

ADAPTIVE Q-LEARNING-BASED ROUTING WITH CONTEXT-AWARE METRICS FOR ROBUST MANET ROUTING (AQLR)

P. Tamilselvi ¹, S. Suguna Devi ¹, M. Thangam ² and P. Muthulakshmi ³

¹ Department of Information Technology, Cauvery College for Women (Autonomous), (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu, India.

² Department of Computer Science, Mount Carmel college (Autonomous), Bangalore

³ Department of Computer Science, Cauvery College for Women (Autonomous), (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu, India.

ABSTRACT

Mobile Ad Hoc Networks encounter persistent challenges due to dynamic topologies, limited resources and high routing load. As these problems continue, the network's overall performance declines as the network scales. To address these challenges Adaptive Q-Learning-Based Routing with Context-Aware Metrics for Robust MANET Routing (AQLR), a routing protocol that uses context-aware data and reinforcement learning to choose the best route for connected mobile devices. AQLR considers four essential routing metrics such as Coverage Factor, RSSI-Based Link Stability, Energy Weighting and Broadcast Delay. AQLR uses Q-learning agent at each node to enable adaptive learning of optimal next-hop decisions based on past history. Composite Routing Metric (CRM) helps to obtain smart decision in the absence of prior learning. Simulation performed with OMNeT++ across varying node densities from 50 to 500 the simulation results shows that AQLR outperforms recent machine learning-based routing protocols including QLAR, RL-DWA, and DRL-MANET. Specifically, AQLR achieves up to 95.8% packet delivery ratio, reduces average end-to-end delay by 25–35%, lowers routing overhead by 20–30%, and improves network lifetime by over 15% in dense scenarios. These results affirm the effectiveness of combining reinforcement learning with context-aware metric computation for scalable and energy-efficient MANET routing.

KEYWORDS

Q-Learning, Context-Aware Metrics, Coverage Factor, Link Stability, Energy Aware Routing, Broadcast Delay, Reinforcement Learning, Adaptive Protocols

1. INTRODUCTION

Mobile Ad Hoc Networks (MANETs) consist of wireless mobile nodes that self-organize to create communication network without any centralized infrastructure [5]. They are inherently very flexible, and thus they can be used in cases such as military applications, disaster recovery, and in vehicular networks [2] where time is of the essence and no infrastructure is available. But, despite these advantages, several inherent characteristics of MANETs like the dynamic topology, the limited energy availability and the limited transmission range, make routing reliably and efficiently a non-trivial problem [7]. Together these elements cause variations in the link quality that are unpredictable, frequent breaks in the routes, an increase in routing overhead, and overall non-optimized performance of the network as it grows. The majority of current routing protocols like AODV, DSR, or OLSR adopt fixed or reactive strategies and use simple metrics (mostly hop count) for route selection. These protocols perform well in static or moderately dynamic scenarios, but tend to struggle under high mobility, link stability variations, or energy imbalance in the

nodes. As a result, these schemes lack learning capabilities and fail to adapt based on past outcomes [3][7][8], thus being unable to improve either the path selection or to decrease Control Messages that are duplicated. Thus these procedures suffer from broadcast storms, premature energy depletion of the nodes as well as poor route discovery in a highly dynamic environment [6]. Towards addressing such limitations, this contribution presents AQLR, which aims at intelligently controlling routing in MANETs through Q-learning and an entire set of metrics that take performance into account [1][4]. AQLR routing protocol embeds four fundamental measurements: Coverage Factor (CF) to prefer nodes that broaden the coverage of the network, RSSI Based Link Stability (LS) to give preference to stable links, Energy Weight (EW) to have an even participation of the nodes and to maximize the overall network lifespan and Broadcast Delay (BD) to reduce rebroadcasting overhead and control routing overhead. Each node employs a Q-learning agent to learn the best routing action as a function of the state, action and reward associated with it, which enables the protocol to be self-adaptive over time. A Composite Routing Metric (CRM) fallback mechanism during the initial learning phase always leads to intelligent and redundant paths being selected even when the Q-table has not been completely populated.

The major contributions of this work are:

- A multi-metric multi-agent-based machine learning approach to MANET routing.
- Deploying Q-learning agents at each node which learn dynamically and update optimal next-hop choices based on feedback from the environment.
- A stable backup method for choosing CRMs based on experience when not enough data available for learning.
- A scalable and lightweight protocol design as performance evaluated by means of OMNeT++ showing a higher packet delivery ratio, lower end-to-end delay, lower routing overhead and higher energy efficiency when compared with current state-of-the-art machine learning-based routing protocols.

The remainder of this paper is structured as follows: section 2 surveys related work on routing in MANETS and use of machine learning techniques. The methods such as the definition of metrics, formulation of states, reward prediction, and algorithm for making routing decisions are described in section 3. In section 4 the simulation set up is described as well as performance results compared with benchmark protocols. Finally, Section 5 concludes the paper and provides ideas for future work.

2. RELATED WORKS

To enhance network routing decision-making, Serhani et al. (2016) developed QLAR, a technique that leverages reinforcement learning. For best results, QLAR uses real-time feedback learning to dynamically adjust network routing policies but lacks in the usage of multiple metrics. [1]. RL-DWA [2] enhances routing with deep reinforcement learning and unsupervised learning to minimize update costs, yet it does not incorporate energy or link stability metrics explicitly. DRL-MANET [3] leverages multi-agent deep Q-learning to support real-time, scalable routing in highly dynamic topologies, although it introduces computational overhead. Deep ADMR [4] focuses on anomaly detection using deep learning to improve routing security, but gives less emphasis to energy efficiency or rebroadcast control. Kumar and Mallick [5] outline MANET challenges like limited energy, mobility, and bandwidth, supporting the need for intelligent protocols. Lin et al. [6] emphasize the importance of MANETs in D2D and IoT communications, making routing reliability critical. Sharma and Sahu [7] propose a hybrid reinforcement learning approach for QoS and energy-aware routing, while Zeng et al. [8] provide a comprehensive review of deep reinforcement learning for wireless network routing. Through the Symbionts Search Algorithm (SSA), Tabatabaei et.al. introduce a novel MANET routing technique that improves routing

performance by leveraging natural symbiotic behaviors. Through adaptive path selection, which reduces energy consumption and improves transmission reliability, the technique offers network stability [9]. Abdullah et.al. proposed an improved On-demand Distance Vector routing protocol. The improved protocol takes link quality and node mobility into account to improve route stability. The enhancement offers mobile environments greater adaptability, guaranteeing reliable communication even in the event of abrupt topology changes [10]. Hamad et.al. designed a routing framework that combines reliable connectivity with energy sustainability. The goal of the work is to show how energy-aware routing increases network endurance while preserving dependable data transmission by striking a balance between energy efficiency and link stability [11]. To create adaptive, energy-efficient routing strategies for MANETs, Chettibi and Chikhi investigated the use of reinforcement learning in conjunction with fuzzy logic, fuzzy logic allows the protocol to handle ambiguous information when making decisions [12]. Sumathi et.al.[13] use the nature-inspired Grey Wolf Optimization (GWO) technique to reduce routing latency. It chooses routes that minimize transmission delay while optimizing factors like energy consumption and route stability. To maximize network efficiency Priyambodo.et.al. [14] examine many MANET routing protocols. Ali et al. suggested the LSTDA routing protocol for Flying Ad-Hoc Networks (FANETs) [15]. Two main issues are high mobility, and unpredictable link stability, which are taken into account to attain reliable communication in aerial networks. The routing protocol finds stable links to minimize latency, which lowers transmission delay and allows for reliable communication. A resource management framework to promote energy efficiency of mobile ad hoc clouds dedicated to mobile cyber-physical system applications is introduced by Shah [16]. The framework improves resource utilization, application efficiency and energy preserving features in dynamic mobile environments by optimizing the parameters on the network layer and middleware layers. Muneeswari et.al. integrate secure routing with energy efficient clustering approaches by utilizing reinforcement learning in a 3D MANET. The approach leverages the implications of clustering routers to improve reliability and security while maintaining energy efficiency, which is of importance in sensor networks [17]. Abdellaoui et al. provide a multi-criteria algorithm for optimal multipoint relay (MPR) selection problem in MANETs. MPRs are selected in the algorithm based on several desirable properties including link quality, stability, and energy consumption, that improve both throughput and stability of the network [18]. Vishwanath Rao et al. apply reinforcement learning to develop a protocol aimed at improving MANET energy efficiency. Their work demonstrates the effectiveness of machine learning methodologies to significantly alter routing decisions as a function of the network state, achieving remarkable energy efficiency with no negative impact on performance [19]. Alsalmi et al. explore the application of deep reinforcement learning (DRL) for the enhancement of mobile wireless sensor networks (WSNs). The DRL approach can incorporate prior knowledge and thus, optimize energy and throughput efficiently and flexibly which is beneficial for the deployment of sensor networks [20]. Maleki et.al. present routing in self-sustaining nodes in MANET, which uses a predictive model to integrate energy replenishment into the decision-making process. The routing decision follows a learning paradigm that continuously monitors the available energy of nodes. [21]. In wireless sensor systems most, effective routes are identified by applying a game theory-based algorithm for inter cluster node communication. Interaction between the nodes immediately notifies the most effective routes which improves the performance of the network and conserves the energy [22]. Due to dynamic topology and frequent link failures occurs which may result in data loss and degrade network performance, to address this issue, Lin and Sun designed a routing protocol that places a high priority on choosing reliable links to ensure consistent communication [23]. Salim and Ramachandran considered the link quality as a crucial factor for maintaining stable communication paths, particularly in highly mobile environments. The proposed protocol prioritizes the most efficient routes based on link stability, thus boosting data communications efficiency [24]. Ahirwar et.al., introduce a chaotic gazelle routing protocol whereby including chaotic mechanisms into the optimization mechanism optimizes routing in MANETs. Based on network conditions, this technique dynamically changes the routing path; as such, it enhances

security and energy efficiency [25]. Pandey and Singh (2024) proposed a modified RREQ broadcasting scheme, ASTMRS [26] uses a fitness-based threshold to reduce routing overhead in multipath MANET routing. While it improves PDR and reduces delay compared to AODV/AOMDV, it lacks adaptive learning and relies on static metric thresholds. The paper proposes a hybrid QRFSODCNL model combining deep convolutional neural learning, quantile regression, and artificial fish swarm optimization for routing in the Internet of Vehicles (IoV). While it improves packet delivery and reduces delay, its reliance on deep learning and swarm-based optimization introduces high computational overhead [27].

2.1. Motivation for the Proposed Protocol

Though recent solutions use machine learning to improve routing in MANETs, the vast majority of them present significant drawbacks. Most of them depend on single metrics like hop count, or residual energy, and do not account for the multiple aspects of the dynamic nature. Some use centralized models or models that require a lot of computational power, such as deep reinforcement learning, which are not suitable for resource constrained nodes. In addition, many routing protocols Q-learning and others fail to consider important features namely broadcast suppression, variable link stability or coverage awareness of the nodes that lead to a greater routing overhead or even unstable paths. On top of that, their approaches fail to provide a strong back-up when the learning model is not trained enough, thus producing very poor performance in the initial phases of route discovery. Such limitations highlight the existence of a lightweight, fully distributed, context-aware learning protocol capable of considering intelligently multiple performance metrics and capable of adapting in real-time to network modifications.

3. PROPOSED METHODOLOGY

AQLR is a multi-metric, Q-learning-based routing approach for dynamic mobile ad hoc networks (MANETs). This approach combines real-time metric analysis with reinforcement learning to allow intelligent and adaptive route planning in a dispersed and resource-limited environment. Every mobile node is an independent agent able to learn from its local surroundings via feedback-driven interactions. State representation in the Q-learning paradigm is based on four important network statistics: Coverage Factor (CF), Link Stability (LS), Energy Weight (EW), and Broadcast Delay (BD). It supports scalability, energy efficiency, and route reliability by allowing each node to separately assess routing decisions and update it based on learned Q-values and rewards. The architecture of AQLR is illustrated in Figure 1.

The following section describe AQLR's adaptive behaviour driven by its internal components, decision-making mechanism, and dynamic feedback system.

3.1. Network Environment

The AQLR protocol is designed to operate in a MANET scenario where the network topology is continuously changing due to the normal mobility patterns of nodes and existing signal interference. Nodes use mobility models such as Random Waypoint or RPGM, where the connectivity on the network is constantly changing. RSSI is a measure of signal strength which also varies depending on environmental factors and thus contributes to reliability of the interconnection. Each node sends out the periodic signals to communicate with its neighbors to identify which other nodes are within its range. These environment states generate dynamic inputs to the Q-learning agent's state secretariat at each node.

3.2. Node i (Agent)

Each node acts as an intelligent agent composed of three critical modules, Metric Manager, Q Agent, and Routing Core, which work together to enable adaptive routing. The metric computations are illustrated in Figure 2.

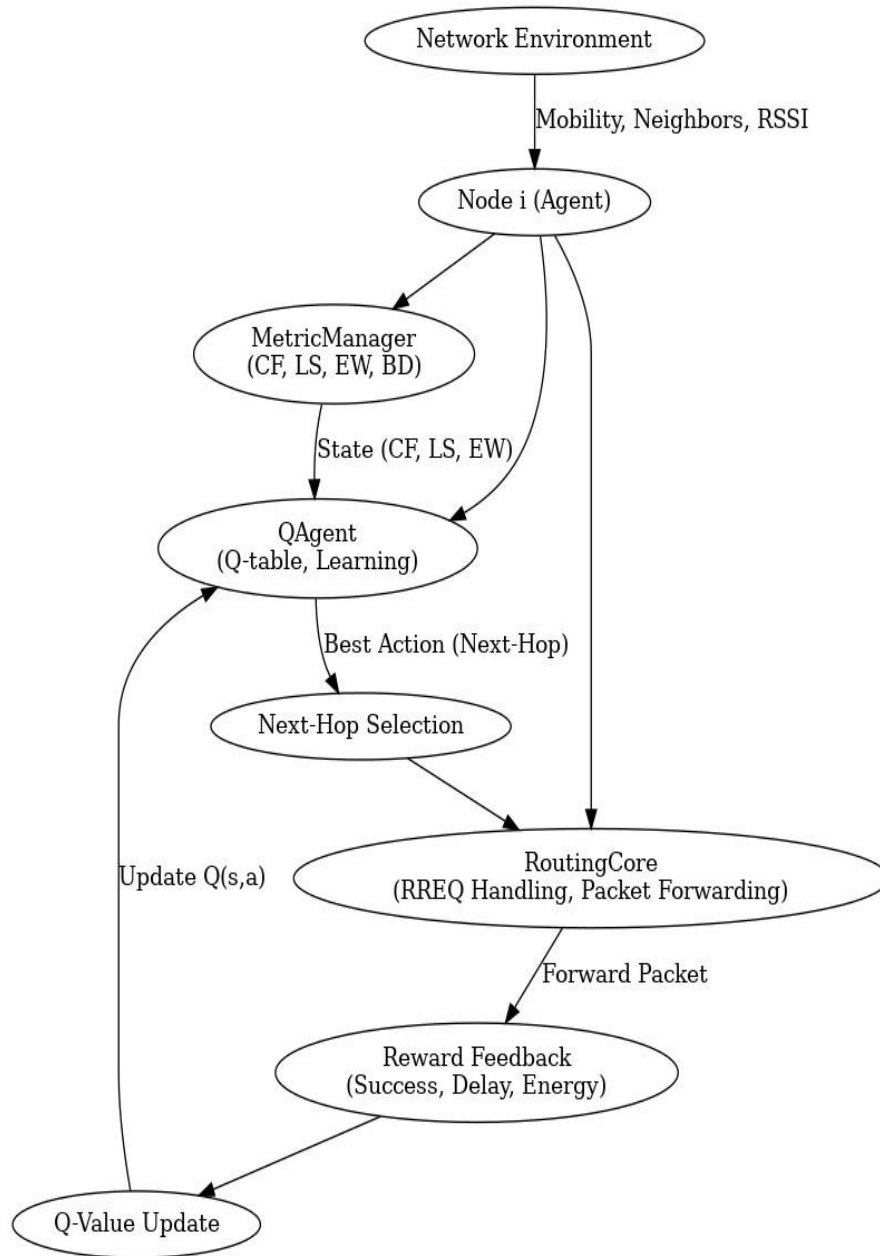


Figure 1: Functional Design of AQLR

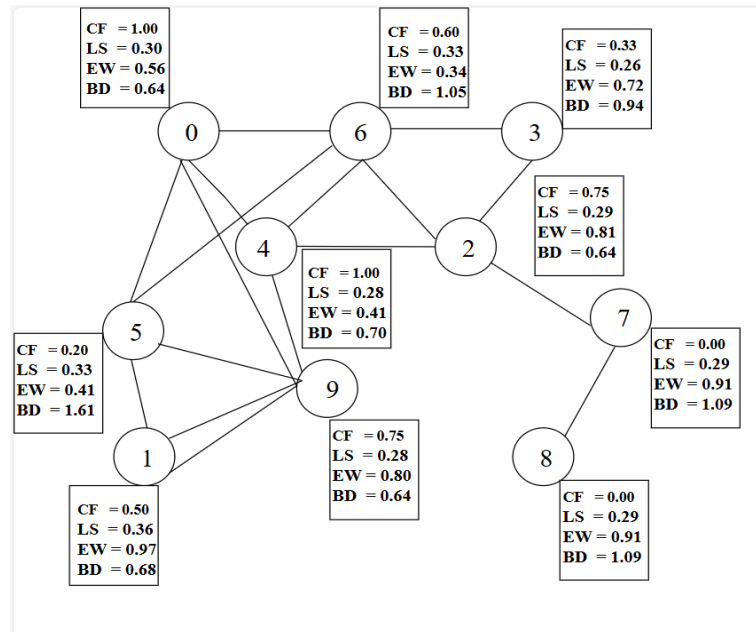


Figure 2. Metric Computation for Coverage Factor, Link Stability, Energy Weight and Broadcast Delay

3.2.1. Metric Manager

Metric Manager is responsible for gathering local and neighbour metrics that determine the current status of a node. The coverage factor (CF) is calculated using Eq.1

$$CF_i = \frac{N_{uncovered}(i)}{N_{Total}} \quad (1)$$

$N_{uncovered}(i)$ refers to the uncovered neighbours that node i can reach and potentially rebroadcast. The total set of neighbours of a node (N_i) represents all nodes within its communication range. CF_i range exist from 0 to 1. A higher CF value means the node can reach more uncovered neighbours and is prioritized for rebroadcast. The link stability (LS) is based on the signal strength variation, which represents the quality of the link. Reliability of the connection is estimated Eq.2.

$$LS_i = 1 - \frac{\sigma(RSSI_i)}{RSSI_{max}} \quad (2)$$

as the Standard deviation of the Received Signal Strength Indicator $\sigma(RSSI_{i,j})$ divided by the maximum receiving signal strength $RSSI_i$. LS_i attains a normalized value ranging from 0 to 1. Energy Weight of node (EW) normalizes the energy level of a node by dividing its current energy ($E_{residual}(i)$) by the maximum initial energy ($E_{initial}(i)$) represented in Eq.3

$$EW_i = \frac{E_{residual}(i)}{E_{initial}(i)} \quad (3)$$

The normalized value ranges from 0 to 1. The value 1 indicates that the node has maximum energy or more remaining energy, and the value 0 indicates that the node is nearer to depletion. The higher value holding nodes are prioritized for forwarding packets to enhance the network's life span. The lower value holding nodes are avoided to prevent premature energy depletion.

which helps to avoid excessive use of nodes which depletes the energy. To minimize unnecessary retransmission, the Broadcast delay (BD) strategy is designed to ensure that the most influential nodes only offer robocasting of RREQ packets. BD_i is dynamically adjusted based on the values of coverage factor and energy weight the broadcast delay is defined as using Eq.4

$$BD_i = \frac{K}{CF_i + EW_i + \epsilon} \quad (4)$$

The delay value shorter or longer. Thus, a node with high CF and residual energy will have a shorter delay and will rebroadcast quicker, resulting in a more responsive network. Nodes, on the other hand, that have low CF and energy, have long delays thus avoiding unnecessary retransmissions. The delay adopted is scaled by a constant K so that it does not get too large and ϵ is employed to avoid division by zero. The advantage of this is that it avoids duplication of information, saves energy, and enhances the lifetime of the network as well as the efficacy of routing.

3.3. Q Agent

To improve adaptive decision making in dynamic environment Q learning based reinforcement learning module is introduced. Each node is equipped with an autonomous agent called Q Agent that learns the routing behaviour based on the metric value observation. Q Agent keeps track of Q-Table that maps state-action pairs (s, a) to expected rewards $Q(s, a)$. The state values (S) are learned values such as (Coverage Factor, Energy Weight, Link Stability, delay), action (A) , corresponds to selecting a next-hop neighbor, in a particular state, which is defined by the routing metrics. A ϵ -greedy action is used next-hop selection. With ϵ probability it explores a random neighbor, otherwise takes the action for exploitation using the Eq.5

$$a = \arg \max_a Q(s, a) \quad (5)$$

If the packet is forwarded, then the agent gets some feedback (reward) and the Q-values are updated according to the standard update rule using Eq.6.

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma * \max_{a'} Q(s', a') - Q(s, a)] \quad (6)$$

where α is the learning rate, γ is the discount factor, r is the immediate reward (e.g., successful packet forwarding or minimized delay), and s' is the resulting state after action.

3.4. Routing Core

The Routing Core handles the operational part of processing RREQ and RREP messages, deciding next-hop based on outputs of the Q Agent, and keeping routes when necessary due to mobility or link failure. It is the bridge between the packet processing at protocol level intelligent decision-making that Q Agent can provide.

3.5. Next-Hop Selection

If the node needs to forward a packet, then it selects the next-hop node according to action selection based on the Eq.5. If no experience has been recorded for state in the Q-table, the node falls back to heuristic-based Composite Routing Metric (CRM), calculated as Eq.7.

$$CRM_i = W_1 \cdot CF_i + W_2 \cdot LS_i + W_3 \cdot EW_i + W_4 \cdot \frac{1}{BD_i} \quad (7)$$

Where W_1 , W_2 , W_3 and W_4 are weights assigned to each metric. This guarantees that nodes will still make informed decisions when they encounter states with which they have no prior experience.

3.6. Reward Feedback

Upon forwarding a packet, the node receives feedback based on the packet's delivery outcome. If the packet arrives successfully at its destination, then a positive reward r is assigned to the Agent, which promotes the routing path that was taken. On the other hand, negative rewards are given in the event of link failures, packet drops, or delays, leading to the eventual penalty of the Q-values of bad actions over time. This feedback is used to update the Q-table and refine future routing actions.

3.7. Q Learning Update and Learning Parameters

Q-values are adjusted during the learning process using the Eq.6. The use of distributed and dynamic learning techniques allows AQLR to improve the routing performance adaptively in real time environment. The pseudocode below outlines the routing decision logic executed at each node. The pseudocode (Figure.3) outlines the routing decision logic executed at each node.

Input: $N_i \leftarrow$ Set of 1-hop neighbors, Metrics: CF, LS, EW, BD, Q-table $Q(s, a)$
 $\epsilon \leftarrow$ exploration rate. $\alpha \leftarrow$ learning rate, $\gamma \leftarrow$ discount factor
Output: nextHop \leftarrow selected neighbor node

```

1: function AQLR_Routing ()
2:   for each neighbor n in  $N_i$  do
3:     Compute CF(n), LS(n), EW(n)
4:     Form state  $s \leftarrow [CF(n), LS(n), EW(n)]$ 
5:   end for
6:   if random () <  $\epsilon$  then
7:     nextHop  $\leftarrow$  random neighbor from  $N_i$  // Exploration
8:   else
9:     for each neighbor n in  $N_i$  do
10:      Evaluate  $Q(s, n)$  // Q-value for state-action pair
11:    end for
12:    nextHop  $\leftarrow$  neighbor with highest  $Q(s, n)$  // Exploitation
13:  end if
14:  Send RREQ or Data Packet to nextHop
15:  Receive reward r based on:
    - Delivery success (+1)
    - Delay minimization
    - Energy-aware forwarding
16:  Update  $Q(s, \text{nextHop})$  using:
     $Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma * \max_{a'} Q(s', a') - Q(s, a)]$ 
17:  return nextHop

```

Figure 3: Pseudocode of routing decision logic executed at each node

4. RESULTS AND DISCUSSIONS

4.1. Comparative Feature Analysis

A comprehensive comparative description between the main characteristics of Adaptive Q-Learning-based Routing AQLR, and the other routing protocols QLAR, RL-DWA, and DRL-MANET is represented in Protocol Feature Comparison Table 1. This comparison is intended to highlight the specific design features and improvements that allow AQLR to outperform others in highly dynamic Mobile Ad Hoc Network (MANET) scenarios. While traditional RL approaches usually select a single metric as input and they are less aware of the context of the decision to be made, AQLR takes advantage of a multi-metric model that integrates information about the energy of the network, the delay and the stability of the link into a composite Q-learning framework. Ablation analyses were performed to assess the contribution of each of the CF, LS, EW and BD metrics in AQLR's performance. Having CF disabled increased routing overhead due to the uncontrolled retransmissions, and permanent LS implied unstable routes with higher delay. Failure to implement EW resulted in asymmetric energy consumption and decreased network lifespan. If BD was not included, the number of collisions increased and the PDR decreased. Together, these findings validate the contribution of each measure and the importance of their collective use to the impact of AQLR.

Table 1: Comparative Feature Analysis of AQLR and Baseline Protocols'

| Feature | AQLR | QLAR | RL-DWA | DRL-MANET |
|----------------------------|------------|------------|----------|-----------|
| Learning Model | Q-Learning | Q-Learning | RL | Deep RL |
| Multi-Metric Input | Yes | No | Partial | Partial |
| Energy Awareness | Yes | No | Yes | Yes |
| Broadcast Delay Adaptation | Yes | No | No | No |
| Composite Metric Fallback | Yes | No | No | No |
| Scalability (Node Density) | High | Moderate | Moderate | Moderate |

4.2. Computational Efficiency Analysis

The computational complexity of AQLR is much lower than that of DRL-MANET. DRL based approaches require continuous training using deep neural network which demand higher memory, processing power and energy consumption, AQLR only needs to periodically update a simple and light-weight Q-table. In AQLR, the Q-learning algorithm state-action updates take $O(1)$ time, while the next-hop selection process is $O(N)$ where N represents the number of one-hop neighbours. This straightforwardness also makes AQLR lightweight enough to be executed by low power MANET nodes, without the need for GPU acceleration and/or external training data. Therefore, AQLR is better adapted for on-the-fly use in a highly dynamic, resource-constrained environment such as in tactical/emergency ad hoc networks.

4.3. Experimental Setup

The result of AQLR and other conventional methods is implemented in OMNeT++ (Version 6.0.3) network simulator. Table 2 provides a comprehensive view of the simulation setup for AQLR.

Table 2: Simulation Setup

| Parameter | Value |
|-------------------------|-------------------------|
| Simulation Tool | OMNeT++ (version 6.0.3) |
| Network Area | 1000m X 1000m |
| Total Nodes | 50 – 250 |
| Mobility Model | Random Waypoint Model |
| Node Speed | 1 – 10 m/s |
| Pause Duration | 0-5 s |
| Transmission Range | 250 m |
| Traffic Pattern | Constant Bit Rate |
| Data Packet Size | 512 Bytes |
| Transmission Interval | 1s |
| Propagation Model | Two Way Ground |
| Simulation Duration | 300s |
| Initial Energy per Node | 100J |

Considering the need of fine-tuning parameters in order to strike the right balance between Exploration and Exploitation in respect to good adaptability and convergence speed, the Q-learning parameters were set as $\alpha=0.5$ (learning rate), $\gamma=0.9$ (discount factor), and $\epsilon=0.1$ (exploration rate), which were chosen through empirical tuning based on stability and convergence criteria. The simulations were executed with varying seeds, and results are reported as mean with standard deviation in order to be statistically robust. AQLR's performance was then compared to that of QLAR, RL-DWA and DRL-MANET, all trained and tested under the same conditions.

4.4. Performance Metrics and Measurement

In order to prove the efficiency of the AQLR routing protocol obtained results are contrasted with three recent machine learning routing protocols QLAR, RL-DWA and DRL-MANET. It presents an assessment using five parameters; Packet Delivery Ratio (PDR), End-to-End Delay, Routing Overhead, Energy Consumption and Network Lifetime at different node densities between 50 and 500 nodes. To ensure statistical reliability, all performance graphs include standard deviation (SD) bars derived from multiple simulation runs. These bars represent the variability or spread of performance values around the average (mean). A smaller SD bar indicates consistent behaviour across different trials, while a larger SD suggests more variability in results. Including SD bars strengthens the experimental validity by illustrating that observed trends are not artefacts of isolated simulations, but reflect repeatable, dependable outcomes.

4.4.1. Packet Delivery Ratio (PDR)

The reliability and effectiveness of data transmission are evaluated using the Packet Delivery Ratio parameter. It is calculated using Eq. (8) as the percentage of data packets that are successfully delivered to the intended destination of the total number of packets originally sent by the source

$$PDR = \frac{\text{Total packets received at the destination}}{\text{Total packet sent by the source}} \times 100 \quad (8)$$

AQLR demonstrates the highest PDR among all other protocols in all node densities illustrate in Figure 4, with a maximum of 95.8% in a 50-node density, and 89.6% in dense 500-nodes scenario. On the contrary, DRL-MANET and RL-DWA experience lower delivery ratios in denser scenarios, as the number of collisions and broadcast storms tends to increase. AQLR which applies Q-learning through metrics that consider context information, such as CF and LS, selects routes in a more stable way and performs a more efficient control of rebroadcasting packets, leading to less dropped packets.

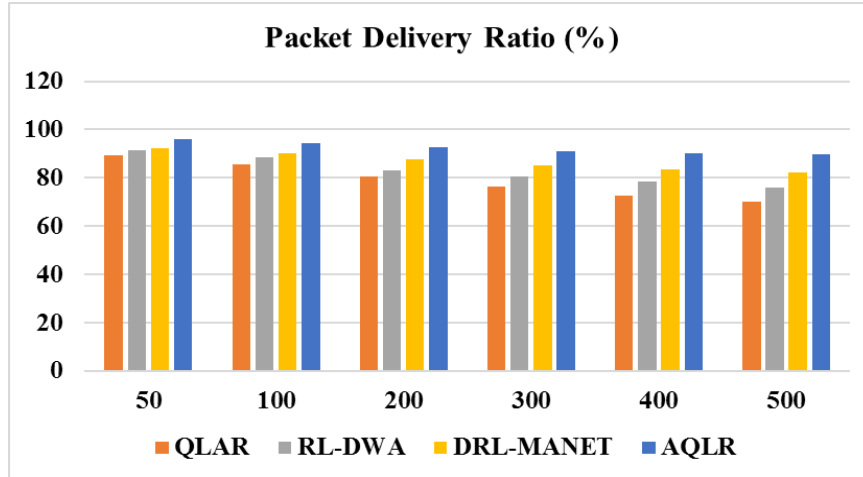


Figure 4: Packet Delivery Ratio Vs Node Density

The quantitative variation in Packet Delivery Ratio is visually illustrated using error bars in the Figure 5. AQLR consistently maintains higher delivery rates with narrower deviation bands ($\pm 0.5\%$ to $\pm 0.8\%$), while QLAR and RL-DWA exhibit wider error margins (up to $\pm 1.4\%$), indicating more unstable performance.

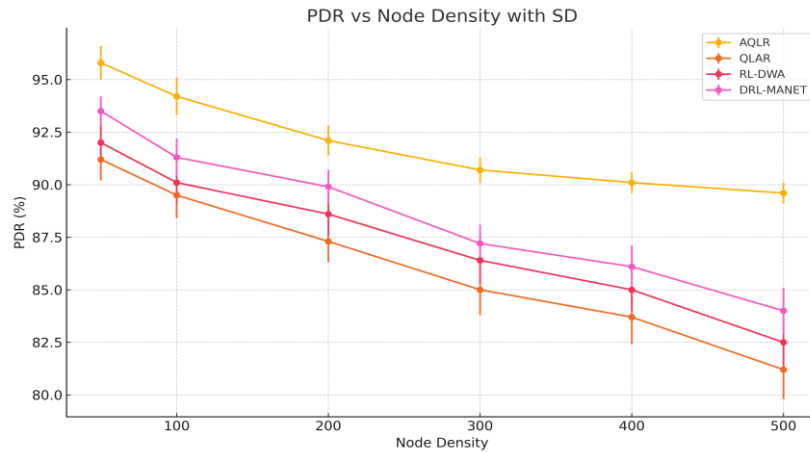


Figure 5: Packet Delivery Ratio Vs Node Density with Standard Deviation

4.4.2. End-to-End Delay

It Measures the average time taken for packets to travel from source to destination. It is calculated using the Eq. 9 and measured as milliseconds(ms)

$$\text{End-to-End Delay} = \frac{\sum (\text{Packet Arrival Time at Destination} - \text{Packet Generation Time at Source})}{\text{Total packets Received}} \quad (9)$$

AQLR achieves the lowest average end-to-end delay among all methods, but the difference with some of the baselines is not bigger than 50% shown in Figure 6. For example, with 300 nodes AQLR has a delay of 101ms, while DRL-MANET and RL-DWA have delays of 127ms and 140ms respectively. The reason for this is that broadcast delay metric (BD) favours to forwarders that are efficient and makes these redundant RREQs to be less broadcasted, reducing thus congestion at MAC level and queuing delays.

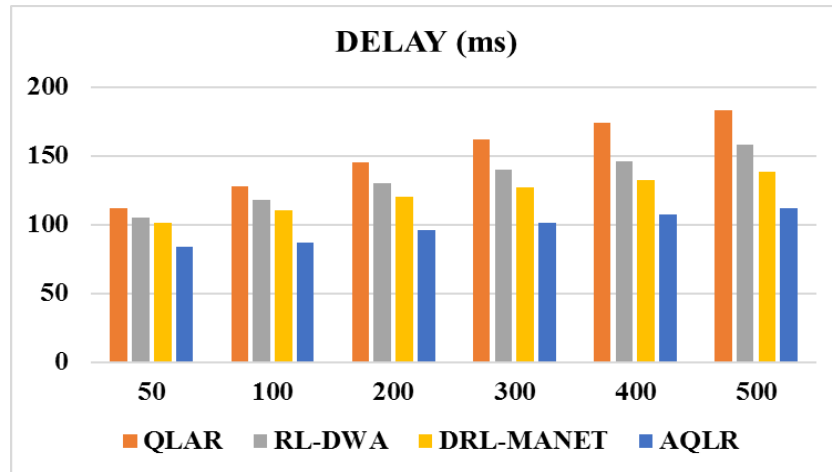


Figure 6: End – to – End Delay Vs Node Density

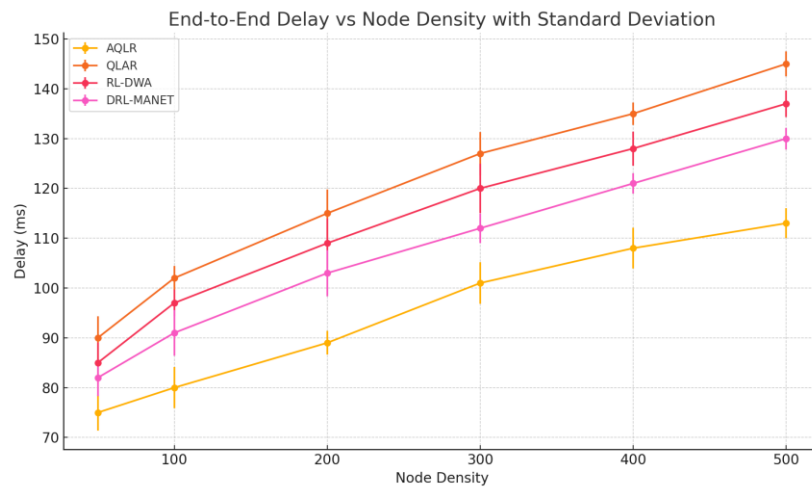


Figure 7: End – to – End Delay Vs Node Density with Standard Deviation

The error bars for delay in Figure 7 demonstrated that AQLR achieves consistent low delay with error margins under ± 5 ms, indicating predictability in data transmission time. In contrast, QLAR and DRL-MANET displayed higher variability ($\pm 6-8$ ms), pointing to performance inconsistency under mobility.

4.4.3. Routing Overhead

Routing overhead measures the efficiency of routing by evaluating the proportion of control packet transmissions relative to data packet delivery. It is computed using the formula (10).

$$\text{Routing Overhead} = \frac{\text{Total Control Packets Exchanged}}{\text{Total Data Packets Delivered}} \quad (10)$$

AQLR also has the least routing overhead in the entire testbed shown in Fig 8. For the case of 500 nodes, the control packets overhead of AQLR is 790 control packets per successfully delivered data packet, while QLAR and RL-DWA impatient have 1180 and 1060 control packets respectively.

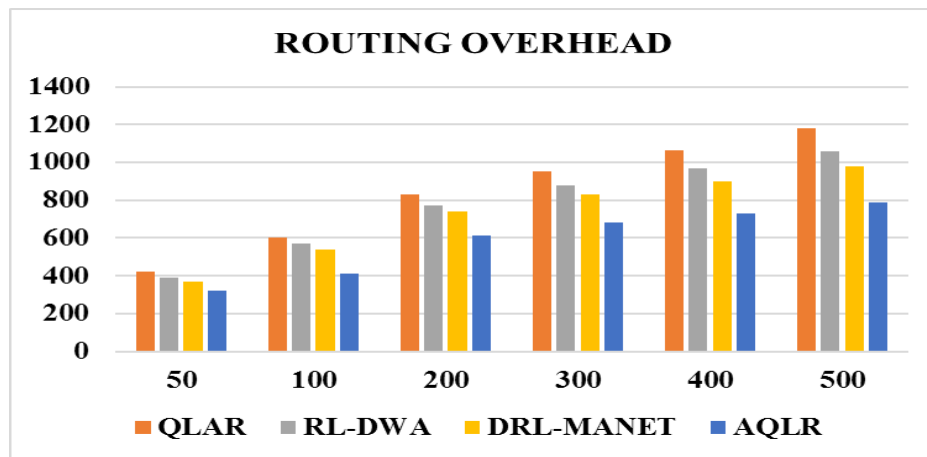


Figure 8: Routing Overhead Vs Node Density

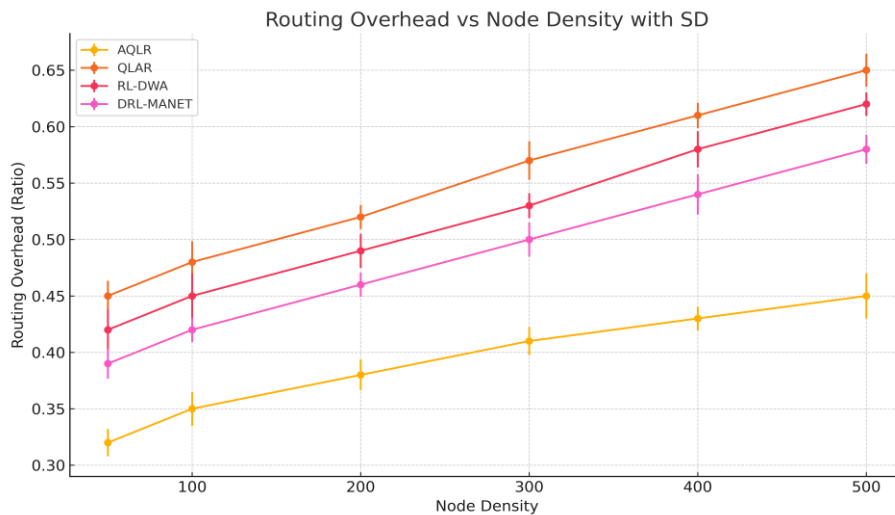


Figure 9: Routing Overhead Vs Node Density with Standard Deviation

The Q-learning model controls EW and CF metrics that effectively limits the participation of the non-ideal nodes to spare control traffic. Error bars for routing overhead shown in Figure 9 measure fluctuations in control message load. AQLR showed low overhead variability (± 0.01 –

0.02), whereas QLAR and RL-DWA had larger spreads (up to ± 0.03), indicating less efficient control packet management under stress.

4.4.4. Energy Consumption

Energy consumption measures the total energy expended by all nodes for transmitting, receiving, and processing data packets. It is determined using the eq. (11) and measured as Joules (J).

$$\text{Energy} = \sum_{i=1}^N (E_{\text{initial},i} - E_{\text{residual},i}) \quad (11)$$

An important benefit of AQLR is energy efficiency. The nodes preselected for routing are filtered in terms of remaining energy EW, which prevents the low energy nodes being transmitted. AQLR showed in Figure 10 a consumption of 785 J with 500 nodes, while DRL-MANET used 860 J and QLAR exceeded 950 J. Such energy-aware design directly contributes to the sustainability of the network in the long run.

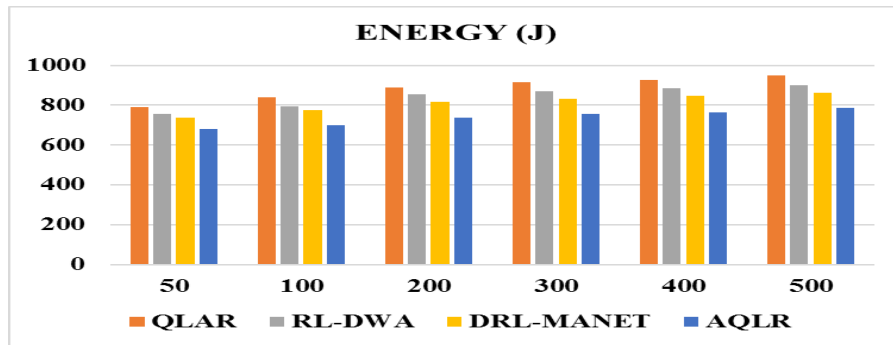


Figure 10: Energy Consumption Vs Node Density

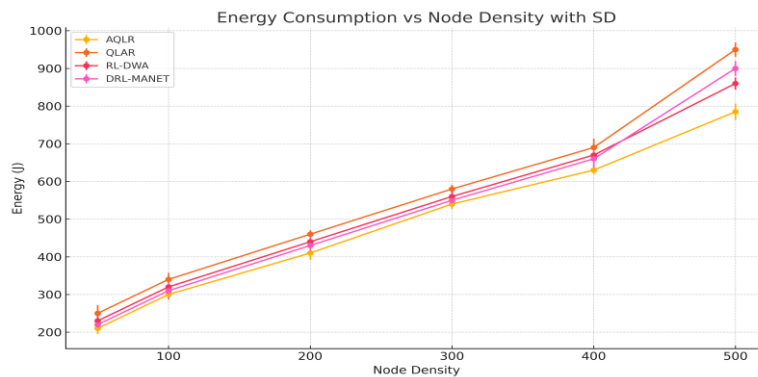


Figure 11: Energy Consumption Vs Node Density with Standard Deviation

Variation in energy usage, as shown by error bars in Figure 11, reflects differences in routing path lengths and node participation. AQLR had narrower bars ($\pm 3\text{--}4\%$), indicating efficient energy balancing, while QLAR and RL-DWA had broader spreads (up to $\pm 6\%$), revealing less uniform energy depletion across the network.

4.4.5. Network Lifetime

It is defined as the time until the first or last node exhausts its energy resources.

$$\text{Life time} = t_{\text{end}} - t_{\text{start}} \quad (12)$$

t_{start} indicates simulation start time and t_{end} represents time when the last node dies. Longer network lifetime indicates better energy-aware load balancing. AQLR has a longer network lifetime, which is the time in which the first node dies, as a result of balanced use of their energy and intelligent use of routes. In the scenario with high density, the AQLR increases the lifetime as well, achieving 930 seconds, higher than the 830 s of DRL-MANET and 800 s of the RL-DWA shown in Figure 8. This reveals that the protocol successfully avoids energy holes and ensures uniform participation of the nodes.

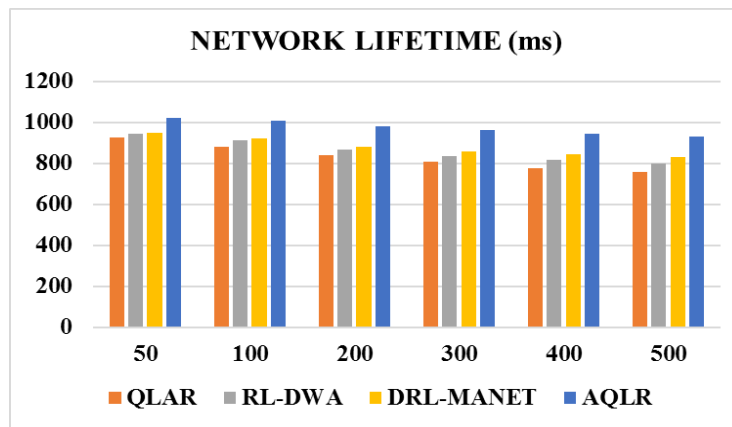


Figure 12: Network Lifetime Vs Node Density

The error bars for network lifetime illustrates in Figure 13 depicts AQLR kept a consistent performance with small deviations of around +/- 10-15 seconds, while RL-DWA and DRL-MANET showed a larger fluctuation indicating an unstable distribution of power among the nodes.

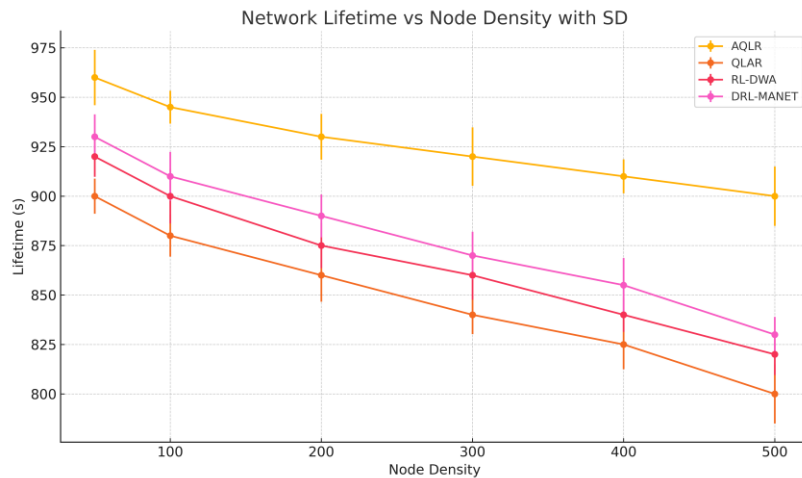


Figure 13: Network Lifetime Vs Node Density with Standard Deviation

The above metrics highlight the strength of AQLR, since it not only provides the best average but also the lowest dispersion, which is very important in MANETs when deployed in practice.

5. CONCLUSION

AQLR is a context-aware, Q-learning-based routing protocol that adaptively optimizes routing performance in MANETs. By integrating multi-metric evaluations into the Q-learning framework, the protocol dynamically adjusts to mobility, congestion, and energy constraints. Evaluations indicate substantial improvements in delivery ratio, delay, overhead, and energy efficiency over state-of-the-art ML protocols. Future enhancements of AQLR protocol can be extended by incorporating deep reinforcement learning models such as DQNs or actor-critic methods for continuous state-action optimization. Additionally, the framework can be adapted for heterogeneous and large-scale environments such as Vehicular Ad Hoc Networks (VANETs) and UAV-based networks. Incorporating federated learning and multi-agent collaboration across clusters could further reduce learning convergence time and improve decision decentralization.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Serhani, Abdellatif, Najib Naja and Abdellah Jamali. "QLAR: A QLAR based adaptive routing for MANETs." 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA) (2016): 1-7.
- [2] Suh, B., Akobir, I., Kim, J., Park, Y., & Kim, K.-I.. A resilient routing protocol to reduce update cost by unsupervised learning and deep reinforcement learning in mobile ad hoc networks. *Electronics*, 14(1), 166,2025
- [3] Kaviani, S., Ryu, B., Ahmed, E., Larson, K. A., Le, A., Yahja, A., & Kim, J. H. Robust and scalable routing with multi-agent deep reinforcement learning for MANETs. *arXiv preprint arXiv:2101.03273*, 2021
- [4] Yahja, A., Kaviani, S., Ryu, B., Kim, J. H., & Larson, K. A. (2023). DeepADMR: A deep learning-based anomaly detection for MANET routing. *arXiv preprint arXiv:2302.13877*. <https://arxiv.org/abs/2302.13877>
- [5] Kumar, A., & Mallick, P. K. (2020). A review on MANET: Characteristics, challenges, applications and routing protocols. *International Journal of Information Technology*, 12, 229–238. <https://doi.org/10.1007/s41870-019-00331-x>
- [6] Lin, X., Andrews, J. G., Ghosh, A., & Ratasuk, R. (2013). An overview of 3GPP device-to-device proximity services. *IEEE Communications Magazine*, 52(4), 40–48. <https://doi.org/10.1109/MCOM.2013.6588662>
- [7] Sharma, M., & Sahu, P. (2022). Energy-efficient and QoS-aware hybrid routing protocol using reinforcement learning for MANETs. *Computer Communications*, 188, 67–78. <https://doi.org/10.1016/j.comcom.2021.11.020>
- [8] Zeng, Z., Zhang, Y., & Yu, R. (2021). Deep reinforcement learning for wireless network routing: A review. *IEEE Access*, 9, 97666–97690.
- [9] S. Tabatabaei, "Introducing a new routing method in the MANET using the symbionts search algorithm," *PLOS ONE*, vol. 18, no. e0290091, 2023.
- [10] A. Abdullah "Enhanced-AODV Routing Protocol to Improve Route Stability of MANETs," *The International Arab Journal of Information Technology*, vol. 19, no. 5, Sept. 2022.
- [11] S. Hamad, S. Belhaj "Average Link Stability with Energy-Aware Routing Protocol for MANETs," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 1, 2018.
- [12] S. Chettibi, S. Chikhi, "Dynamic fuzzy logic and reinforcement learning for adaptive energy efficient routing in mobile ad-hoc networks," *Applied Soft Computing*, vol. 38, pp. 321–328, 2016.

- [13] K. Sumathi, "Minimizing Delay in Mobile Ad-Hoc Network Using Ingenious Grey Wolf Optimization Based Routing Protocol," *International Journal of Computer Networks and Applications (IJCNA)*, 2019.
- [14] P. T. K. Priyambodo, D. Wijayanto, M. S. Gitakarma, "Performance Optimization of MANET Networks through Routing Protocol Analysis," *Computers*, vol. 10, no. 1, p. 2, 2021.
- [15] F. Ali, K. Zaman, B. Shah, T. Hussain, H. Ullah, A. Hussain, D. Kwak, "LSTDA: Link Stability and Transmission Delay Aware Routing Mechanism for Flying Ad-Hoc Network (FANET)," 2023.
- [16] S. C. Shah, "An Energy-Efficient Resource Management System for a Mobile Ad Hoc Cloud," *IEEE Access*, vol. 6, pp. 62898–62914, 2018.
- [17] B. Muneeswari, M. S. K. Manikandan, "Energy Efficient Clustering and Secure Routing Using Reinforcement Learning for Three-Dimensional Mobile Ad Hoc Networks," *IET Communications*, 2019.
- [18] A. Abdellaoui, Y. Himeur, O. Alnaseri, S. Atalla, W. Mansoor, J. Elmhamdi, H. Al-Ahmad, "Enhancing Stability and Efficiency in Mobile Ad Hoc Networks (MANETs): A Multicriteria Algorithm for Optimal Multipoint Relay Selection," *Information*, vol. 15, no. 12, p. 753, 2024.
- [19] V. B. Rao, P. A. Vikhar, "Reinforcement Machine Learning-based Improved Protocol for Energy Efficiency on Mobile Ad-Hoc Networks," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 8s, pp. 654–670, 2023.
- [20] N. Alsalmi, K. Navaie, H. Rahmani, "Energy and throughput efficient mobile wireless sensor networks: a deep reinforcement learning approach," *IET Networks*, vol. 13, no. 5–6, pp. 413–433, 2024.
- [21] M. Maleki, V. Hakami, M. A. Dehghan, "A Model-Based Reinforcement Learning Algorithm for Routing in Energy Harvesting Mobile Ad-Hoc Networks," *Wireless Personal Communications*, vol. 95, pp. 3119–3139, 2017.
- [22] Z. Xin, Z. Xin, G. Hongyao, L. Qian, "The Inter-Cluster Routing Algorithm in Wireless Sensor Network Based on the Game Theory," in *Proceedings of the 4th International Conference on Digital Manufacturing & Automation*, Shanan, China, 2013, pp. 1477–1480. [Online]. Available: <https://doi.org/10.1109/ICDMA.2013.353>.
- [23] Z. Lin, J. Sun, "Routing Protocol Based on Link Stability in MANET," in *Proc. of the 2021 World Automation Congress (WAC)*, Taipei, Taiwan, 2021, pp. 260–264. [Online]. Available: <https://doi.org/10.23919/WAC50355.2021.9559469>.
- [24] R. P. Salim, R. Ramachandran, "Link stability-based optimal routing path for efficient data communication in MANET," *International Journal of Advances in Intelligent Informatics*, vol. 10, no. 3, pp. 471–489, Aug. 2024.
- [25] G. K. Ahirwar, R. Agarwal, A. Pandey, "Secured Energy Efficient Chaotic Gazelle Based Optimized Routing Protocol in Mobile Ad-Hoc Network," *Sustainable Computing: Informatics and Systems*, 2025.
- [26] Pandey, Singh "Modified route request broadcasting for improving multipath routing scheme performance in MANET", *International Journal of Computer Networks & Communications*, Vol.16, No.4, 2024.
- [27] S.SugunaDevi and A.Bhuvanewari, "Quantile Regressive Fish Swarm Optimized Deep Convolutional Neural Learning for Reliable Data Transmission in IoV", *Vol.13, No.2, 2021*.

AUTHORS

Dr. P. Tamilselvi is an Associate Professor in the Department of Information Technology at Cauvery College for Women (Autonomous), Tiruchirappalli. She has over 18 years of academic experience, with primary research interests in Mobile Ad Hoc Networks (MANETs). She has published eight research papers, including four in Scopus-indexed journals. Her academic contributions reflect a strong commitment to research, student mentorship, and technological advancement.



Dr. S. Suguna Devi is an Associate Professor in the Department of Information Technology at Cauvery College for Women (Autonomous), Tiruchirappalli. With 20 years of teaching experience, she has contributed extensively to research in the field of the Internet of Vehicles (IoV) and is the author of a published book. She has successfully guided numerous students through their academic and technical projects across diverse domains.



Dr. M. Thangam is an Assistant Professor in the Department of Computer Science at Mount Carmel College (Autonomous), Bangalore. She brings over 18 years of academic experience with expertise in Machine Learning and the Internet of Things (IoT). She has published five research papers in reputed journals indexed in Scopus and Web of Science, and has presented her work at international conferences. Her academic career is marked by a dedication to research excellence, student engagement, and innovation.



Dr. P. Muthulakshmi is an Associate Professor in the Department of Computer Science at Cauvery College for Women (Autonomous), Tiruchirappalli. With more than 16 years of teaching and research experience, her interests centre on Big Data Analytics. She has authored six research papers, including publications in Scopus and Web of Science indexed journals. She is known for her passion for teaching, student mentoring, and ongoing contributions to emerging technologies.

