A Reinforcement Learning-Based Energy Aware Routing Algorithm for WSNs

Prajesh.V.Kuchekar, Ryan Pandit
Department of Computer Science and Engineering
National Institute of Technology Karnataka
Surathkal, Mangalore, India
+91 9172469676, +91 7483952376
prajeshkuchekar.221ec125@nitk.edu.in, ryanpandit.221ec147@nitk.edu.in

April 17, 2025

Abstract

Routing algorithms play a crucial role in determining efficient paths for data transmission in wireless sensor networks (WSNs), which are often deployed in energy-constrained environments. With the growth of IoT, smart cities, and environmental monitoring applications, energy conservation has become vital to ensure long-term, reliable network operations. The primary challenge lies in designing routing protocols that minimize energy consumption while maintaining acceptable performance in terms of latency, throughput, and reliability. Traditional approaches such as LEACH, PEGASIS, and DEAR aim to improve energy efficiency through clustering, chain-based transmission, or distance-aware strategies. However, these protocols still face challenges like uneven energy depletion, static role assignment, and limited adaptability to dynamic network conditions. These issues often lead to premature node failures and reduced network lifetime. This work addresses these limitations by proposing a new energy-efficient routing approach that considers key factors such as residual energy, distance to the sink, and balanced energy usage across nodes. The proposed algorithm aims to extend network lifetime by optimizing route selection and load distribution dynamically. The main objectives of the proposed work are: (a) to develop an energy-aware routing protocol that minimizes overall energy consumption, and (b) to improve network sustainability through adaptive and balanced routing decisions.

1 Introduction

Routing algorithms are essential in network communication, as they determine the most efficient paths for data transmission across interconnected nodes [1]. These algorithms improve delivery performance by considering factors such as hop count, bandwidth, and congestion. Traditional networks rely on protocols such as Open Shortest Path First (OSPF) [2] and Border Gateway Protocol (BGP) [3] to ensure reliability and low latency. In mobile and wireless environments, protocols such as Ad hoc On-Demand Distance Vector (AODV) [4] and Dynamic Source Routing (DSR) [5] support dynamic topologies. However, these routing protocols typically prioritize metrics like speed and reliability while overlooking energy efficiency, which is a growing concern in battery-operated wireless systems.

The major challenge in Wireless Sensor Networks (WSNs) [6] is the energy constraint of sensor nodes. In environments such as Mobile Ad Hoc Networks (MANETs) [7], IoT deployments [8], and MEMS-based sensor systems [9], energy consumption directly impacts network lifespan. Shortest-path routing often leads to overuse of certain nodes, resulting in uneven battery drainage and early node death. Furthermore, high-power transmissions increase interference and reduce network stability. Static routing methods cannot adapt to changes in energy levels or topology, and load balancing techniques that ignore energy consumption cannot fully optimize network longevity. These issues emphasize the importance of energy-sensitive routing mechanisms that can dynamically adjust to node conditions.

Several energy-sensitive routing protocols [10] have been proposed to address these concerns. LEACH (Low Energy Adaptive Clustering Hierarchy) [11], PEGASIS (Power-Efficient Gathering in Sensor Information System) [12], and DEAR (Distance-based Energy Aware Routing) [13] are notable examples. LEACH uses periodic cluster head (CH) rotation to distribute energy use, PEGASIS employs chain-based routing to minimize transmissions, and DEAR integrates distance and energy metrics for next-hop selection. While these protocols offer energy improvements, they suffer from limitations like random CH selection, high latency in chain structures, and static routing metrics that don't adapt well to dynamic conditions. As a result, energy consumption is often imbalanced, and network lifetime remains suboptimal.

The proposed work aims to address the above issues by designing a new energy-efficient routing protocol that dynamically balances energy consumption and adapts to real-time network changes. This approach builds on the concepts of energy-awareness and optimized path selection while seeking to minimize transmission overhead and improve load distribution across nodes. Unlike existing protocols that rely on fixed strategies, our method focuses on flexible decision-making based on current network parameters. The main objectives of this work are threefold: (1) to develop an energy-efficient routing protocol that optimizes network lifespan while maintaining high data transmission efficiency, (2) to implement adaptive routing techniques that dynamically adjust based on network conditions to reduce unnecessary energy consumption, and (3) to compare the proposed routing strategy with existing protocols such as LEACH, PEGASIS, DEAR, RLBR, and DADF by evaluating key performance metrics like energy consumption, throughput, and network lifetime.

The rest of the report is structured as follows: Section 2 reviews the related literature on energy-efficient routing protocols. Section 3 provides the comparative analysis of the literature review. Section 4 discusses the proposed methodology and the designed algorithm. Section 5 provides the implementation details along with the simulation parameters and setup. Section 6 presents simulation results and a comparative analysis of the traditional algorithms with the proposed one. Finally, Section 7 concludes the report with future research directions.

2 Literature Survey

This section presents a review of significant contributions made by researchers in the field of energy-efficient routing protocols for Wireless Sensor Networks (WSNs) [14]. The discussion primarily focuses on well-known and widely used protocols such as LEACH [11], DEAR [13], PEGASIS [12], RLBR, and DADF. These protocols have been developed with the aim of optimizing energy usage, prolonging network lifetime, and enhancing data delivery in resource-constrained wireless environments.

Despite their effectiveness, most existing WSN routing protocols suffer from limitations such as uneven energy consumption, static routing structures, and limited adaptability to dynamic network conditions. These drawbacks often result in early node failures, increased latency, and reduced overall performance in large or energy-critical deployments.

2.1 Low Energy Adaptive Clustering Hierarchy (LEACH) Protocol

This review is based on the paper titled Low Energy Adaptive Clustering Hierarchy (LEACH) Protocol: A Retrospective Analysis [11], which has been selected as the base reference for this study. The paper explores the LEACH protocol, its operational phases, and several variations developed to enhance energy-aware routing in Wireless Sensor Networks (WSNs). Heinzelman et al. [**?**] originally proposed LEACH as a self-organizing, adaptive clustering protocol that distributes energy consumption evenly among nodes by rotating the Cluster Head (CH) role. LEACH divides communication into rounds, each comprising a setup and steady-state phase. In the setup phase, nodes probabilistically elect themselves as CHs and advertise their status. Other nodes join clusters based on signal strength. A DAMMA schedule is used for data transmission, allowing nodes to conserve energy. In the steady-state phase, data is aggregated by the CH and sent to the base station. CDMA codes reduce inter-cluster interference, and the protocol can support hierarchical clustering for large-scale deployments. Despite its energy-saving approach, LEACH has notable limitations. It assumes direct CH-to-base station communication, which is energy-intensive in large networks. Its CH selection does not factor in residual energy, risking early node failure. Moreover, inefficient cluster formation can lead to energy imbalances and reduced network longevity.

The paper also proposes several variants of LEACH to overcome these limitations. Yi *et al.* (2007) [15] proposed the PEACH protocol, which reduces energy consumption by forming clusters adaptively using wireless overhearing, avoiding extra overhead and supporting multi-level clustering for both location-aware and location-unaware WSNs. Shi *et al.* (2002) [16] proposed LEACH-C, a centralized clustering protocol where the base station collects location and residual energy information from all sensor nodes and applies a Simulated Annealing algorithm to form energy-efficient and balanced clusters, minimizing transmission distances and enhancing network lifetime. Song *et al.* (2010) [17] proposed the AFSO protocol, a hierarchical routing enhancement to LEACH, where the base station utilizes an Artificial Fish Swarm Optimization algorithm for more effective cluster head (CH) selection based on nodes' location and residual energy. While this approach improves the setup phase efficiency, it introduces communication overhead due to centralized decision-making and has limited impact on the steady-state phase, which remains similar to LEACH. The following table shows a comparative analysis of various variant of leach algorithms that are discussed in the table.

Ref	Protocol	Scalability	Communication Overhead	Energy Efficiency
[11]	LEACH	Moderate	High (due to random CH selection)	Low (due to lack of energy consideration)
[15]	PEACH	High (due to priority-based selection)	Moderate (optimized selection reduces overhead)	High (residual energy and proximity considered)
[16]	LEACH-C	Low (centralized control limits scalability)	Low (centralized selection reduces overhead)	Moderate (energy considered but limited by centralization)
[17]	LEACH-AFSO	High (adaptive to network conditions)	Low (optimized cluster formation reduces overhead)	Very High (energy, distance, and network conditions optimized)

Table 1: The Various LEACH Algorithms Proposed

2.2 A Distance-Based Energy Aware Routing (DEAR) Protocol

This paper introduces a novel energy-efficient routing algorithm designed to extend the operational lifetime of wireless sensor networks (WSNs). The core contributions of this work include: **Optimal Multi-Hop and Distance Calculation** – a theoretical model that determines the ideal number of multi-hop transmissions and spacing between nodes, considering various traffic scenarios such as time-based and event-driven communication; **The DEAR Algorithm** – the proposed Distance-based Energy Aware Routing (DEAR) [13] strategy, which functions through route setup and maintenance stages. It prioritizes the physical distance between nodes for path selection while using residual energy levels as a secondary parameter, promoting balanced energy depletion across the network; and **Simulation Results** – extensive performance evaluations that demonstrate DEAR's effectiveness in reducing energy consumption and significantly increasing the network's lifetime compared to conventional routing schemes. The system under study models a typical WSN as an undirected graph $G = \langle V, E \rangle$, where V denotes the set of sensor nodes and E represents the communication links between them. Nodes are randomly distributed across a two-dimensional monitoring area, and any two nodes can establish a connection if the Euclidean distance between them is within the defined communication radius R. The undirected nature of the graph implies bidirectional communication, ensuring that if node i can send data to node j, then node j can also send data back to node i. The objective is to identify optimal or near-optimal routing distances so that energy consumption remains balanced among all participating nodes.

The network considers multiple traffic patterns including time-based, event-driven, query-based, and hybrid traffic, each with distinct energy implications. For instance, time-based traffic involves periodic reporting, while event-driven traffic generates burst transmissions in response to detected stimuli. The DEAR algorithm dynamically adjusts routing strategies based on these patterns to balance load across nodes. Each sensor maintains a *routing table*—storing information about previous, current, and next hops—and a *neighbor table*—containing details like distance, energy level, and degree of nearby nodes. This structure enables adaptive route formation and local route maintenance in case of link failure, ensuring continuous and efficient operation.

Each sensor node consumes energy to transmit, receive, and forward data. The energy consumption models used in DEAR are defined as follows:

$$E_{Tx}(l, d) = l \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d^2, \quad if d < d_0 \tag{1}$$

$$E_{Tx}(l, d) = l \cdot E_{elec} + l \cdot \varepsilon_{mp} \cdot d^4, \quad if d \geq d_0 \tag{2}$$

Here, E_{elec} denotes the energy consumed per bit by the transmitter/receiver circuitry, while ε_{fs} and ε_{mp} are the amplifier energies for the free-space and multi-path fading models, respectively. The threshold distance d_0 determines which model applies: for $d < d_0$, the free-space model with path loss exponent 2 is used; otherwise,

the multi-path model with exponent 4 applies. The equations (1) and (2) form the foundation for estimating transmission and reception energy in DEAR, guiding energy-aware routing decisions to extend network lifetime.

2.3 PEGASIS Routing Protocol

The following summary reviews the PEGASIS protocol for energy-efficient routing in wireless sensor networks (WSNs), along with its notable variants. This review is based on the reference paper Energy Efficient PEGASIS Routing Protocol for Wireless Sensor Networks [18]. Stephanie Lindsey *et al.* [12] (2002) proposed PEGASIS as a chain-based hierarchical routing protocol designed to improve energy efficiency in Wireless Sensor Networks. Unlike LEACH, which uses cluster-based CH selection, PEGASIS connects sensor nodes in a chain using a greedy algorithm where each node communicates only with a nearby neighbor. One node in the chain is then selected as the leader to transmit the aggregated data to the sink. This approach reduces long-range transmissions and balances energy usage across nodes. The chain formation is done locally using a greedy strategy that selects the closest neighbor, creating an efficient path with minimal transmission overhead. Once the chain is formed, a token-based mechanism controls communication order along the chain, and the leader node is dynamically selected based on its proximity to the sink and remaining energy. Data fusion occurs as information is passed along the chain, reducing transmission volume and conserving energy.

To address energy imbalance and further enhance performance, variations like Energy-Efficient PEGASIS (EE-PEGASIS) [18] and PEGASIS-E [19] have been proposed . Vinod Kumar *et al.* [18] further demonstrates that the EE-PEGASIS protocol begins by dividing the sensor network field into four parts, with nodes distributed across these regions. For example, 100 nodes might be divided into 25 nodes per part. The greedy algorithm is applied to form a chain, starting with the nodes in the first region. Next, the protocol selects a leader node based on its proximity to the other nodes and the sink node, considering both energy levels and the distance to ensure efficient communication. The leader node is chosen by evaluating the energy of each node relative to its distance from others, with a preference for nodes that have a balance between energy and distance to the sink. Energy consumption of child nodes is then calculated, considering the transmission distance between nodes. If the distance between two nodes is large, more energy is consumed, while shorter distances lead to less energy consumption. This ensures that energy is evenly distributed throughout the network. The process is repeated for all nodes to form the final EE-PEGASIS network, optimizing energy efficiency and prolonging the network's lifespan. Vibha Nehra *et al.* [19] proposed Pegasis-E which is an improved chain-based routing protocol that operates in rounds consisting of chain formation, leader selection, and data transmission. The chain is built starting from the node farthest from the base station (BS), progressively connecting nearby nodes. In each round, a different node is randomly selected as the leader to forward aggregated data to the BS. Data transmission follows a token-passing scheme, where nodes send fused data along the chain to the leader using TDMA scheduling. The chain is reconstructed dynamically in case of node failures.

Ref	**Protocol**	**Chain Formation**	**Leader Selection**	**Energy Efficiency**
[12]	PEGASIS	Formed using a greedy algorithm based on nearest neighbor	Based on proximity to sink	Reduces transmission range and balances energy better than LEACH
[18]	EE-PEGASIS	Divides the network into regions and applies greedy chain formation region-wise	Based on both residual energy and distance to sink and other nodes	Further improves energy distribution by selecting optimal leaders and regional partitioning
[19]	PEGASIS-E	Chains are formed based on energy levels and node proximity to the sink	Leader selection is dynamic, based on both energy levels and node proximity	Enhances energy efficiency by balancing load

Table 2: Different Variants of PEGASIS

2.4 Reinforcement Learning Based Routing

The RLBR (Reinforcement Learning-Based Routing) protocol [20] introduces a hybrid routing approach that blends both reactive and proactive elements to improve reliability in wireless sensor networks (WSN). The RLBR scheme allows a child sensor node to dynamically search for a new reliable parent node with more residual energy. This dynamic adaptation strategy can alleviate the energy hole problem as stated in [10][23]. It equips each sensor node with the ability to quickly switch to a new parent node when the existing one fails—either due to energy depletion or poor link quality—by using readily available neighborhood routing tables. RLBR leverages hardware-level Channel State Information (such as RSSI from radios like CC1000), alongside software-estimated parameters such as packet reception ratio (PRR), residual energy, and node-level hierarchy. These factors are integrated into a dynamic cost function that helps the node choose the best possible parent based on current network conditions. Essentially, each sensor chooses a parent from its own or a lower level in the routing tree, based on criteria like link quality, hop count, latency, and energy ensuring that the route to the base station remains both energy efficient and stable.

What sets RLBR apart is its ability to quickly recover from disruptions. When a sensor detects that its current parent node is no longer viable, it enters a brief waiting period to discover a new route. If no suitable parent is found within this timeframe, the node consults its proactively maintained routing table—a backup built using overheard routing information from neighboring nodes. This allows the network to adapt quickly without relying on lengthy route rediscovery processes. The routing tree, originally constructed using a simple shortest path method, evolves dynamically as nodes update their parent selections based on improved cost values. This proactive-reactive blend reduces data loss, improves energy distribution, and ensures the network remains responsive to real-time topology changes caused by energy depletion or signal degradation.

2.5 Delay-aware data fusion: A Q-Learning Approach

The proposed Delay-Aware Data Fusion (DADF) [21] algorithm optimizes data transmission in duty-cycled wireless sensor networks (WSNs) [22] by incorporating two key phases: Hierarchical Data Fusion (HDF) and Forwarding Node Selection (FNS). In the HDF phase, data is pre-processed at the source node to remove duplicates and inconsistencies, reducing energy consumption and unnecessary transmissions. This is accomplished through statistical measures such as mean and standard deviation, which help identify and discard erroneous data before transmission to the base station (BS). On the other hand, the FNS process selects the best forwarding node by utilizing a reinforcement learning (RL) approach, specifically Q-learning [23], to dynamically choose neighboring nodes that minimize energy consumption and end-to-end delay. The algorithm continuously evaluates possible nodes and updates its knowledge base through an iterative process, making it adaptable to the varying conditions of the network.

During operation, the BS oversees the training and control of node behavior. It gathers information about the network, such as node activity cycles and energy levels, which are then used to train the system using the Q-learning algorithm. By iterating this process, the system gradually refines its ability to select the optimal routing path based on factors like the nodes' active times and their proximity to the BS. The forwarding node selection adapts in real-time, helping maintain efficient routing despite changes in node availability due to varying duty cycles. The algorithm proves robust in maintaining network connectivity and ensuring energy-efficient routing, even as the duty cycle decreases, which could otherwise lead to isolated nodes. The overall computational complexity of the DADF algorithm is a combination of the complexities of the HDF and FNS phases, with the worst-case scenario involving a cubic time complexity for node selection due to the reinforcement learning process.

3 Analysis of literature review

This section presents a detailed discussion and comparative analysis of the key protocols and algorithms reviewed in the previous section, including LEACH [11], PEGASIS [12], DEAR [13], RLBR [20], and DADF [21], along with their respective variants. The comparison focuses on their core mechanisms, strengths, limitations, and suitability for energy-efficient and reliable routing in Wireless Sensor Networks (WSNs). The protocols are compared based on several critical parameters such as energy efficiency, scalability, load balancing, reliability, adaptability to topology changes, duty cycling support, and the incorporation of reinforcement learning techniques.

From the comparative study of existing routing protocols in wireless sensor networks (WSNs), it becomes evident that different protocols are designed with distinct trade-offs in mind. LEACH and PEGASIS, for instance, show good energy efficiency but lack robustness in areas like load balancing and adaptability to topology changes. LEACH performs reasonably well in scalability due to its clustering mechanism, but it doesn't dynamically adapt when the network topology evolves, nor does it incorporate mechanisms like duty

Protocol / Variant	Energy Efficiency	Scalability	Load Balancing	Reliability	Adaptive to Topology Changes	Duty Cycling Support	RL
LEACH	✓	✓	✓	✗	✓	✗	✓
PEGASIS	✓	✗	✗	✓	✗	✗	✗
DEAR	✗	✓	✓	✓	✗	✗	✗
RLBR	✓	✓	✓	✓	✓	✗	✓
DADF	✓	✓	✓	✓	✓	✓	✓

Table 3: Comparative Analysis of WSN Routing Protocols

cycling. PEGASIS, although energy-aware, struggles with scalability and load distribution due to its chain-based structure, which becomes inefficient as the network grows.

More recent protocols such as DEAR, RLBR, and DADF exhibit a more well-rounded approach to routing in WSNs. DEAR improves significantly by addressing reliability and load balancing while maintaining energy efficiency, though it still lacks adaptability and advanced energy-saving strategies like duty cycling. RLBR makes further strides by introducing reinforcement learning principles, improving adaptiveness and routing intelligence, but still omits certain features like sleep scheduling. Among all, DADF stands out by integrating nearly all desirable traits: energy efficiency, scalability, reliability, topology adaptability, duty cycling, and reinforcement learning. This suggests a clear evolution in protocol design, moving from simple energy-saving models to more intelligent, flexible systems capable of long-term, sustainable performance in dynamic WSN environments. These insights underline the importance of designing protocols that balance performance with adaptability, particularly as WSNs become increasingly complex and application-specific.

4 Presented Methodology

4.1 Overview

The proposed algorithm, i.e., a reinforcement-learning based energy aware algorithm integrates conventional wireless sensor network routing strategies with modern machine learning techniques. It operates through three core phases: routing, sleep scheduling, and data transmission control, as illustrated in Fig. 5. The routing phase leverages a reinforcement learning approach, while the data transmission control and sleep scheduling phases are built upon traditional methodologies.

4.2 Main Algorithm

The algorithm proposed combines the reinforcement-learning based adaptive routing policies of RLBR, and the parametric considerations of DEAR, i.e., residual energies, link quality and the distance between the nodes. There are two main scenarios observed: one at the non-cluster head nodes, and that of the cluster heads, as shown in Fig 1.

For regular nodes, the process begins by checking for commands from the sleep scheduling unit. Based on the command, the node either enters sleep mode or activates. Active nodes collect sensor data and update their cache. If transmission is permitted by the restricted data unit, the node sends the data to the cluster head using the IEEE 802.11 protocol. [24]

For cluster heads, incoming packets are checked—those from other clusters are prioritized, while packets from cluster members are aggregated with previous data. The cluster head then runs the routing algorithm to select the next hop. If a forwarder is found, the packet is sent; otherwise, it is dropped.

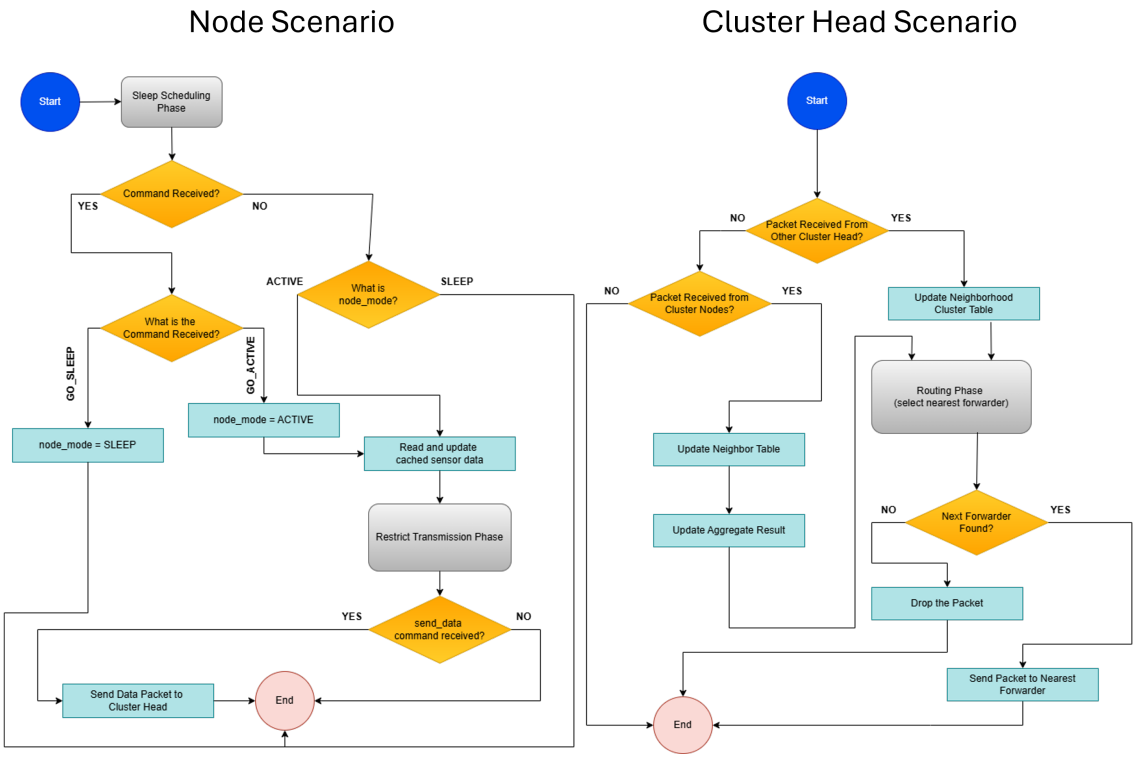

Figure 1: Protocol Node Scenarios

4.3 Network Initialization

In the initial stage of setting up the Wireless Sensor Network (WSN), a predefined number of sensor nodes are deployed randomly across a two-dimensional field, emulating real-world deployment scenarios, such as aerial drops or irregular placement in an open area. A static sink node is positioned centrally within the field, functioning as the primary hub for data aggregation and forwarding. Each sensor node is assigned a unique identifier, random coordinates in the field, and an initial energy level that simulates a battery-powered device. After deployment, nodes begin discovering their neighbors by identifying other nodes within their transmission range. This neighbor discovery process establishes local communication links, forming the basic topology that will support data transmission and routing within the network. It ensures that each node has a clear understanding of its immediate surroundings, enabling efficient and localized communication.

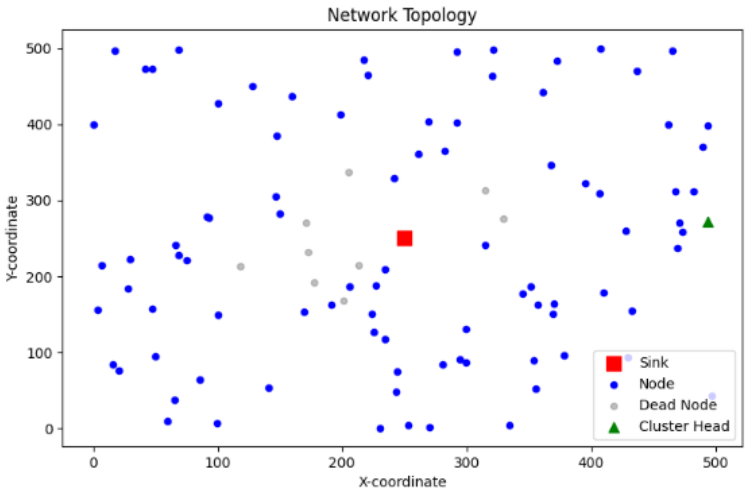

Figure 2: Deployed WSN Topology

Once the neighbor relationships are formed, a specific number of non-sink nodes are randomly selected to act as Cluster Heads (CHs), which play a critical role in aggregating data from neighboring nodes and forwarding the processed information to the sink node. This helps reduce redundancy in data transmission,

ultimately conserving energy and improving the overall network efficiency. After the CHs are selected, each regular sensor node assesses its distance to the available CHs using the Euclidean distance formula and associates itself with the nearest CH. This step ensures that each node communicates only with its closest CH, minimizing long-range communication and energy consumption. By organizing the network into clusters and assigning CHs, the WSN forms a more energy-efficient and scalable communication structure, enabling the network to handle larger areas and longer operational lifetimes. This phase lays the groundwork for the more advanced routing algorithms and energy-aware mechanisms that will be implemented later, aiming to further optimize the network's performance.network operation.

4.4 Network Discovery Phase

Once the foundational structure of the network is set up, the next critical phase involves ensuring that sensor nodes can effectively communicate with one another, and it lays the groundwork for introducing a learning-based routing mechanism. This stage is vital because it equips each node with the necessary tools and information to make intelligent, adaptive decisions regarding the forwarding of data. In this phase, the nodes begin to explore their local environment and establish communication links with nearby nodes, which is the first step in creating a dynamic and self-sustaining network. Each sensor node starts by calculating the Euclidean distance to all other nodes within its vicinity, checking whether each node falls within its transmission range. When a node detects another within its range, it recognizes the other node as a "neighbor" and records it as part of its local connectivity map. This process helps establish a network topology where nodes know which other nodes are within reach for potential data exchange. The neighbor discovery process is crucial because it forms the foundation for all future routing decisions, allowing the network to communicate locally before expanding to wider areas. This local connectivity mapping plays a pivotal role in ensuring the network's efficient operation, as it enables nodes to collaborate seamlessly while laying the groundwork for the self-organization necessary for scalable and energy-efficient routing.

With neighbor relationships established, the next task is to initialize Q-values for each connection. These Q-values represent the expected reward or cost associated with selecting a particular neighbor as the next hop for data forwarding. Initially, the Q-values may be assigned randomly or uniformly, but they evolve as the network operates, guided by a reinforcement learning framework. The Q-values are updated iteratively based on the feedback received during data transmission. To further enhance the routing decisions, each node calculates its hop count to the central sink node using the Breadth-First Search (BFS) algorithm, providing an understanding of how far it is from the sink. This hop count helps in shaping the reward function for Q-value updates, favoring shorter or more energy-efficient paths. With all the necessary components in place, the network enters the learning phase, where nodes can refine their routing decisions by balancing the exploration of new routes with the exploitation of known optimal paths. This process gradually allows the network to adapt to dynamic changes, enhancing its energy efficiency and overall performance in a self-organizing manner.

4.5 Sleep Scheduling Phase

To ensure energy efficiency, an adaptive sleep scheduling mechanism is implemented, which helps to reduce unnecessary energy consumption by the sensor nodes. Each node in the network is tasked with continuously monitoring environmental data, but rather than transmitting data constantly, it only sends updates when significant changes are detected. The sensor nodes read data from their respective environmental sensors, which are modeled with a normal distribution to mimic realistic conditions. When a node detects a change in the environmental data that exceeds a predefined threshold known as the (CHANGE THRESHOLD), it triggers a transmission process to send the updated data to the sink node. This approach prevents excessive communication and conserves energy, as nodes only transmit when there is a significant change in the sensed data.

However, if a node does not detect any significant change over a series of consecutive rounds, it enters a sleep state to further save energy. The duration for which a node stays in this sleep state is determined by a SLEEP THRESHOLD, which specifies how many rounds a node must remain inactive before entering sleep mode. During the sleep phase, the node halts all data sensing and transmission activities, allowing it to conserve energy. After a predetermined period, known as SLEEP DURATION, the sleeping nodes are scheduled to wake up and resume their operations. This ensures that the network operates in a more energy-efficient manner without sacrificing the integrity of the data. By balancing active sensing and transmission with periods of rest, this sleep scheduling mechanism helps to prolong the overall lifetime of the sensor network while ensuring that data updates are still timely and relevant when significant changes occur.

This approach not only improves the overall efficiency of the network but also guarantees that the sensor nodes are not overburdened with constant data sensing and transmission tasks, which would otherwise deplete their energy resources quickly.

4.6 Routing Phase

In this phase of the network operation, an adaptive routing strategy driven by Q-learning is utilized to maximize energy efficiency during data transmission. The routing procedure varies depending on the role each node plays within the network. Cluster heads (CHs) are responsible for forwarding aggregated data directly to the sink node, ensuring that the data reaches its destination without unnecessary detours. Non-cluster head nodes, on the other hand, send their collected data to their associated CH, which then forwards it towards the sink. The routing decisions are based on a Q-learning algorithm, where each node maintains a Q-value table for its neighboring nodes. These Q-values reflect the expected cumulative reward of choosing a particular neighbor as the next hop for forwarding data. When a node is ready to transmit, it selects the neighbor with the highest Q-value, indicating the most optimal path in terms of energy efficiency.

The Q-value table is updated using the standard Q-learning update rule, which is given by [25]:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right] \tag{3}$$

Here in eqn(3), $Q(s,a)$ represents the Q-value for the current state s and action a, α is the learning rate, γ is the discount factor, r is the immediate reward obtained after taking action a, and s' denotes the next state. The immediate reward r is calculated by considering multiple routing factors, such as the residual energy E_{res} of the neighbor node, the Euclidean distance d to the neighbor, and the hop count h to the sink node. The reward function is structured as:

$$r = \beta_1 \frac{E_{res}}{E_{init}} - \beta_2 d - \beta_3 h \tag{4}$$

Here in eqn(4) $\beta_1, \beta_2, \beta_3$ are coefficients that adjust the significance of each factor, and E_{init} is the initial energy of the node. After determining the routing decision, the transmission takes place, and the energy consumption for sending and receiving data is subtracted from the energy reserves of the involved nodes. These energy costs are typically calculated using the radio energy dissipation model:

$$E_{tx}(k,d) = E_{elec} \cdot k + \varepsilon_{amp} \cdot k \cdot d^2 \tag{5}$$

$$E_{rx}(k) = E_{elec} \cdot k \tag{6}$$

Here in eqn(4) k is the size of the data packet in bits, d is the distance to the receiving node, E_{elec} is the energy consumed per bit by the transmission and reception circuitry, and ε_{amp} is the energy required for the transmitter's amplifier. Over time, as the nodes continue to interact with their environment and adjust their Q-values, the network progressively learns the most efficient routing paths that balance energy consumption and path efficiency. This dynamic learning process allows the network to adapt to changes in topology and variations in the energy levels of nodes, ensuring a robust, energy-aware communication protocol.

4.7 Monitoring and Collection Phase

During the simulation, several key performance indicators (KPIs) are continuously monitored and assessed to gain insights into the network's performance. The number of alive nodes is tracked at each round to understand the overall health of the network, indicating how many nodes are still operational and actively participating in data transmission. The total number of packets delivered successfully to the sink node is recorded to evaluate the effectiveness of the data routing mechanism. Additionally, energy consumption is measured, both in total and per round, to ensure that the network is operating efficiently and to identify any energy inefficiencies that may arise during the simulation. One important metric, the First Node Death (FND), marks the round when the first node depletes its energy, which serves as an indicator of the network's energy sustainability. The remaining energy of each node is also computed and plotted over time, providing a visual representation of energy distribution across the network.

In addition to these basic performance measures, the simulation also includes an adaptive cluster head (CH) rotation mechanism to ensure a balanced energy load among the nodes. CHs are rotated at predefined intervals, specified by the CLUSTER ROTATION INTERVAL, to prevent any single node from prematurely depleting its energy due to the increased responsibility of aggregating and forwarding data. This dynamic approach helps prolong the network's operational lifetime. At the conclusion of the simulation, all the gathered results are visualized through various plots and graphs, which allow for a comprehensive evaluation of the network's longevity, energy efficiency, and overall resilience to failures. These visual tools provide valuable insights into the effectiveness of the routing algorithm and energy management strategies, helping to determine the network's performance under varying conditions and load scenarios.

In conclusion, the simulation provides a detailed evaluation of the wireless sensor network's performance by tracking essential metrics such as node vitality, data transmission success, energy consumption, and the

timing of the first node failure. The use of adaptive cluster head rotation ensures a more balanced energy distribution, enhancing the overall longevity of the network. The visualization of energy distribution and other KPIs offers a clear picture of the network's efficiency and resilience, highlighting the effectiveness of the implemented strategies. These insights are crucial for identifying potential areas of improvement, guiding future optimizations to achieve more energy-efficient, scalable, and robust wireless sensor networks.

5 Implementation and Simulations

5.1 Simulation Interface

The proposed algorithm was simulated with the Python language. Python provides implementations of the most popular libraries of data science and machine learning algorithms, as well as various tools for visualizing the results It is a very effective platform for network and RL simulations. In the performed simulations, a NumPy library is used to properly process and structure the data. The MatPlotlib library is also a very effective tool in visualizing graphs and trends in the data.

5.2 Simulation Metrics

Our measurement criteria in the simulation are selected in such a way that we can use them to properly measure the network lifetime and to show that the proposed method can be considered an efficient protocol for controlling wireless sensor networks. Our first performance measurement metric is the **number of nodes alive over time**. This parameter is an indicator of the lifetime of the sensor network under the various routing algorithms used. The second measurement parameter in this simulation is a count of the **cumulative packets delivered by the network**. This is a strong indicator of the throughput of the network. The third parameter of the simulation is the **energy consumption** of each algorithm. To gain a fair idea, the iteration with the maximum energy used was considered. This is an indicator of the energy efficiency and thus the viability of the algorithm in a WSN.

Hyper-Parameters	**Value**
Transmission Range	100 m
Energy for Electronics	50 nJ/bit
Energy for Amplifiers	100 pJ/bit/m^2
Energy for Sensing	0.1 mJ
Energy for Processing	0.2 mJ
Q-Learning Discount Factor	0.9
Width	500 m
Height	500 m
Number of Nodes	100
Number of Clusters	5
Initial Node Energy	0.5 J
Learning Rate	0.1
Number of Rounds	7000

Table 4: Simulation Hyper-Parameters

5.3 Simulation Setup

The simulation was run under the following assumptions:

1. The network spanned over an area of 500m x 500m, with around 100 sensor nodes scattered into 5 clusters.

2. The simulation ran through 7000 iterations, i.e., 7000 time cycles, with the cluster heads rotating every 20 cycles. The nodes were also considered to function for 5 cycles, then sleep for 3.

3. Each sensor node was given a maximum transmission range of 100m, as well as an initial energy status of 0.5J. Each network packet transmitted contained 500 bits.

There were multiple other hyperparameters considered, with regards to the energy used per action, as well as some of the key values for the RL Agent, i.e., the learning rate and the discount factors. They have been listed in Table 4.

6 Results and Comparative Analysis

The graph presented in Figure 3 offers a clear visual comparison of how different routing protocols perform in terms of maintaining node availability over time. It effectively highlights the resilience and energy management capabilities of each method by plotting the number of alive nodes during the simulation rounds. From the trends observed, it's evident that the proposed routing strategy outperforms others by sustaining a higher count of active sensor nodes, especially in the later stages of the simulation. This prolonged activity suggests that the proposed method not only conserves energy more efficiently but also distributes workload more evenly across the network, thereby delaying the onset of node failures.While other approaches like DADF, LEACH, and PEGASIS perform moderately better, they still lag behind the proposed scheme in maintaining long-term network viability.

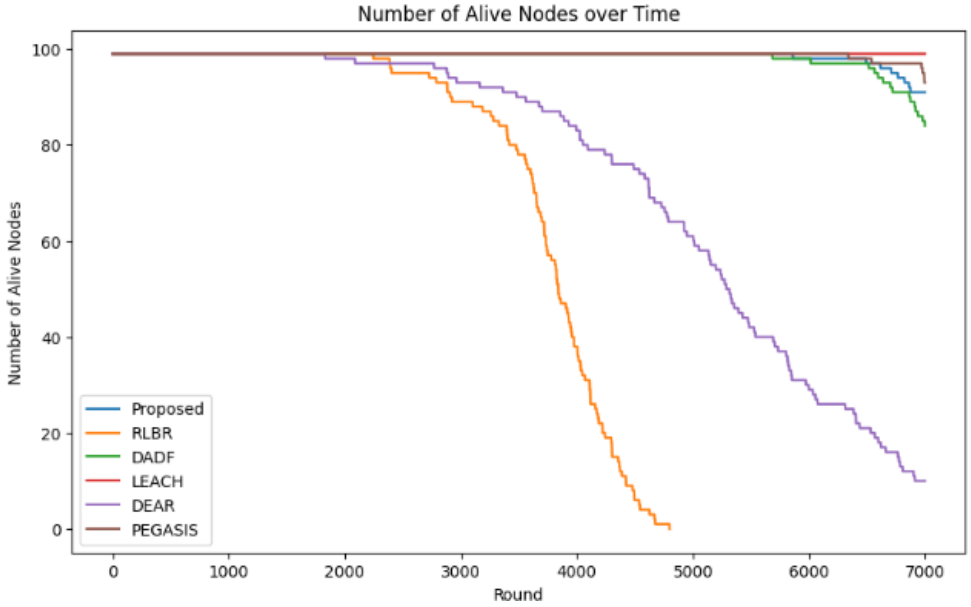

Figure 3: Number of Alive Nodes over Time

We then took a look at the comparison of throughput of the network, i.e., the cumulative packets delivered until the end of the period. It was observed that DEAR has the highest throughput by far, followed by the RLBR and then the proposed algorithm, closely followed by the remaining three approaches, as can be observed in Figure 4. It is important to note that the RLBR algorithm elapses prematurely due to the death of all the nodes in the network.

Figure 5 presents a bar chart that compares the maximum energy consumed by each routing protocol during the simulation, offering valuable insights into how efficiently each approach manages power usage under demanding conditions. Among the six strategies evaluated, the proposed method stands out by achieving the lowest peak energy consumption. This not only reflects a more energy-conscious routing behavior but also indicates that the protocol effectively balances communication tasks across nodes, preventing any one node from becoming overly burdened and quickly draining its energy.

On the other end of the spectrum, the DEAR protocol exhibits the highest energy usage, suggesting that it may struggle with load distribution or lack sufficient mechanisms to adapt to peak traffic scenarios efficiently. The remaining protocols—RLBR, DADF, LEACH, and PEGASIS—show intermediate levels of energy usage.

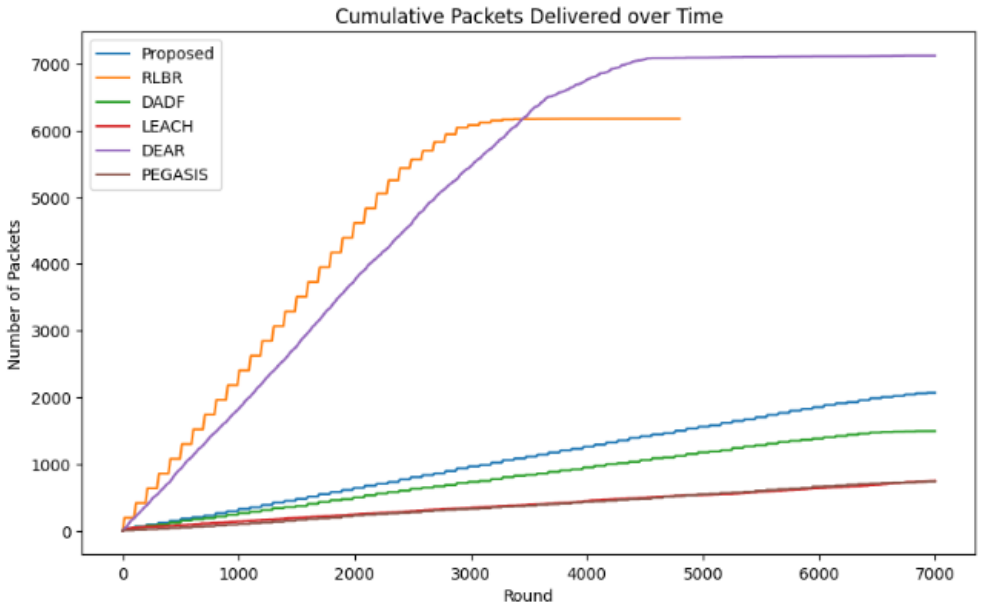

Figure 4: Cumulative Packets Delivered over Time

While these methods demonstrate moderate efficiency, none of them manage to achieve the same level of optimization as the proposed approach. This comparative analysis reinforces the strength of the proposed scheme in minimizing energy expenditure while ensuring consistent network performance, particularly when the system is under stress.

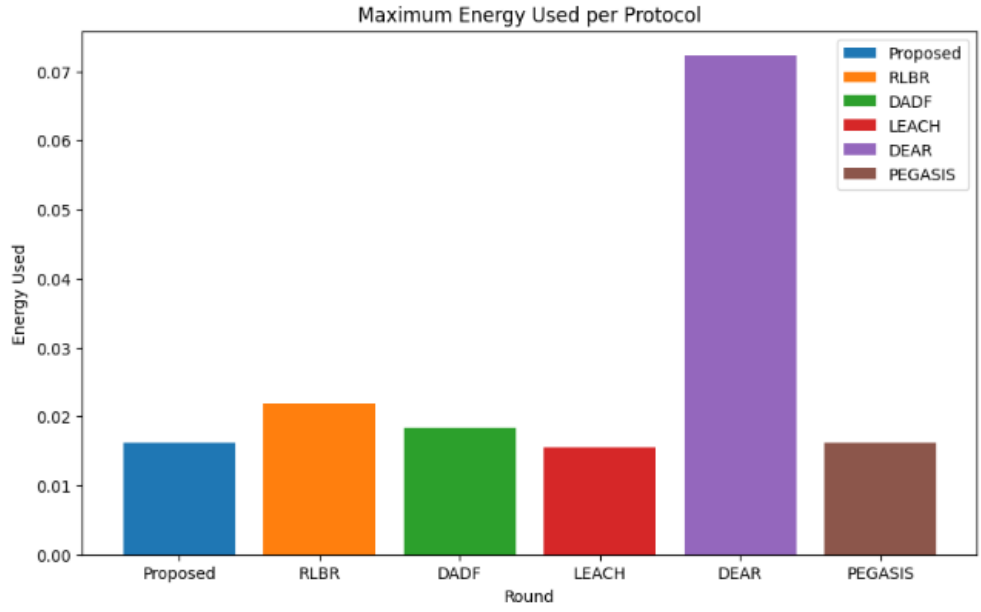

Figure 5: Energy Efficiency of the Algorithms

7 Conclusion and Potential Future Work

In this work, we have proposed a Reinforcement Learning-based Energy-Aware routing algorithm for Wireless Sensor Networks (WSNs), designed to optimize energy consumption while maintaining stable data transmission. Through extensive evaluation, the proposed method demonstrated substantial improvements in network lifetime when compared with other reinforcement learning-based routing protocols, specifically RLBR and DADF. Among the six routing algorithms tested under identical network conditions, our method consistently outperformed the rest in terms of energy efficiency, reliability, and overall network stability.

Unlike RLBR, which provides high throughput but suffers from a significantly reduced lifespan, our approach achieves a balanced trade-off by ensuring steady data flow while prolonging node activity. This performance gain is largely due to the hybrid nature of the algorithm, which combines traditional energy-efficient strategies like LEACH and PEGASIS with adaptive decision-making enabled by reinforcement learning. Moreover, the incorporation of simplified sleep scheduling techniques and minimal reliance on data fusion further contributes to energy conservation during communication, reducing unnecessary overhead and extending the operational duration of the network.

Overall, the proposed approach not only maximizes throughput and minimizes energy loss but also maintains network equilibrium by intelligently managing both transmission paths and node states. However, despite the promising results, there remains potential for further enhancement. Future work could focus on refining the reward function and learning dynamics in the reinforcement learning model, as well as improving energy management policies and transmission protocols. These enhancements would allow for even more robust, scalable, and energy-conscious WSN implementations moving forward.

References

[1] J. Al-Karaki, A. Kamal, Routing techniques in wireless sensor networks: a survey, https://ieeexplore.ieee.org/document/1368893 (2004).

[2] J. T. Moy, OSPF: The Anatomy of an Internet Routing Protocol, 8th Edition, Addison-Wesley, 2004.

[3] I. van Beijnum, Building Reliable Networks with the Border Gateway Protocol, 1st Edition, O'Reilly, 2002.

[4] C. E. Perkins, E. M. Roye, Ad-hoc On-Demand Distance Vector Routing, https://ieeexplore.ieee.org/document/749281 (2021).

[5] D. B. Johnson, D. A. Maltz, J. Broch, DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks, https://www.cse.iitb.ac.in/~mythili/teaching/cs653_spring2014/references/dsr.pdf (2014).

[6] I. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey , https://doi.org/10.1016/S1389-1286(01)00302-4 (2002).

[7] D. Ramphull, A. Mungur, S. Armoogum, S. Pudaruth, A Review of Mobile Ad hoc NETwork (MANET) Protocols and their Applications, https://ieeexplore.ieee.org/document/9432258 (2021).

[8] K. Georgiou, S. X. de Souza, K. Eder, The IoT Energy Challenge: A Software Perspective , https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8012513 (2018).

[9] W. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks , https://ieeexplore.ieee.org/document/926982 (2000).

[10] S. Lonare, G. Wahane, A Survey on Energy Efficient Routing Protocols in Wireless Sensor Network, https://ieeexplore.ieee.org/document/6726591 (2013).

[11] N. Palan, B. Barbadekar, S. Patil, Low Energy Adaptive Clustering Hierarchy (LEACH) Protocol: A Retrospective Analysis, https://ieeexplore.ieee.org/document/8068715 (2017).

[12] C. S. R. Stephanie Lindsey, PEGASIS: Power-Efficient Gathering in Sensor Information Systems , https://ieeexplore.ieee.org/document/1035242 (2003).

[13] L. Shu, Y. Niu, S. Lee, J.-U. Kim, A Distance-Based Energy Aware Routing Algorithm for Wireless Sensor Networks, https://www.mdpi.com/1424-8220/10/10/9493 (2010).

[14] M. Liu, J. Cao, G. Chen, X. Wang, An Energy-Aware Routing Protocol in Wireless Sensor Networks , https://www.mdpi.com/1424-8220/9/1/445 (2009).

[15] S. Yi, J. Heo, Y. Cho, J. Hong, PEACH: Power-efficient and adaptive clustering hierarchy protocol for wireless sensor networks, https://doi.org/10.1016/j.comcom.2007.05.034 (2007).

[16] S. Shi, X. Liu, X. Gu, An Energy-Efficiency Optimized LEACH-C for Wireless Sensor Networks , https://ieeexplore.ieee.org/document/6417532 (2012).

[17] X. Song, C. Wang, J. Wang, B. Zhang, A hierarchical routing protocol based on AFSO algorithm for WSN, https://ieeexplore.ieee.org/document/5541265 (2010).

[18] V. Kumar, A. Khunteta, Energy Efficient PEGASIS Routing Protocol for Wireless Sensor Networks, https://ieeexplore.ieee.org/document/8742817 (2018).

[19] V. Nehra, A. K. Sharma, PEGASIS-E: Power Efficient Gathering in Sensor Information System Extended, https://www.researchgate.net/profile/Vibha-Nehra/publication/276204021_PEGASIS-E_Power_Efficient_Gathering_in_Sensor_Information_System_Extended/links/5589417508ae2affe7140fe5/PEGASIS-E-Power-Efficient-Gathering-in-Sensor-Information-System-Extended.pdf (2013).

[20] Z. Mammeri, Reinforcement Learning Based Routing in Networks: Review and Classification of Approaches , https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8701570 (2019).

[21] P. K. Donta, T. Amgoth, C. S. R. Annavarapu, Delay-aware Data Fusion in Duty-cycled Wireless Sensor Networks: A Q-learning Approach , https://www.sciencedirect.com/science/article/pii/S2210537921001256 (2021).

[22] J. Hao, B. Zhang, H. T. Mouftah, Routing Protocols for Duty Cycled Wireless Sensor Networks: A Survey , https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6384460 (2012).

[23] C. J. C. H. Watkins, P. Dayan, Q-Learning , https://link.springer.com/article/10.1007/bf00992698 (1992).

[24] S. Banerji, R. S. Chowdhury, On IEEE 802.11: Wireless LAN Technology , https://arxiv.org/pdf/1307.2661 (2013).

[25] T. Hu, Y. Fei, QELAR: A Machine-Learning-Based Adaptive Routing Protocol for Energy-Efficient and Lifetime-Extended Underwater Sensor Networks , https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5408367 (2010).

[26] R. Shah, J. Rabaey, Energy aware routing for low energy ad hoc sensor networks , https://ieeexplore.ieee.org/document/993520 (2002).