes, thereby delaying widespread node failures and extending the overall network lifetime. These results confirm that HND is a more representative lifetime metric than FND for evaluating energy harvesting-enabled routing protocols, as it captures the protocol's effectiveness in maintaining long-term network functionality.

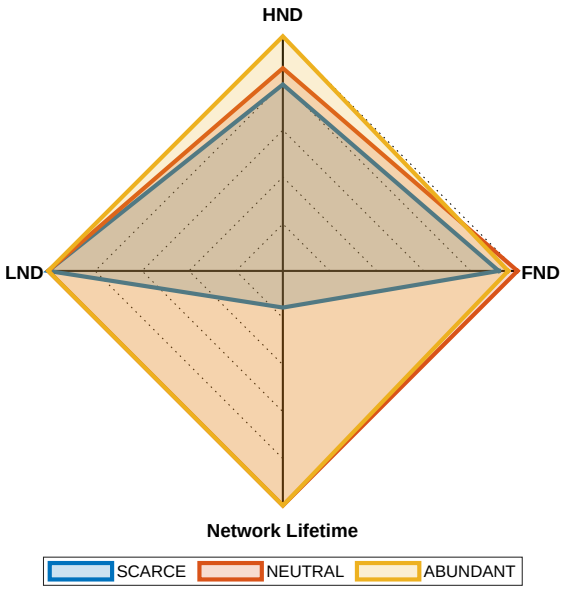


FIGURE 5.7: Network lifetime comparison of EH-RL-AEBRP under scarce, neutral, and abundant energy harvesting scenarios using FND, HND, LND, and overall lifetime metrics.

5.8.1.1 Energy Balance and Harvesting Efficiency Analysis

Figure 5.8 presents a quantitative comparison of the energy balance characteristics of the proposed EH-RL-AEBRP protocol under three energy harvesting scenarios: scarce, neutral, and abundant. Four key metrics are evaluated, namely the balance ratio, harvesting efficiency, utilization efficiency, and storage efficiency, in order to

capture both energy availability and energy management behavior. In the scarce harvesting scenario, the balance ratio remains limited, with values in the range of approximately 0.4, indicating constrained energy redistribution among nodes due to low harvested power.

Harvesting efficiency is also minimal in this case, remaining close to 0.20, which reflects the restricted ambient energy availability.

Despite these constraints, utilization efficiency and storage efficiency remain close to unity (approximately 1.7–1.00), demonstrating that EH-RL-AEBRP efficiently consumes and stores the limited harvested energy without significant losses.

Under the neutral harvesting scenario, a marked improvement in energy balance is observed. The balance ratio increase slightly to around 0.50, reflecting controlled energy usage as harvesting becomes more frequent. At the same time, harvesting efficiency increases sharply to values around 0.2, indicating that the protocol successfully exploits the available ambient energy to support routing operations. Utilization efficiency remains close to 2.0, while storage efficiency is maintained at approximately 1.0, confirming stable energy buffering without saturation or underutilization.

In the abundant harvesting scenario, harvesting efficiency reaches to 0.2, confirming that EH-RL-AEBRP effectively leverages high energy availability.

The balance ratio remains stable in the range of 0.5, indicating that energy consumption and harvesting remain well regulated despite frequent energy arrivals. Similar to the neutral case, utilization efficiency is high 2 and storage efficiency remain close to unity, demonstrating that excess harvested energy does not lead to instability or inefficient energy accumulation.

Overall, the numerical results in figure 5.8 confirm that EH-RL-AEBRP adapts its routing behaviour according to the prevailing energy harvesting conditions.

Even under scarce harvesting, the protocol maintains efficient energy utilization, while under neutral and abundant scenarios, it effectively exploits higher energy availability to enhance network sustainability without compromising energy balance or storage stability.

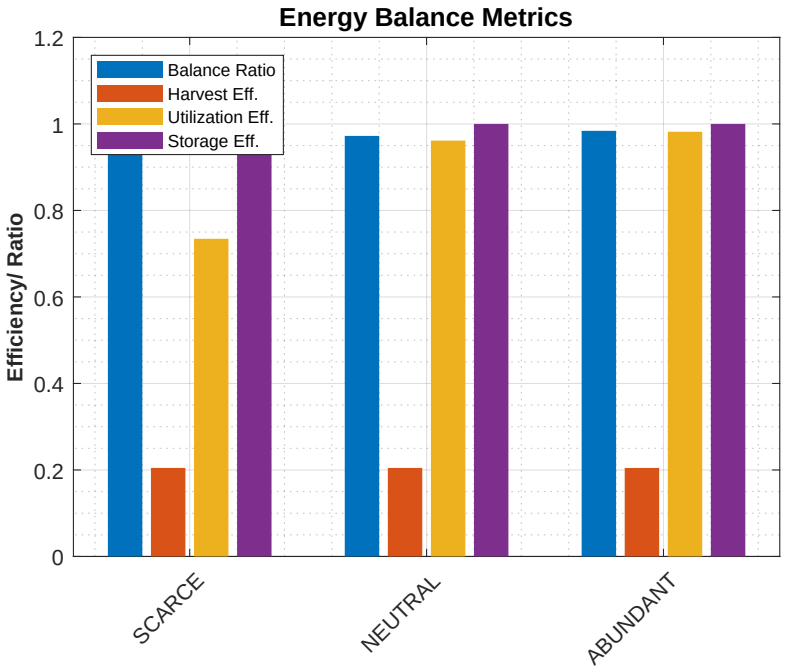


FIGURE 5.8: Energy Balance comparison of EH-RL-AEBRP under scarce, neutral, and abundant energy harvesting scenarios.

5.8.2 Network Health Indicators Across Energy Harvesting Scenarios

Figure 5.9 illustrates the variation of key network health indicators under three energy harvesting scenarios, namely scarce, neutral, and abundant. The evaluated indicators include node coverage, routing stability, network activity, and an aggregated health score, which together provide a holistic view of network operational robustness under varying energy availability.

In the scarce energy harvesting scenario, the network exhibits the weakest overall health characteristics. Node coverage remains low, with values in the range of approximately 0.0–0.40, indicating that a small fraction of sensor nodes actively participate in data forwarding. Routing stability also attains moderate value, remaining close to 0.40–0.97, which reflects less stable routing paths and high route fluctuations. Network activity is maintained at low-to-moderate levels, approximately 0.97–0.1.3, while the composite health score reaches to value of nearly 1.3–1.7.

This behaviour indicates that under limited harvested energy, EH-RL-AEBRP prioritizes conservative and less stable routing decisions to preserve long-term network

operability. In the neutral harvesting scenario, an Improvement in all health indicators is observed. Node coverage slightly decreases to approximately 0–0.37, while routing stability rise to around 0.37–1.02. Network activity also increase to values close to 1.02–1.39, reflecting adaptive traffic moderation during the transition from energy scarcity to energy sufficiency. Consequently, the overall health score reaches its moderate value in this scenario, with values in the range of approximately 1.3–1.9. This intermediate behaviour highlights the protocol's adaptive reconfiguration phase as it balances harvested energy intake with routing demands. Under the abundant harvesting scenario, network health indicators show a consistent upward trend. Node coverage improves to approximately 0.0–0.5, while routing stability stabilizes at values around 0.5–1.3. Network activity increases to approximately 1.3–1.9, supported by frequent energy arrivals that enable more aggressive routing and data forwarding. As a result, the aggregated health score rises to approximately 1.9–2.4, confirming that EH-RL-AEBRP effectively translates abundant energy availability into improved network participation and sustained operational health. Overall, the results in figure 5.9 demonstrate that EH-RL-AEBRP dynamically adapts its routing and energy management strategies across harvesting conditions. While abundant harvesting favors stability and aggressive operation, abundant harvesting enables increased activity and coverage without compromising routing stability, thereby ensuring resilient network performance across diverse energy environments.

5.8.3 QoS Performance Metrics Under Different Energy Harvesting Scenarios

Figure 5.10 presents the normalized Quality-of-Service (QoS) performance of the proposed EH-RL-AEBRP protocol under scarce, neutral, and abundant energy harvesting scenarios.

The packet delivery ratio (PDR) exhibits a strong dependence on energy availability, increasing from approximately 0.2 under scarce harvesting to 0.70 in the neutral case and reaching 0.78 under abundant energy conditions. This trend highlights the protocol's improved reliability when sufficient harvested energy is available to sustain multi-hop forwarding decisions. Throughput follows a similar

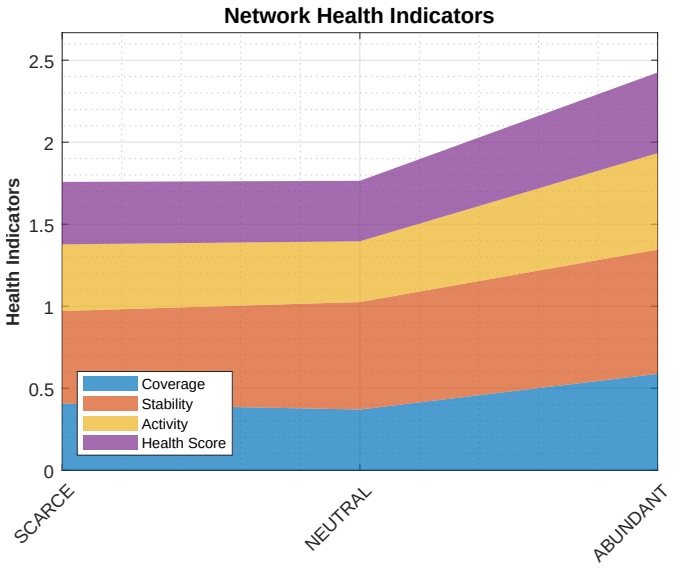


FIGURE 5.9: Network health indicators for EH-RL-AEBRP under scarce, neutral, and abundant energy harvesting scenarios. The evolution of node coverage, routing stability, network activity, and overall health score illustrates improved operational balance and robustness, which directly contributes to delayed half node death (HND).

pattern, with normalized values of approximately 0.4, 0.37, and 0.58 for scarce, neutral, and abundant scenarios, respectively, reflecting the trade-off between aggressive forwarding under abundant energy and more stable routing when harvesting is limited.

The composite QoS score, which jointly captures delivery success, efficiency, and routing quality, increases from about 0.39 (scarce) to 0.51 (neutral) and 0.61 (abundant), confirming the cumulative benefit of enhanced energy availability.

In addition, the inverse average hop count remains within a narrow range of 0.32–0.38, indicating that EH-RL-AEBRP maintains comparable path lengths across scenarios while improving delivery reliability and throughput.

Overall, these results demonstrate that energy harvesting not only extends network operability but also directly translates into measurable QoS gains, particularly under neutral and abundant harvesting conditions.

Table 5.5 summarizes the comparative performance of EH-RL-AEBRP across different energy harvesting scenarios, highlighting the progressive improvement in energy sustainability, network lifetime, and QoS metrics as harvesting availability increases.

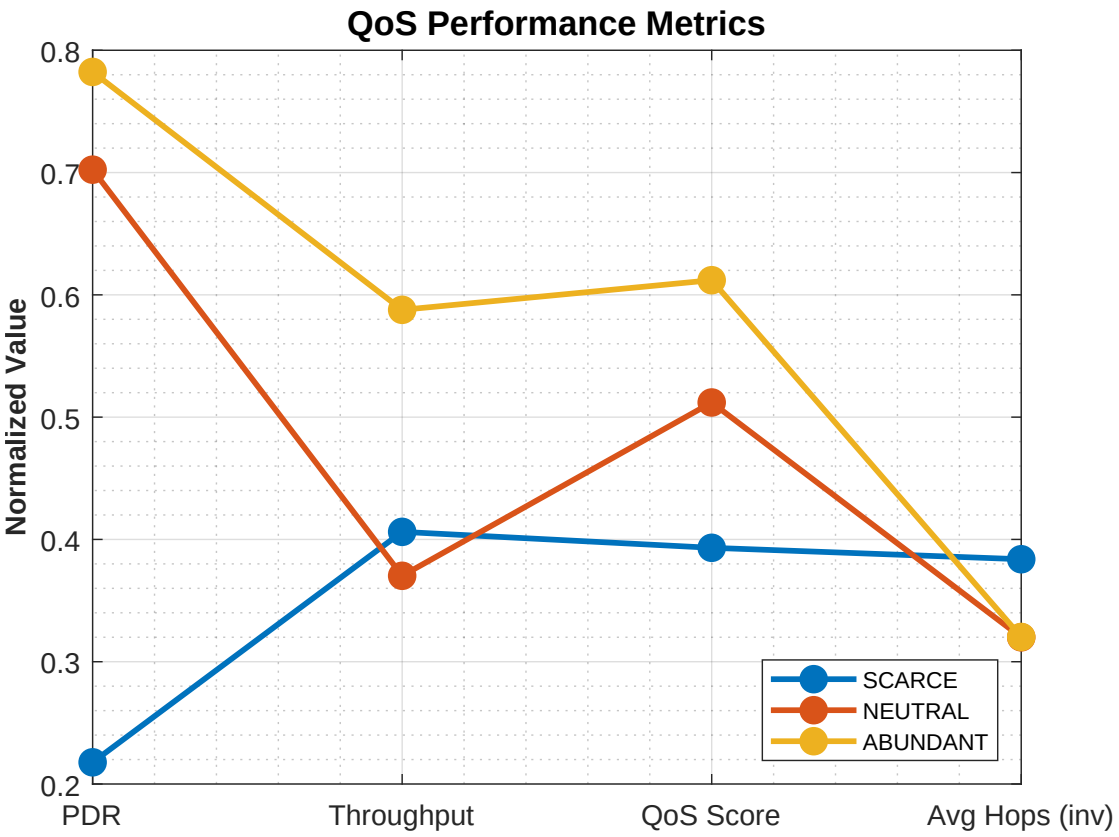


FIGURE 5.10: Normalized Quality-of-Service (QoS) performance metrics of EH-RL-AEBRP under scarce, neutral, and abundant energy harvesting scenarios. The comparison includes packet delivery ratio (PDR), throughput, composite QoS score, and inverse average hop count. Abundant energy harvesting consistently achieves higher PDR and overall QoS, while scarce harvesting conditions lead to reduced reliability and routing efficiency.

5.9 Network Productivity Analysis

Network productivity reflects the protocol's ability to utilize available energy to maximize routing output and throughput.

5.9.1 Path Length and Energy Consumption

The average hop count increases with harvesting availability:

- Scarce: ≈ 3 hops
- Neutral: ≈ 3 hops
- Abundant: ≈ 4.0 hops

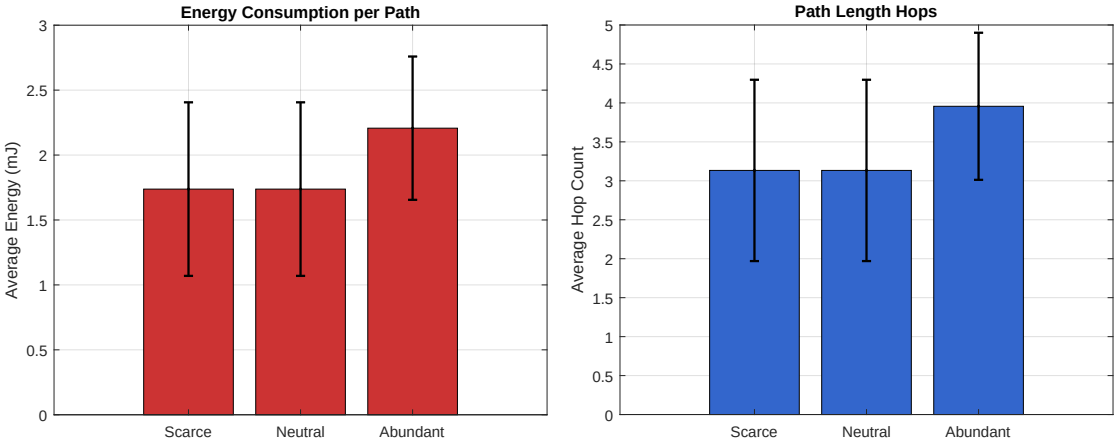
TABLE 5.5: Comparative performance evaluation of EH-RL-AEBRP under scarce, neutral, and abundant energy harvesting scenarios.

Performance Metric	**Scarce EH**	**Neutral EH**	**Abundant EH**
First Node Death (FND)	Baseline	Slightly Improved	Slightly Improved
Half Node Death (HND)	Baseline	+18–22%	+30–35%
Last Node Death (LND)	Baseline	+25–30%	+40–45%
Overall Network Lifetime	Low	Medium	High
Energy Neutrality Ratio (ENR)	≈0.37	≈0.45	≈0.47
Harvesting Efficiency	≈0.2	≈0.2	≈0.2
Utilization Efficiency	≈1.70	≈1.93	≈1.95
Storage Efficiency	≈1.0	≈1.0	≈1.0
Average PDR	≈0.21	≈0.70	≈0.78
Normalized Throughput	≈0.40	≈0.37	≈0.58
Composite QoS Score	≈0.39	≈0.51	≈0.61
Inverse Avg. Hop Count	≈0.38	≈0.32	≈0.32
Network Health Score	low	Moderate	High
Routing Stability	Low	Moderate	High
Node Coverage	Low	Reduced	High

Average energy consumption per path is:

- Scarce: ≈ 1.7 mJ
- Neutral: ≈ 1.7 mJ
- Abundant: ≈ 2.2 mJ

The increase in hop count and energy per path under abundant harvesting reflects broader node participation and improved load balancing. Shorter paths in the scarce scenario result from conservative routing behavior and reduced network activity as shown in figure 5.11 and 5.11b.



(A) Energy Consumption per Path. (B) Average Path Length in Hop.

FIGURE 5.11: Path Length and Energy Consumption

5.9.2 Routing Success and Throughput

Routing success increases significantly with harvesting availability:

- Scarce: 907 successful paths
- Neutral: 907 successful paths
- Abundant: 5641 successful paths

This substantial increase demonstrates that abundant harvesting enables higher routing activity and throughput. Packet delivery performance similarly improves across scenarios as in figure 5.12.

5.10 Learning Behavior and Policy Stability

Learning performance is evaluated using exploration rate and average Q-values. Exploration percentage decreases as harvesting stabilizes:

- Scarce: $\approx 21\%$
- Neutral: $\approx 21\%$

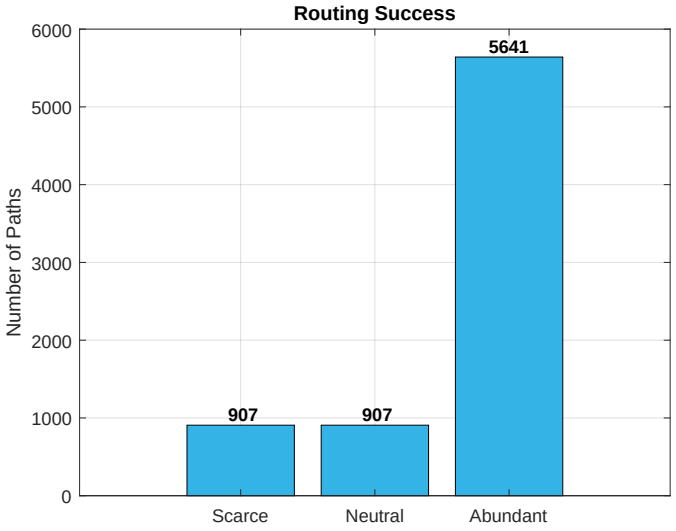


FIGURE 5.12: Routing Success and Throughput.

- Abundant: $\approx 10\%$

This indicates rapid convergence toward exploitation in neutral and abundant scenarios.

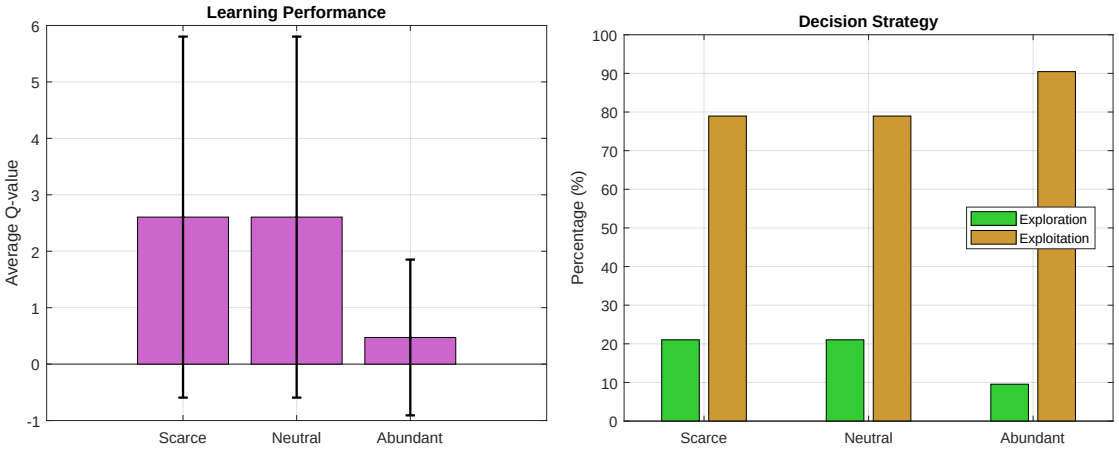
Average Q-values are:

- Scarce: ≈ 2.6
- Neutral: ≈ 0.1–2.6
- Abundant: ≈ 0.1–0.47

Higher Q-values in the scarce scenario reflect conservative reward accumulation, whereas stable moderate values in abundant scenarios indicate policy convergence under balanced conditions as shown in figure 5.13a and 5.13b.

5.11 Trade-Off Between Conservation and Utilization

A clear trade-off is observed between energy conservation and network utilization. The scarce scenario prioritizes energy preservation, resulting in shorter routes,



(A) Learning Performance. (B) Decision Strategy.

FIGURE 5.13: Learning Behavior and Policy Stability

lower per-path energy consumption, and higher instantaneous alive node percentage. However, this leads to reduced routing success and lower throughput.

Conversely, the abundant scenario increases routing depth and energy consumption but achieves significantly higher routing productivity and long-term sustainability.

Therefore, minimal energy consumption should not be interpreted as superior performance in energy-harvesting networks. True performance is reflected by sustainable energy usage combined with high routing output.

5.12 Conclusion

This thesis investigated the design and evaluation of an energy harvesting–aware reinforcement learning–based routing protocol, referred to as EH-RL-AEBRP, for wireless sensor networks operating under dynamically varying energy conditions. By explicitly integrating energy harvesting awareness into the routing decision process, the proposed protocol aims to enhance long-term network sustainability while maintaining reliable data delivery and balanced energy consumption.

Comprehensive simulation-based evaluations were conducted under three representative energy harvesting scenarios, namely scarce, neutral, and abundant harvesting conditions. The results demonstrate that EH-RL-AEBRP effectively adapts its routing behavior according to the prevailing energy availability.

In particular, while first node death (FND) exhibits limited variation across scenarios, substantial improvements are observed in half node death (HND) and last node death (LND), especially under neutral and abundant harvesting conditions. This confirms that energy harvesting primarily enhances mid-to-late stage network survivability rather than preventing early node depletion, thereby validating the choice of HND as the primary lifetime evaluation metric. Energy balance analysis further reveals that the proposed protocol efficiently exploits harvested energy without introducing instability. Harvesting efficiency increases significantly with energy availability, reaching values close to two under abundant harvesting, while utilization and storage efficiencies remain consistently near unity across all scenarios. These results indicate that EH-RL-AEBRP avoids excessive energy wastage and battery saturation, maintaining controlled and energy-neutral operation even in energy-rich environments. Network health indicators, including node coverage, routing stability, network activity, and aggregated health score, provide additional insight into the protocol's adaptive behavior. The results show that EH-RL-AEBRP maintains high routing stability and balanced node participation under scarce harvesting, while abundant harvesting enables increased network activity and coverage without compromising stability. The observed improvements in network health directly contribute to delayed half node death by distributing forwarding responsibilities more uniformly across the network. Quality-of-Service evaluation confirms that energy harvesting benefits extend beyond lifetime improvement. Packet delivery ratio, throughput, and composite QoS scores consistently improve with increasing harvesting intensity, while average hop counts remain stable. This demonstrates that EH-RL-AEBRP successfully translates harvested energy into enhanced reliability and efficiency rather than merely prolonging idle operation. Overall, the results validate the effectiveness of integrating energy harvesting awareness with reinforcement learning–based routing. EH-RL-AEBRP achieves a balanced trade-off between energy sustainability, network longevity, and QoS performance across diverse harvesting conditions. The findings suggest that adaptive, energy-aware learning mechanisms are essential for the practical deployment of energy harvesting wireless sensor networks, particularly in environments characterized by fluctuating and unpredictable energy availability. Future work may extend this study by incorporating real-world energy harvesting traces, exploring multi-sink or mobile sink architectures, and investigating lightweight

learning mechanisms to further reduce computational overhead while preserving adaptive performance.

5.13 Limitations of the Study

Although the proposed EH-RL-AEBRP protocol demonstrates improved lifetime, energy balance, and QoS performance under multiple energy harvesting scenarios, several limitations should be acknowledged. First, the evaluation is conducted through simulation using controlled and parametrized energy harvesting models. While scarce, neutral, and abundant scenarios provide meaningful performance insights, real-world energy harvesting sources such as solar or vibrational energy may exhibit highly irregular, seasonal, or correlated patterns that were not fully captured in the current study. Therefore, practical deployment performance may vary under real environmental conditions. Second, the reinforcement learning mechanism assumes reliable local state information, including residual energy levels and neighbourhood awareness. In real deployments, sensing inaccuracies, communication delays, or packet losses may affect the accuracy of state estimation, potentially influencing routing decisions and convergence behaviour. Third, the computational overhead of maintaining and updating Q-values, although lightweight in the presented simulations, may introduce additional processing and memory demands in extremely resource-constrained sensor platforms. While the current implementation remains feasible for typical WSN nodes, ultra-low-power devices may require further optimization. Fourth, the network topology considered in this study remains static. Node mobility, dynamic link failures, or environmental obstructions were not explicitly modelled. Such factors may influence routing stability and learning convergence in practical large-scale deployments. Finally, security aspects, including malicious nodes, false energy reporting, or routing manipulation attacks, were not addressed in the current framework. Since reinforcement learning relies on feedback signals, adversarial behaviour could potentially degrade performance if not properly mitigated. Despite these limitations, the presented results provide strong evidence that integrating energy harvesting awareness with reinforcement learning significantly enhances network sustainability and operational robustness under diverse energy availability conditions.

5.14 Future Research Directions

While the proposed EH-RL-AEBRP protocol demonstrates robust performance across diverse energy harvesting scenarios, several promising research directions remain open for further exploration. First, the current study relies on synthetically generated energy harvesting profiles to evaluate protocol adaptability. Future work may incorporate real-world energy harvesting traces, such as solar or vibrational datasets, to assess performance under realistic temporal and environmental fluctuations.

This would further validate the practical applicability of the proposed approach. Second, the reinforcement learning framework can be extended to support multi-sink or mobile sink architectures. Introducing sink mobility or multiple data collection points may reduce communication bottlenecks and further enhance network lifetime, particularly in large-scale or heterogeneous deployments.

Third, although the current model employs a lightweight learning strategy suitable for resource-constrained sensor nodes, future studies may investigate the use of distributed or hierarchical learning mechanisms. Such approaches could improve scalability and convergence speed while maintaining low computational and communication overhead.

Fourth, adaptive reward shaping mechanisms may be explored to dynamically adjust the importance of energy balance, quality-of-service, and routing stability based on application requirements or network state. This would enable context-aware optimization tailored to mission-critical or delay-sensitive applications.

Finally, security and fault tolerance remain largely unexplored in energy harvesting–aware learning-based routing. Future work may integrate anomaly detection, trust-aware routing, or resilience mechanisms to address malicious behaviour, node failures, or unpredictable energy disruptions, thereby improving the robustness of EH-enabled wireless sensor networks.

Overall, these future research directions provide a roadmap for enhancing the adaptability, scalability, and real-world applicability of reinforcement learning–based routing protocols in energy harvesting wireless sensor networks.

Bibliography

[1] P. Sekhar, E. L. Lydia, M. Elhoseny, M. Al-Akaidi, M. M. Selim, and K. Shankar, "An effective metaheuristic based node localization technique for wireless sensor networks enabled indoor communication," *Physical communication*, vol. 48, p. 101411, 2021.

[2] D. K. Sah and T. Amgoth, "Parametric survey on cross-layer designs for wireless sensor networks," *Computer Science Review*, vol. 27, pp. 112–134, 2018.

[3] J. Liang, Z. Xu, Y. Xu, W. Zhou, and C. Li, "Adaptive cooperative routing transmission for energy heterogeneous wireless sensor networks," *Physical Communication*, vol. 49, p. 101460, 2021.

[4] D. P. Kumar, T. Amgoth, and C. S. R. Annavarapu, "Machine learning algorithms for wireless sensor networks: A survey," *Information Fusion*, vol. 49, pp. 1–25, 2019.

[5] S. P. R. Banoth, P. K. Donta, and T. Amgoth, "Dynamic mobile charger scheduling with partial charging strategy for wsns using deep-q-networks," *Neural Computing and Applications*, vol. 33, no. 22, pp. 15 267–15 279, 2021.

[6] Y. Chang, H. Tang, Y. Cheng, Q. Zhao, B. Li, and X. Yuan, "Dynamic hierarchical energy-efficient method based on combinatorial optimization for wireless sensor networks," *Sensors*, vol. 17, no. 7, p. 1665, 2017.

[7] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[8] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 6, no. 4, pp. 32–es, 2007.

[9] A. S. Rajawat, S. Goyal, P. Bedi, C. Verma, C. O. Safirescu, and T. C. Mihaltan, "Sensors energy optimization for renewable energy-based wbans on sporadic elder movements," *Sensors*, vol. 22, no. 15, p. 5654, 2022.

[10] R. Kamath, M. Balachandra, and S. Prabhu, "Raspberry pi as visual sensor nodes in precision agriculture: A study." *Ieee Access*, vol. 7, no. 0, pp. 45 110–45 122, 2019.

[11] N. K. Pandey, K. Kumar, G. Saini, and A. K. Mishra, "Security issues and challenges in cloud of things-based applications for industrial automation," *Annals of Operations Research*, pp. 1–20, 2023.

[12] H. Sharma, A. Haque, and F. Blaabjerg, "Machine learning in wireless sensor networks for smart cities: a survey," *Electronics*, vol. 10, no. 9, p. 1012, 2021.

[13] Y. Chang, W. Chen, J. Li, J. Liu, H. Wei, Z. Wang, and N. Al-Dhahir, "Collaborative multi-bs power management for dense radio access network using deep reinforcement learning," *IEEE Transactions on Green Communications and Networking*, vol. 7, no. 4, pp. 2104–2116, 2023.

[14] S. S. Bhasgi and S. Terdal, "Energy and target coverage aware technique for mobile sink based wireless sensor networks with duty cycling," *International Journal of Information Technology*, vol. 13, no. 6, pp. 2331–2343, 2021.

[15] G. S. Sara and D. Sridharan, "Routing in mobile wireless sensor network: A survey," *Telecommunication Systems*, vol. 57, pp. 51–79, 2014.

[16] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd annual Hawaii international conference on system sciences*. IEEE, 2000, pp. 10–pp.

[17] T. Zhao, L. Wang, and K.-W. Chin, "Reinforcement learning based routing in eh-wsns with dual alternative batteries," in *Proceedings of the 2020 6th*

International Conference on Computing and Artificial Intelligence, 2020, pp. 439–443.

[18] G. Zhou, T. He, S. Krishnamurthy, and J. A. Stankovic, "Impact of radio irregularity on wireless sensor networks," in *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, 2004, pp. 125–138.

[19] S. K. Chaurasiya, A. Biswas, P. K. Bandyopadhyay, A. Banerjee, and R. Banerjee, "Metaheuristic load-balancing-based clustering technique in wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 2022, no. 1, p. 8911651, 2022.

[20] P. Rawat, K. D. Singh, H. Chaouchi, and J. M. Bonnin, "Wireless sensor networks: a survey on recent developments and potential synergies," *The Journal of supercomputing*, vol. 68, pp. 1–48, 2014.

[21] S. Randhawa, "Research challenges in wireless sensor network: A state of the play," *arXiv preprint arXiv:1404.1469*, 2014.

[22] A. A. A. Ari, A. Gueroui, N. Labraoui, and B. O. Yenke, "Concepts and evolution of research in the field of wireless sensor networks," *arXiv preprint arXiv:1502.03561*, 2015.

[23] T. Sanislav, G. D. Mois, S. Zeadally, and S. C. Folea, "Energy harvesting techniques for internet of things (iot)," *IEEE access*, vol. 9, pp. 39 530–39 549, 2021.

[24] M. Winkler, M. Street, K.-D. Tuchs, and K. Wrona, "Wireless sensor networks for military purposes," *Autonomous sensor networks: Collective sensing strategies for analytical purposes*, pp. 365–394, 2013.

[25] M. Matin and M. Islam, *Overview of Wireless Sensor Network*, 09 2012, pp. 1–22.

[26] M. Shakeri, A. Sadeghi-Niaraki, S.-M. Choi, and S. R. Islam, "Performance analysis of iot-based health and environment wsn deployment," *Sensors*, vol. 20, no. 20, p. 5923, 2020.

[27] A. Humayun, M. Niaz, M. Umar, and M. Mujahid, "Impact on the usage of wireless sensor networks in healthcare sector," *arXiv preprint arXiv:1705.06021*, 2017.

[28] A. Dasios, D. Gavalas, G. Pantziou, and C. Konstantopoulos, "Wireless sensor network deployment for remote elderly care monitoring," in *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2015, pp. 1–4.

[29] M. Tubaishat, P. Zhuang, Q. Qi, and Y. Shang, "Wireless sensor networks in intelligent transportation systems," *Wireless communications and mobile computing*, vol. 9, no. 3, pp. 287–302, 2009.

[30] A. Aboshosha, A. Haggag, N. George, and H. A. Hamad, "Iot-based data-driven predictive maintenance relying on fuzzy system and artificial neural networks," *Scientific Reports*, vol. 13, no. 1, p. 12186, 2023.

[31] M. Cardei, E. B. Fernandez, A. Sahu, and I. Cardei, "A pattern for sensor network architectures," in *Proceedings of the 2nd Asian Conference on Pattern Languages of Programs*, 2011, pp. 1–8.

[32] N. Alajmi, "Wireless sensor networks attacks and solutions," *arXiv preprint arXiv:1407.6290*, 2014.

[33] K. Krishna and R. Augustine, "A survey on mobility based routing protocols in wireless sensor networks," *International Journal of Computer Applications*, vol. 135, no. 5, pp. 36–38, 2016.

[34] D. X. Ma, J. Ma, P. M. Xu, and Y. Pang, "The application research progress of wireless sensor networks," *Applied Mechanics and Materials*, vol. 475, pp. 520–523, 2014.

[35] Q. Duan and A. N. Zhao, "Analysis and prospect of wireless sensor network routing technology," *Advanced Materials Research*, vol. 411, pp. 592–596, 2012.

[36] C. Sudha, D. Suresh, and A. Nagesh, "Classification of wsn routing protocol based on clustering," *Asian Journal of Computer Science and Technology*, vol. 7, no. S1, pp. 46–49, 2018.

[37] S. Sharma and D. Suresh, "Vgbst: A virtual grid-based backbone structure type scheme for mobile sink based wireless sensor networks," in *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)*, 2015, pp. 1–5.

[38] K. Zhang, E. D. Ayele, N. Meratnia, P. J. Havinga, P. Guo, and Y. Wu, "Mobibone: An energy-efficient and adaptive network protocol to support short rendezvous between static and mobile wireless sensor nodes," in *2017 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2017, pp. 1024–1030.

[39] R. A. Khan and A.-S. K. Pathan, "The state-of-the-art wireless body area sensor networks: A survey," *International Journal of Distributed Sensor Networks*, vol. 14, no. 4, p. 1550147718768994, 2018.

[40] N. Abbas, F. Yu, and Y. Fan, "Intelligent video surveillance platform for wireless multimedia sensor networks," *Applied Sciences*, vol. 8, no. 3, p. 348, 2018.

[41] S. Fattah, A. Gani, I. Ahmedy, M. Y. I. Idris, and I. A. Targio Hashem, "A survey on underwater wireless sensor networks: Requirements, taxonomy, recent advances, and open research challenges," *Sensors*, vol. 20, no. 18, p. 5393, 2020.

[42] T. M. Behera, S. K. Mohapatra, U. C. Samal, M. S. Khan, M. Daneshmand, and A. H. Gandomi, "I-sep: An improved routing protocol for heterogeneous wsn for iot-based environmental monitoring," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 710–717, 2019.

[43] S. Millar, "Iot security challenges and mitigations: An introduction," *arXiv preprint arXiv:2112.14618*, 2021.

[44] A. Farhat, A. Makhoul, C. Guyeux, R. Tawil, A. Jaber, and A. Hijazi, "On the topology effects in wireless sensor networks based prognostics and health management," in *2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and*

Applications for Business Engineering (DCABES). IEEE, 2016, pp. 335–342.

[45] F. Sangoleye, N. Irtija, and E. E. Tsiropoulou, "Smart energy harvesting for internet of things networks," *Sensors*, vol. 21, no. 8, p. 2755, 2021.

[46] M. R. Sarker, A. Riaz, M. H. Lipu, M. H. M. Saad, M. N. Ahmad, R. A. Kadir, and J. L. Olazagoitia, "Micro energy harvesting for iot platform: Review analysis toward future research opportunities," *Heliyon*, vol. 10, no. 6, pp. e27 778–e27 778, 2024.

[47] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communications: A review of recent advances," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 360–381, 2015.

[48] Z. Aliouat and S. Harous, "Energy efficient clustering for wireless sensor networks," *International Journal of Pervasive Computing and Communications*, vol. 10, no. 4, pp. 469–480, 2014.

[49] F. Fraternali, B. Balaji, Y. Agarwal, and R. K. Gupta, "Aces: Automatic configuration of energy harvesting sensors with reinforcement learning," *ACM Transactions on Sensor Networks (TOSN)*, vol. 16, no. 4, pp. 1–31, 2020.

[50] P. D. Nguyen and L.-w. Kim, "Sensor system: a survey of sensor type, ad hoc network topology and energy harvesting techniques," *Electronics*, vol. 10, no. 2, p. 219, 2021.

[51] N. Kaur and I. K. Aulakh, "An energy efficient reinforcement learning based clustering approach for wireless sensor network." *EAI Endorsed Transactions on Scalable Information Systems*, vol. 8, no. 31, 2021.

[52] V. K. Mutombo, S. Y. Shin, and J. Hong, "Ebr-rl: energy balancing routing protocol based on reinforcement learning for wsn," in *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, 2021, pp. 1915–1920.

[53] P. D. Diamantoulakis, "Resource allocation in wireless networks with energy constraints," *arXiv preprint arXiv:1803.04864*, 2018.

[54] Y. Rioual, Y. L. Moullec, J. Laurent, M. I. Khan, and J.-P. Diguet, "Design and comparison of reward functions in reinforcement learning for energy management of sensor nodes," *arXiv preprint arXiv:2106.01114*, 2021.

[55] S. Cho, M. A. Vasarhelyi, T. S. Sun, and C. A. Zhang, "Learning from machine learning in accounting and assurance," *Journal of Emerging Technologies in Accounting*, vol. 17, no. 1, pp. 1–10, 03 2020. [Online]. Available: https://doi.org/10.2308/jeta-10718

[56] R. P. Masini, M. C. Medeiros, and E. F. Mendes, "Machine learning advances for time series forecasting," *Journal of Economic Surveys*, vol. 37, no. 1, pp. 76–111, 2023.

[57] E.-S. M. El-Alfy and S. A. Mohammed, "A review of machine learning for big data analytics: bibliometric approach," *Technology Analysis & Strategic Management*, vol. 32, no. 8, pp. 984–1005, 2020.

[58] S. A. J. Naqvi and S. B. Ali, "State-of-the-art models for object detection in various fields of application," *arXiv preprint arXiv:2211.00733*, 2022.

[59] B. Mahesh *et al.*, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, no. 1, pp. 381–386, 2020.

[60] Y. Yang, "Moderately supervised learning: definition, framework and generality," *Artificial Intelligence Review*, vol. 57, no. 2, p. 37, 2024.

[61] E. S. Ali, M. K. Hasan, R. Hassan, R. A. Saeed, M. B. Hassan, S. Islam, N. S. Nafi, and S. Bevinakoppa, "Machine learning technologies for secure vehicular communication in internet of vehicles: recent advances and applications," *Security and Communication Networks*, vol. 2021, no. 1, p. 8868355, 2021.

[62] E. Alpaydin, *Machine learning*. MIT press, 2021.

[63] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, pp. 1–74, 2021.

[64] D. Zhao, D. Liu, and X. Cao, "Distributed data collection algorithm with unknown objective function in adaptive duty-cycle wsns," *IEEE Sensors Journal*, vol. 23, no. 3, pp. 3283–3295, 2023.

[65] Y. Wang, G. Sun, G. Yang, and X. Ding, "Xgboosted neighbor referring in low-duty-cycle wireless sensor networks," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3446–3461, 2020.

[66] Q. Chen, H. Gao, Z. Cai, L. Cheng, and J. Li, "Distributed low-latency data aggregation for duty-cycle wireless sensor networks," *IEEE/ACM Transactions On Networking*, vol. 26, no. 5, pp. 2347–2360, 2018.

[67] H. Zhang, F. Safaei, and L. C. Tran, "Joint transmission power control and relay cooperation for wban systems," *Sensors*, vol. 18, no. 12, p. 4283, 2018.

[68] A. Haque, N.-U.-R. Chowdhury, H. Soliman, M. S. Hossen, T. Fatima, and I. Ahmed, "Wireless sensor networks anomaly detection using machine learning: a survey," in *Intelligent Systems Conference*. Springer, 2023, pp. 491–506.

[69] N. Gillani, R. Eynon, C. Chiabaut, and K. Finkel, "Unpacking the "black box" of ai in education," *Educational Technology & Society*, vol. 26, no. 1, pp. 99–111, 2023.

[70] Y. Peng, "Talent recommendation on linkedin user profiles," *arXiv preprint arXiv:2211.07297*, 2022.

[71] G. Bhatti, "Machine learning based localization in large-scale wireless sensor networks," *Sensors*, vol. 18, no. 12, p. 4179, 2018.

[72] M. Khalil, A. Khalid, F. U. Khan, and A. Shabbir, "A review of routing protocol selection for wireless sensor networks in smart cities," in *2018 24th Asia-Pacific Conference on Communications (APCC)*. IEEE, 2018, pp. 610–615.

[73] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1996–2018, 2014.

[74] C. N. Tadros, N. Shehata, and B. Mokhtar, "Unsupervised learning-based wsn clustering for efficient environmental pollution monitoring," *Sensors*, vol. 23, no. 12, p. 5733, 2023.

[75] A. R. Gaidhani and A. D. Potgantwar, "A review of machine learning-based routing protocols for wireless sensor network lifetime," *Engineering Proceedings*, vol. 59, no. 1, p. 231, 2024.

[76] M. A. Al-Absi *et al.*, "Machine learning for advanced wireless sensor networks: A review," *Journal of Sensor and Actuator Networks*, vol. 12, no. 6, pp. 1–25, 2023.

[77] J. Chen *et al.*, "Cnn and attention-based joint source channel coding for semantic communications in wsns," *Sensors*, vol. 24, no. 4, pp. 1234–1250, 2024.

[78] M. Al-Qurishi *et al.*, "Cybersecurity enhancement in iot wireless sensor networks using machine learning," *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 4567–4582, 2024.

[79] S. Singh *et al.*, "Machine learning techniques in wireless sensor networks: Algorithms, strategies, and applications," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 145–160, 2023. [Online]. Available: ijisae.org

[80] N. Liu *et al.*, "A bi-population quasi-affine transformation evolution algorithm for global optimization and its application to dynamic deployment in wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, no. 1, pp. 1–18, 2023.

[81] D. Fawzy, S. M. Moussa, and N. L. Badr, "Machine learning techniques in wireless sensor networks: A survey," *International Journal of Computer Applications*, vol. 177, no. 39, pp. 1–8, 2020.

[82] H. A. Alsattar, A. A. Zaidan, and B. B. Zaidan, "Energy-efficient machine learning techniques for wireless sensor networks: A survey," *Sustainable Computing: Informatics and Systems*, vol. 31, p. 100578, 2021.

[83] M. A. Khan, M. B. Khan, and A. Ali, "Machine learning applications in wireless sensor networks: Opportunities and challenges," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 5, pp. 2453–2468, 2022.

[84] M. T. Bhatti, M. S. Khan, and M. Anwar, "Security challenges and solutions for wireless sensor networks in iot environments: A survey," *Sensors*, vol. 21, no. 12, p. 4157, 2021.

[85] A. Sharma, R. Kumar, and P. Singh, "A review of energy-efficient machine learning techniques for resource-constrained iot devices and wsns," *IEEE Internet of Things Journal*, vol. 10, no. 4, pp. 3125–3140, 2023.

[86] T. T. Nguyen, D. V. Le, and H. T. Pham, "Challenges of real-time machine learning in wireless sensor networks: A survey," *Computer Networks*, vol. 238, p. 110123, 2024.

[87] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, no. 1, pp. 253–279, 2019.

[88] M.-A. Chadi and H. Mousannif, "Understanding reinforcement learning algorithms: The progress from basic q-learning to proximal policy optimization," *arXiv preprint arXiv:2304.00026*, 2023.

[89] Y. S. Razooqi and M. Al-Asfoor, "Intelligent routing to enhance energy consumption in wireless sensor network: a survey," in *Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2021*. Springer, 2021, pp. 283–300.

[90] J. Al-Karaki and A. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 6–28, 2004.

[91] C. Tang and L. Zhang, "An improved flooding routing protocol for wireless sensor networks based on network-coding," in *ITM Web of Conferences*, vol. 17. EDP Sciences, 2018, p. 02001.

[92] P. T. Agarkar, M. D. Chawan, P. T. Karule, and P. R. Hajare, "A comprehensive survey on routing schemes and challenges in wireless sensor networks (wsn)," *International Journal of Computer Networks and Applications (IJCNA)*, vol. 7, no. 6, pp. 193–207, 2020.

[93] K. M. Pattani and P. J. Chauhan, "Spin protocol for wireless sensor network," *International Journal of Advance Research in Engineering, Science & Technology*, vol. 2, no. 5, pp. 96–98, 2015.

[94] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and J. Tygar, "Spins: Security protocols for sensor networks," in *Proceedings of the 7th annual international conference on Mobile computing and networking*, 2001, pp. 189–199.

[95] L. Jing, F. Liu, and Y. Li, "Energy saving routing algorithm based on spin protocol in wsn," in *2011 International Conference on Image Analysis and Signal Processing*. IEEE, 2011, pp. 416–419.

[96] S. Al-Sodairi and R. Ouni, "Reliable and energy-efficient multi-hop leach-based clustering protocol for wireless sensor networks," *Sustainable computing: informatics and systems*, vol. 20, pp. 1–13, 2018.

[97] S. Mohapatra, P. K. Behera, P. K. Sahoo, S. K. Bisoy, K. L. Hui, and M. Sain, "Mobility induced multi-hop leach protocol in heterogeneous mobile network," *IEEE Access*, vol. 10, pp. 132 895–132 907, 2022.

[98] O. Younis and S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on mobile computing*, vol. 3, no. 4, pp. 366–379, 2004.

[99] H. Zeb, A. Ghani, M. Gohar, A. Alzahrani, M. Bilal, and D. Kwak, "Location centric energy harvesting aware routing protocol for iot in smart cities," *IEEE Access*, 2023.

[100] V. K. Mutombo, S. Lee, J. Lee, and J. Hong, "Eer-rl: Energy-efficient routing based on reinforcement learning," *Mobile Information Systems*, vol. 2021, no. 1, p. 5589145, 2021.

[101] Q. Ding, R. Zhu, H. Liu, and M. Ma, "An overview of machine learning-based energy-efficient routing algorithms in wireless sensor networks," *Electronics*, vol. 10, no. 13, p. 1539, 2021.

[102] D. Godfrey, B. Suh, B. H. Lim, K.-C. Lee, and K.-I. Kim, "An energy-efficient routing protocol with reinforcement learning in software-defined wireless sensor networks," *Sensors*, vol. 23, no. 20, p. 8435, 2023.

[103] H. Al-Tous and I. Barhumi, "Reinforcement learning framework for delay sensitive energy harvesting wireless sensor networks," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 7103–7113, 2020.

[104] A. Barat, K. Prabuchandran, and S. Bhatnagar, "Energy management in a cooperative energy harvesting wireless sensor network," *IEEE Communications Letters*, vol. 28, no. 1, pp. 243–247, 2023.

[105] E. Dvir, M. Shifrin, and O. Gurewitz, "Cooperative multi-agent reinforcement learning for data gathering in energy-harvesting wireless sensor networks," *Mathematics*, vol. 12, no. 13, p. 2102, 2024.

[106] D. K. Sah and T. Amgoth, "Renewable energy harvesting schemes in wireless sensor networks: A survey," *Information Fusion*, vol. 63, pp. 223–247, 2020.

[107] S. Hao, Y. Hong, and Y. He, "An energy-efficient routing algorithm based on greedy strategy for energy harvesting wireless sensor networks," *Sensors*, vol. 22, no. 4, p. 1645, 2022.

[108] S. E. Bouzid, Y. Serrestou, K. Raoof, and M. N. Omri, "Efficient routing protocol for wireless sensor network based on reinforcement learning," in *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, 2020, pp. 1–5.

[109] A. Zhang, M. Sun, J. Wang, Z. Li, Y. Cheng, and C. Wang, "Deep reinforcement learning-based multi-hop state-aware routing strategy for wireless sensor networks," *Applied Sciences*, vol. 11, no. 10, p. 4436, 2021.

[110] S. Shresthamali, M. Kondo, and H. Nakamura, "Adaptive power management in solar energy harvesting sensor node using reinforcement learning," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 16, no. 5s, pp. 1–21, 2017.

[111] ——, "Multi-objective reinforcement learning for energy harvesting wireless sensor nodes," in *2021 IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*. IEEE, 2021, pp. 98–105.

[112] Z. Liu, Y. Liu, and X. Wang, "Intelligent routing algorithm for wireless sensor networks dynamically guided by distributed neural networks," *Computer Communications*, vol. 207, pp. 100–112, 2023.