A Reinforcement Learning Approach for Routing in Marine Communication Network of Fishing Vessels

Simi Surendran*, Alberto Montresor[1] and Maneesha Vinodini Ramesh[2]

*Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, India.
[1]Department of Information Engineering and Computer Science, University of Trento, Trento, Italy.
[2]Amrita Center for Wireless Networks & Applications (AmritaWNA), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India.

*Corresponding author(s). E-mail(s): simisurendran@am.amrita.edu;

Abstract

The lack of low-cost communication facilities to the coast remains a fundamental problem for fishermen engaged in deep-sea fishing. The Offshore Communication Network (OCN) is a wireless network of fishing vessels, whose goal is to provide internet over the ocean. However, extreme weather conditions, the challenge of deploying additional infrastructure, vessel movements caused by waves, and the increased mobility at sea hinder the performance of traditional routing protocols in OCNs. This paper proposes a Q-learning-based routing strategy for OCNs. As the quality of the learning process depends on the reward function, the latter has been designed to account for the most important features, including real-time forecasts of connectivity quality, path probability, link availability duration, and distance to the destination. The performance of the proposed routing strategy OCN-AR is evaluated through simulations under various conditions, including different mobility scenarios, transmission rates, vessel rocking factors, and node densities, and is compared against existing protocols. The results indicate that OCN-AR is well-suited for maritime networks of fishing vessels.

Keywords: Marine communication network, Fishing vessel network, Routing protocol, Reinforcement learning

1 Introduction

The absence of adequate means to communicate with the shore is a fundamental impediment encountered by fishermen in deep-sea fishing. Conventional communication technologies such as cellular networks and marine radio can only provide connectivity up to a distance of about 20 km from the coast. These systems cannot be exploited since fishing vessels travel more than 100 km into the deep sea. On the other hand, satellite phone services are too expensive for fishermen. The Offshore Communication Network (OCN) proposed by Rao et al. aims at resolving this issue by building a wireless network of fishing vessels that provide internet over the ocean [1]. The goal is to enable fishermen to connect to the internet using affordable, handheld devices such as mobile phones, allowing them to communicate with the shore and other vessels.

Although marine networks share some of the features of terrestrial mobile ad-hoc and vehicular networks, they present unique characteristics and research challenges [2]. Unlike terrestrial vehicular networks, which are limited by road infrastructure, OCN benefits from greater freedom of movement in the open ocean. However, OCNs face significant communication challenges due to the inability to deploy additional infrastructure in the marine environment, the impact of extreme weather conditions on wireless signals, and the misalignment of directional links. The most critical challenge arises from the rough sea conditions, which can severely affect link quality. The topology of the network can change rapidly due to antenna orientation, the rocking movement of vessels, and propagation effects that weaken signal strength. For all the above reasons, providing uninterrupted internet connectivity is a difficult problem in OCNs.

Given these challenges, an adaptive routing mechanism capable of responding to the dynamic and unpredictable nature of OCNs is essential for maintaining operational efficiency. Reactive routing protocols face difficulties in establishing reliable end-to-end paths due to issues in route discovery, while proactive protocols are resource-intensive, as they precompute all source-destination routes, including those that are not needed [3, 4]. Location-based routing protocols, which use predetermined parameters to select the next hop, fail to adapt quickly to the dynamic environment of OCNs. Integrating machine learning algorithms into the routing process continuously monitors the network conditions and adjusts the routing paths in real time. This adaptability is important in a marine environment where the links can change within short periods, making precomputed routes ineffective. Reinforcement learning (RL) techniques, in particular, have demonstrated potential, as agents interact with the wireless environment and make decisions based on feedback signals.

Although numerous routing protocols based on RL strategies have been proposed in the literature for terrestrial networks, they must be adapted to meet the unique requirements and challenges of OCNs [5, 6]. A well-designed reward function is essential for learning the optimal next-hop to deliver packets effectively under dynamic conditions. The contribution of this paper goes exactly in this direction. An OCN-specific reward function is proposed, taking into account the real-time connectivity quality estimate, the link availability duration, the path probability, and the geographical distance to the destination of neighbor nodes. Based on a Q-learning strategy, the proposed protocol discovers adaptive routes in a completely distributed manner.

We evaluated the proposed protocol under various network conditions and compared its performance with other Q-learning-based routing schemes. In the simulation environment, the routing strategy achieved significantly higher packet delivery ratio.

The rest of the paper is organized as follows: Section 2 reviews previous works related to routing protocols based on reinforcement learning. In Section 3, we present the architecture of OCN. Section 4 describes the reinforcement learning model and the formulation of the reward function. Section 5 presents simulation results, followed by the concluding remarks of Section 6.

2 Related Works

The routing problem in wireless networks has been extensively studied in the literature. In conventional *proactive* table-driven routing, algorithms predetermine all source-destination paths. However, in OCNs, the control traffic required to update neighborhood information is excessive, and nodes cannot predict in advance whether these paths will be needed for routing. In *reactive* routing, traditional protocols aim to discover a complete path between a source and target node, but gathering neighbor information in OCNs is challenging, often causing reactive protocols to fail in establishing a proper end-to-end path. Similarly, in *location-based* routing, the selection of predefined parameters for the next hop cannot adapt quickly enough to the highly dynamic nature of the environment.

RL approaches have been successfully utilized to introduce intelligence in routing solutions across various wireless applications [6–9]. A comprehensive survey of RL-based routing protocols has been presented by Mammeri [10]. Boyan et al. introduced the first Q-learning-based routing algorithm in a telephone network [11]. While Q-routing outperforms non-adaptive algorithms, it may struggle to find the optimal policy under low or fluctuating network load conditions. To address this issue and improve convergence, several variants of Q-routing have been proposed [12]. Further, RL routing schemes have been applied to different wireless networks such as mobile ad-hoc network (MANET) [13, 14], vehicular adhoc networks (VANET) [15], wireless sensor networks (WSN) [16–18], wireless mesh networks (WMN) [19] and delay tolerant networks (DTN) [20] for performance optimization. Some other contexts like underwater sensor networks [21, 22], software-defined networks [23, 24], and information-centric networks [25] has also been employed RL-based routing.

Several model-free Q-learning based routing approaches have been proposed such as QLAODV [31], MQ routing [27], FROMS [16], SMART [32], QGrid [15], and Q-Geo [30]. Q-learning based routing for flying ad-hoc networks was discussed in [33] and [34]. Lahsen-Cherif et al. proposed a Q-routing approach designed for wireless mesh networks that utilize directional antennas [35]. Meanwhile, Lee et al. introduced a multi-agent Q-learning framework specifically for UAV networks [36]. Wu et al. took a different approach by modifying the AODV protocol to develop QLAODV, a routing scheme for vehicular networks that considers vehicle movements and available channel bandwidth [31]. Nonetheless, QLAODV can encounter delays in collecting link information on multi-hop routes, which extends the time required for route adaptation.

Sl No	Routing Protocol	Application	Performance Parameters						
			p_1	p_2	p_3	p_4	p_5	p_6	p_7
1	Q-routing[11]	Static Network	x						
2	PQ-routing [12]	Static Network	x					x	
3	DRQ [26]	Static Network	x					x	
4	CRL [13]	MANET		x	x				
5	MQ routing [27]	MANET		x		x			
6	ARBR [20]	DTN	x	x					
7	FROMs [16]	WSN		x		x	x		
8	QELAR [28]	Underwater WSN				x			
9	QoE routing [29]	Multimedia Network	x		x				
10	DMARL [21]	Underwater WSN		x			x		
11	QGrid [15]	VANET		x					
12	Q-Geo [30]	Robotic Networks		x			x		
13	DCR [19]	WMN		x					
14	QL-AODV [31]	MANET		x		x			
15	SMART [32]	Cognitive Radio Networks		x			x		

Table 1: Summary of RL-based routing protocols and performance metrics: p_1:delay, p_2 : delivery ratio, p_3: packet error rate, p_4: Lifetime, p_5: overhead reduction, p_6: learning speed improvement, p_7: path length reduction

In another study, Macone et al. discussed a proactive routing technique called MQ-routing, aimed at improving node lifetime in mobile ad-hoc networks used in disaster relief scenarios [27]. This technique combines path availability and energy parameters in its optimization process. However, the proactive nature of MQ-routing makes it difficult to update the neighborhood table in rapidly changing environments. Although MQ-routing accounts for mobility and link availability in updating Q-values, it does not adequately consider link connectivity quality, which may result in high Q-values for nodes with low mobility despite poor connectivity. Forster et al. developed the FROMs protocol, a multicast routing solution for wireless sensor networks [16]. The protocol SMART, proposed by Saleem et al., applies Q-learning in a cluster-based framework to enhance stable route selection and optimize secondary user performance in cognitive radio networks [32].

The Q-Grid protocol [15], a geographic routing scheme for vehicular networks, uses Q-learning techniques. However, it does not guarantee effective routing performance in networks with intermittent links, as it fails to utilize information on link status or availability duration. Jung et al. proposed another geographic routing protocol called Q-Geo for robotic networks based on Q-learning [30]. To ensure reliable data transfer, Q-Geo's reward function takes into account packet travel speeds, distance, link status, and location status. However, the protocol's fixed learning rate results in a uniform Q-value update rate across all network conditions. This static learning rate may not

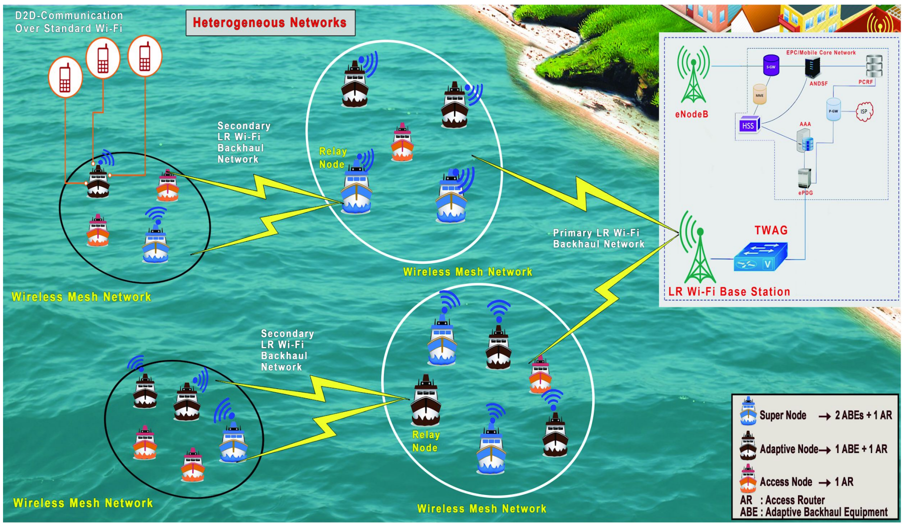


Fig. 1: Architecture of Offshore Communication Network[1]

be effective in networks with time-varying and context-dependent topology changes. Table 1 summarizes the RL routing schemes applied in different types of networks along with the performance metrics.

Conventional routing approaches such as static, proactive, reactive, and geographic routing protocols can not adapt to rapid network changes and discover reliable paths. Although many adaptive RL routing algorithms have been suggested, all protocols mentioned above have been developed based on the distinctive features of each network. The learning parameters and reward function must be customized to meet the specific requirements of each network. Since OCNs operate in harsh environments and communication to the shore remains an essential factor, an adaptive model is necessary to enhance network connectivity. In previous work, we focused on packet status and signal strength when designing the reward function [37]. This paper revises the reward function by incorporating node-level connectivity quality, link availability, and distance to the destination, aiming to enhance the packet delivery ratio.

3 Network Architecture

The OCN is a fishing vessel network that provides Internet access well beyond 100 km from the shore. Its goal is to let fishermen access the internet through their mobile phones, using off-the-shelf apps such as WhatsApp for calls and messages. The fishing vessels forming the OCN act as edge nodes that locally analyze routing data and perform multiple network functions.

Based on the resources present in the fishing vessels, OCN nodes are classified into three types: access nodes, adaptive nodes, and supernodes.

- An *access node* is a fishing vessel that only provides a wireless access router (AR) and communicates using Wi-Fi omnidirectional antennas.
- An *adaptive node* is a vessel equipped with adaptive backhaul equipment (ABE) and one AR.
- A node equipped with two ABEs and one AR is categorized as a *supernode*.

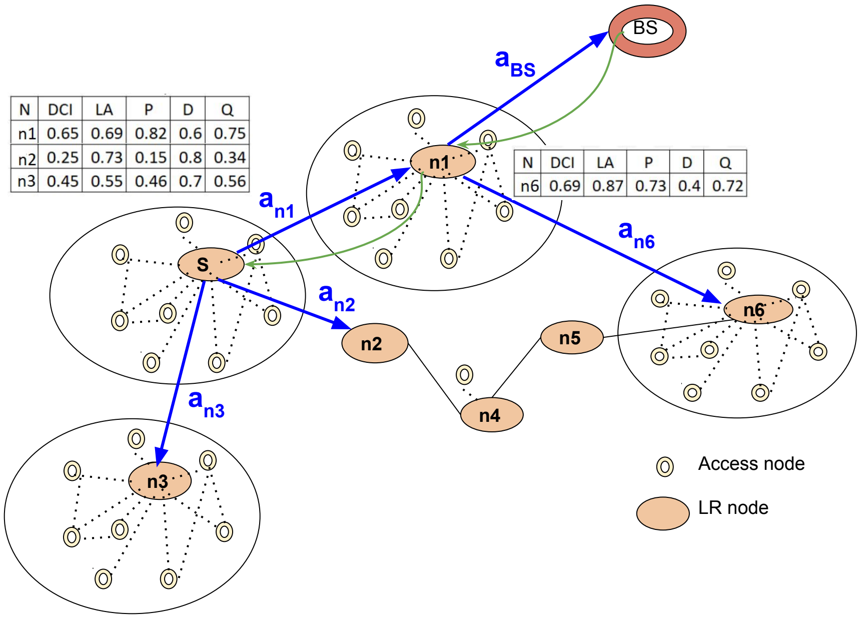
Adaptive nodes and supernodes use 120^{o} sectored long-range Wi-Fi links. These nodes are also termed *long range* (LR) nodes and are the ones forming the backbone network. LR Wi-Fi link's range is between 15 and 20 km. AR nodes with Wi-Fi links form a wireless mesh network.

The network architecture is shown in Figure 1. A detailed description of the architecture and packet forwarding strategies can be found in previous papers [1, 38, 39], [37, 40, 41]. The validation of the OCN architecture has been carried out over the Arabian Sea from a coastal village in Kerala, India. LR Wi-Fi equipment from Ubiquiti Networks and Cisco Linksys access routers were deployed for sea trials. A single onshore base station is located 56 m above sea level, and the vessel's ABE is placed at 9 m above sea level. In these sea trial experiments, the network provided a 40+ km range in the first hop and 20 km in every succeeding hop.

4 RL Based Routing Model for OCN

We consider an OCN routing scenario consisting of LR nodes, access nodes, and onshore base stations. All nodes, except the onshore base stations, are mobile. The transmission range of LR nodes spans 15-20 km, whereas access nodes offer connectivity within a radius of 150-250 meters. To reach the final destination, multi-hop communication is necessary. We assume that all nodes are equipped with GPS devices that provide their current positions. Each node generates messages with varying priorities, such as emergency, audio, and video messages. Additionally, all nodes periodically transmit beacons to their neighboring nodes, containing location data and other routing-related information.

Routing in OCN is formulated as an RL problem. In this scenario, the entire network, including access nodes, LR nodes, and base stations—constitutes the environment for the agent. Each packet within the network acts as an individual agent. The agent receives rewards or penalties based on the success or failure of each transmission, and it uses these feedback signals to learn the optimal policy for selecting the best next hop. Each node in the network maintains information about its neighbors' connectivity features, such as connectivity quality, link availability duration, path probability, distance to the destination, and Q-value. These features are periodically updated through beacon messages exchanged between nodes. To determine the next hop for packet forwarding, we employ a temporal difference off-policy Q-learning algorithm [42]. This algorithm uses an ϵ-greedy policy, where ϵ represents a



N	DCI	LA	P	D	Q
n1	0.65	0.69	0.82	0.6	0.75
n2	0.25	0.73	0.15	0.8	0.34
n3	0.45	0.55	0.46	0.7	0.56

N	DCI	LA	P	D	Q
n6	0.69	0.87	0.73	0.4	0.72

Fig. 2: State-Action model in OCN routing: Node S have three possible actions: a_{n_1}, a_{n_2}, a_{n_3}. It selects the best action a_{n_1} and forwards the packet to node n_1. Node S will receive a reward for action a_{n_1} and update its Q-value accordingly. Similarly, the best action will be selected from node n_1 and this process will continue until the packet reaches its destination.

small probability of introducing randomness into action selection, allowing the agent to explore new potential routing solutions. As wireless links fluctuate over time, this approach ensures that new links can be incorporated into the routing path, making it adaptable to the dynamically changing environment of OCNs.

Figure 2 illustrates the routing scenario in OCN-AR. Consider a situation where node S needs to communicate with the base station and must select a next-hop neighbor. Node S has three possible options for backbone connections: n_1, n_2, and n_3. The actions of selecting these neighbors are denoted by a_{n_1}, a_{n_2}, and a_{n_3}, respectively. Node S will choose the action with the highest Q-value with a probability of $1-\epsilon$, for instance, a_{n_1}. Upon executing this action, node S receives a reward that is calculated using both local and remote information. At node n_1, the process repeats, with the next-hop being selected based on the highest Q-value. In this scenario, the base station has the highest Q-value, leading node n_1 to choose that path. Each node maintains a Q-table that records the Q-values of its neighbors, which are updated after feedback is received. These updated Q-values are used in future packet forwarding decisions, enabling the selection of the most optimal route over time.

A Markov Decision Process with the following state, action, and reward function is used to model the routing process.

- *States*: Each packet in the network corresponds to an agent, and the agent's environment is defined as the collection of all nodes in the network. The state of an agent is represented by the node where the packet is currently located. In other words, if a node i receives a packet for forwarding or generates a new packet, the agent's state is identified as i. Consequently, the state space is defined as the set of all nodes in the network, along with their associated features.
- *Actions*: Let $\mathcal{N}_i = \{n_1, n_2, ..., n_k\}$ be the neighbors of node i. An action in node i at time t is the selection of one of the neighbors from $\mathcal{N}_i$ for forwarding packet. The action space comprises the set of possible actions that all nodes can take in an OCN.
- *Reward function*: The reward function is the most important factor in determining the effectiveness of Q-routing. The environment returns a reward to the agent for indicating the impact of the neighborhood selection. This reward comprises a function of local information computed in the node and a component of remote information received as feedback from other nodes, obtained through the acknowledgment scheme. The parameters to be considered at the local (node) level include the connectivity quality of neighboring nodes, the link availability duration, the probability of the path to the destination, and the distance to the destination.

The following subsections detail the first three components of the proposed reward function, followed by the reward function itself and a description of the Q-Learning algorithm.

4.1 Node Connectivity Quality

To study the characteristics of marine wireless links, we collected data on how signal strength varies with distance in different sea trials[41]. This data was used to examine the factors affecting connectivity. Sea wave-induced movement and propagation effects directly impact the quality of signals in OCN. These parameters change the topology from time to time and create a severe challenge in routing packets. Hence, we developed a machine learning framework using historical and online data to predict the link connectivity probability.

A metric called *dynamic connectivity index (DCI)* has been defined to compute the level of node connectivity by employing this link prediction model [43]. Assume that connectivity is determined solely based on the local link quality. In this case, there is a possibility of selecting a neighbor with good link quality. However, if the chosen neighbor has poor overall connectivity, it is less likely to successfully deliver the message. DCI is a node-level measure of a node's connectivity to clusters and base stations. DCI helps to decide the most suitable next-hop and minimizes the possibility of node isolation. To define the DCI of a node, we compute the DCI of its next-hop neighbors and link probability to those neighbors as shown in equation 1.

$$DCI(x) = \sum_{i \in AN(x)} w_i \cdot p_i \cdot DCI(i) \tag{1}$$

where w_i is the weight of next-hop i, p_i is the link probability from node x to neighbor i, $AN(x)$ is the list of neighbor nodes whose link probability is greater than a threshold

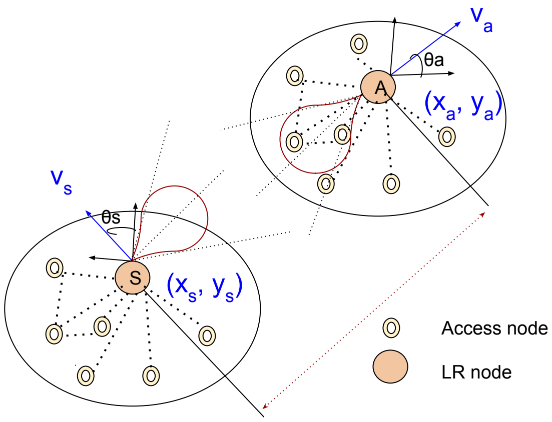


Fig. 3: Communication scenario between two mobile LR nodes with movement vectors v_s and v_a.

value 0.25 and $DCI(i)$ is the dynamic connectivity index of neighbor i. A dynamic weighting scheme is used to prioritize the neighbors of a node [43].

4.2 Link Availability Duration

Consider a communication scenario between two mobile LR nodes as shown in Figure 3. Let nodes S and A be separated by a distance d^{τ} after time τ. Let v_s and v_a be the velocity vectors of nodes S and A. The alignment of directional antennas between the LR nodes affects the communication radius. Let ϕ represent the angle of misalignment among the transmitter and receiver antennas. An increase in ϕ decreases the communication distance between nodes and varies with the environments' dynamics. z_{eff}^{τ} represents the effective communication distance of S and A after time τ as a function of ϕ. The displacement over the x, y directions and the distance between nodes due to the mobility can be computed as:

- Difference in x direction δ_x^{τ}:

$$\delta_x^{\tau} = (x_s - x_a) + (v_a \cos\theta_a - v_s \cos\theta_s)\tau \tag{2}$$

- Difference in y direction δ_y^{τ}:

$$\delta_y^{\tau} = (y_s - y_a) + (v_a \sin\theta_s - v_s \sin\theta_s)\tau \tag{3}$$

- Distance between nodes after time τ:

$$\begin{aligned} d^{\tau} &= \sqrt{(\delta_x^{\tau})^2 + (\delta_y^{\tau})^2} \\ &= \Big[((x_s - x_a) + (v_a \cos\theta_a - v_s \cos\theta_s)\tau)^2 \\ &\quad + ((y_s - y_a) + (v_a \sin\theta_a - v_s \sin\theta_s)\tau)^2\Big]^{1/2} \end{aligned}$$

For effective communication, $d^{\tau} < z_{\mathit{eff}}^{\tau}$. The maximum duration of link availability τ can be obtained by rewriting this equation as:

$$\tau = \frac{-(\Delta_x V_{x\delta} + \Delta_y V_{y\delta}) \pm \sqrt{(V_{x\delta}^2 + V_{y\delta}^2)(z_{\mathit{eff}}^{\tau})^2 - (\Delta_y V_{x\delta} - \Delta_x V_{y\delta})^2}}{V_{x\delta}^2 + V_{y\delta}^2} \tag{4}$$

where

$$\begin{aligned} \Delta_x &= (x_a - x_s), \\ \Delta_y &= (y_a - y_s), \\ V_{x\delta} &= v_a \cos(\theta_a + \phi_a) - v_s \cos(\theta_s + \phi_s), \\ V_{y\delta} &= v_a \sin(\theta_a + \phi_a) - v_s \sin(\theta_s + \phi_s) \end{aligned}$$

4.3 Path Probability

Information regarding the possibility of an end-to-end path between the LR nodes or the base station helps to select next-hop neighbors. There may be multiple paths with different hops and various link properties between two nodes. The connectivity probabilities of these links can be predicted using the machine learning framework in OCN. Additionally, the details of existing paths are available through the feedback mechanism of the routing algorithm. We consider all such paths and link probabilities to estimate the connectivity probability between nodes.

Let $r_1, r_2, ..., r_q$ be the routing paths exist between two nodes a and b. Consider a path r_i that consists of m links, whose availability is expressed by probabilities $l_1, l_2, ..., l_m$. Assume that such probabilities are independent of each other. Then, the probability of the existence of the path r_i can be calculated by the product of its individual link probabilities as in Equation 5.

$$Prb(r_i) = l_1 \cdot l_2 \cdot \ldots \cdot l_m \tag{5}$$

If any one of the links in this path breaks, then $Prb(r_i)$ becomes zero. Then the probability for path w_i to not exist is equal to $1 - Prb(r_i)$.

Assume that the probabilities of multiple paths existing between two nodes are independent. Considering q possible paths between nodes a and b, the probability for the existence of at least one path is equal to

$$\mathcal{P}(a, b) = 1 - \prod_{i=1}^{q}(1 - Prb(r_i)) \tag{6}$$

When the path probability of a neighbor node to the destination is high, the node will earn a larger Q-value. Under certain network conditions, such as when a significant number of links fail, computing path probability can be challenging. In these situations, the reward function will not take this factor into account.

4.4 Reward Function

The reward for forwarding a packet from node i to node j with action a_j at time t is defined as:

$$\mathcal{R}_t = \mathcal{R}_f + \beta_1 \cdot DCI(j) + \beta_2 \cdot \tau + \beta_3 \cdot \mathcal{P}(j, dest) + \beta_4 \cdot \frac{\Delta d(i,j)}{d(i,j)} \tag{7}$$

where $DCI(j)$ is the connectivity index of node j, τ is the link availability time, $\mathcal{P}(j, dest)$ is the probability that a path exists from node j to the destination and $\Delta d(i,j)$ is the difference in distance between i and j to the destination. β_is are the weights given to each of the node level features. Since $DCI(j)$ and τ are the most important features, we used a normalized weight vector $\beta = \langle 0.5, 0.25, 0.1, 0.15 \rangle$ in simulations. $\mathcal{R}_f$ is defined as

$$\mathcal{R}_f = \begin{cases} +5, & \text{ACK received} \\ -5, & \text{ACK not received} \\ -10, & \text{next-hop is local maximum} \\ R_{max}, & \text{next-hop is destination} \end{cases} \tag{8}$$

The node will receive a positive or negative reward, depending whether the packet has been received or not. In addition, a higher negative value will be assigned when the selected neighbor does not have a potential connection for forwarding; the maximum reward will be provided if the subsequent hop from the selected neighbor is the destination. In the simulation experiments we used R_{max} as 100.

4.5 OCN Adaptive Routing Algorithm

The OCN Adaptive Routing (OCN-AR) algorithm is shown in Algorithm 1. Since the state of the agent includes connectivity features of its neighbors j, it can be represented as $\langle DCI(j), \tau, \mathcal{P}(j, dest), \Delta D \rangle$ where ΔD is the difference in distance of current node and neighbor j to the destination. In Figure 2, the state of the agent for neighbor n_1 at node S is $\langle 0.65, 0.69, 0.82, 0.6 \rangle$. Each node develops and maintains a Q-value list of its neighbor nodes. If an entry is not present in the Q-table, the source node creates a new record using information from the target and the neighbor node. We initialize the Q-value of all state-action pairs $\mathcal{Q}(s_t, a_t)$ using a function based on DCI and link probability $\mathcal{LP}$. Each node i periodically sends beacon messages to update the Q-table. This includes Q-values, the current $DCI(i)$, and location information loc. When a node receives a beacon or acknowledgment message, it updates the reward $\mathcal{R}_t$ and the Q-value using the Equation 9.

$$\mathcal{Q}^{new}(s_t, a_t) = (1-\alpha)\mathcal{Q}^{old}(s_t, a_t) + \alpha.[\mathcal{R}_t + \gamma. \max_a \mathcal{Q}((s_{t+1}, a)] \tag{9}$$

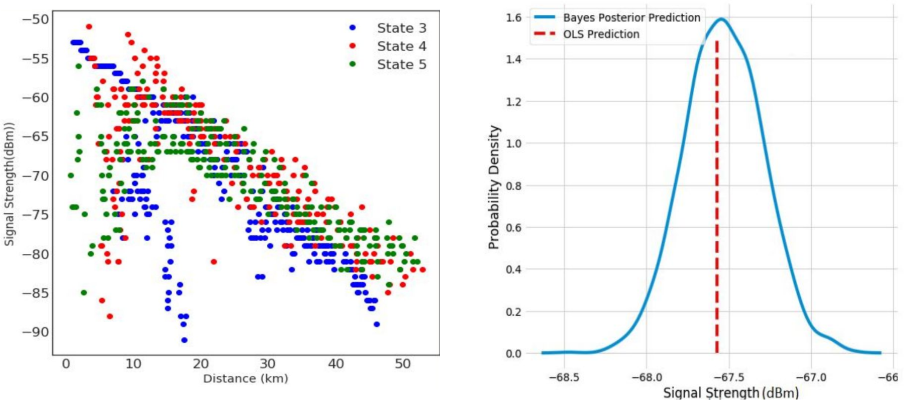
where $\mathcal{Q}^{old}(s_t, a_t)$ is the old Q-value for the action a_t in state s_t. $\mathcal{R}_t$ denotes the reward for the action a_t in state s_t. $\max_a \mathcal{Q}((s_{t+1}, a)$ is the maximum Q-value in the next state s_{t+1} with best action a. The learning rate and the discount factor are represented as α and γ, respectively, with values in the range $[0, 1]$.

The extent of topology changes varies in different sea states and fishing stages. In the roughest sea states, rocking movement will be extremely large and cause frequent topology changes. In such conditions, a higher learning rate is required to prioritize the current data. Moreover, we keep a small discount factor as the future expectations are not accurate. In more calm sea states and during resting or fishing phases, topology changes are smaller compared to other scenarios. In this case, a low learning rate is favored. We apply a context-dependent parameter selection instead of using a constant learning rate and discount factor.

After creating the Q-table, the nodes select a neighbor with the highest $\mathcal{Q}$ as next-hop with probability $1 - \epsilon$. This phase is the exploitation stage of Q-learning. For the exploration of the state space, the source node arbitrarily selects any of the neighbors as next-hop with probability ϵ. LR nodes select the best neighbor only from the list of LR nodes as the transmission radius of these nodes is very large compared to access nodes. Access nodes can choose multiple hops to reach an LR node.

Algorithm 1: OCN Adaptive Routing

```
State $S$ = Set of all nodes in the network;
Action $A$ = Neighbor nodes ;
Exploration coefficient $\epsilon = 0.01$;
for every $s \in S$ do
    for every $a \in A$ do
        $\mathcal{N}$ = neighbors of node $s$ ;
        $\mathcal{Q}(s_t, a_t) = DCI(\mathcal{N}_{a_t}) * \mathcal{P}$;
while true do
    if beacon timer expires in each node $x$ then
        Send beacon $< Q - value,\ DCI(x),\ loc >$
    if receive a beacon/ack message then
        Update reward function $\mathcal{R}_t$ and Q-value $\mathcal{Q}(s_t, a_t)$ as in Eq. 7, 8 and 9
    if Node $x$ has a packet to send then
        Generate a random number $p \in [0, 1]$;
        if $\epsilon \leq p$ then
            next-hop = $random(\mathcal{N})$
        else
            if node type(x) = access then
                next-hop = $\text{argmax}_{a \in \mathcal{N}} \{ a : \mathcal{Q}(s, a))\}$
            if node type(x) = LR then
                next-hop = $\text{argmax}_{a \in \mathcal{N}(\mathcal{LR})} \{ a : \mathcal{Q}(s, a))\}$
```



(a)

(b)

Fig. 4: (a) Data collected from sea-trial experiments on signal strength variation with distance in different vessel rocking stages (b) Prediction of signal strength at distance 25 km and vessel rocking stage = 3.

5 Results

The experimental setup for evaluating the OCN-AR protocol involved both simulations and real-world experiments to collect signal strength data under different vessel rocking conditions. A machine learning framework was then developed to predict signal strength under varying sea conditions. This prediction model is used in the computation of link probability and DCI, which were the important components of the reward function in the OCN-AR protocol. In the simulation phase, NS-2 was used to model the network, which included 50 mobile nodes—comprising 12 Long-Range nodes and 38 access nodes—along with one onshore base station. Packet generation rates varied from 2 to 64 packets per second, with a default packet size of 512 bytes, and traffic was generated using a fixed-rate UDP source. The protocol's performance was assessed based on the packet delivery ratio and compared with three state-of-the-art routing protocols: AODV, GPSR, and Q-Geo.

The signal strength data collected during marine experiments from various vessel rocking stages are shown in Figure 4a. Here, each state represents a different vessel rocking stage. The sea condition is rough in higher-numbered states, and we can observe more signal deviation due to the increased impact of waves on node mobility. We developed a machine learning framework to predict the signal strength under different sea conditions using this data. Figure 4b shows a sample prediction of signal strength over a distance of 25km. Employing this prediction model, link probability and DCI, which are components of the reward function, are computed.

Although the average speed of fishing vessels is 3-5 m/s, the rocking motion leads to more packet drops. This moving effect of the vessels is simulated by setting the node pause time. The shorter the pause time, the higher the degree of mobility. Figure 5 shows the packet delivery ratio for various pause times. Link failures often occur when

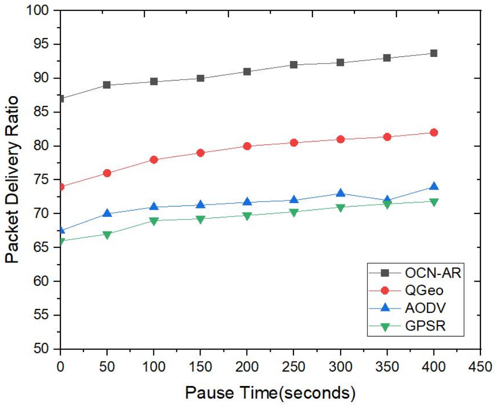
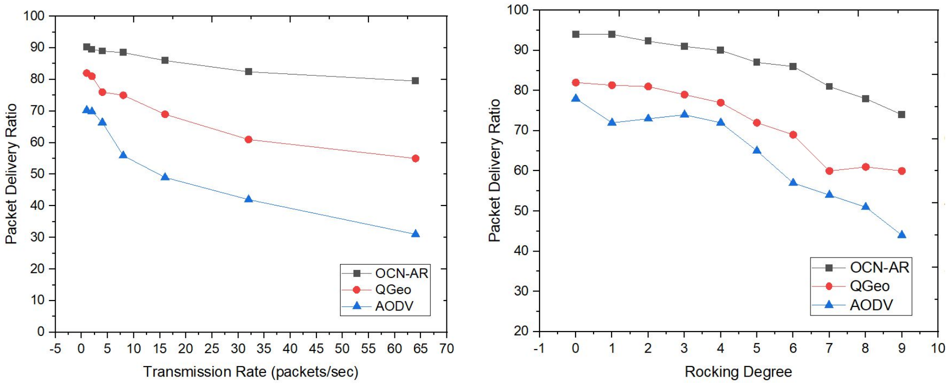


Fig. 5: Comparison of packet delivery ratio for varying mobility degree.



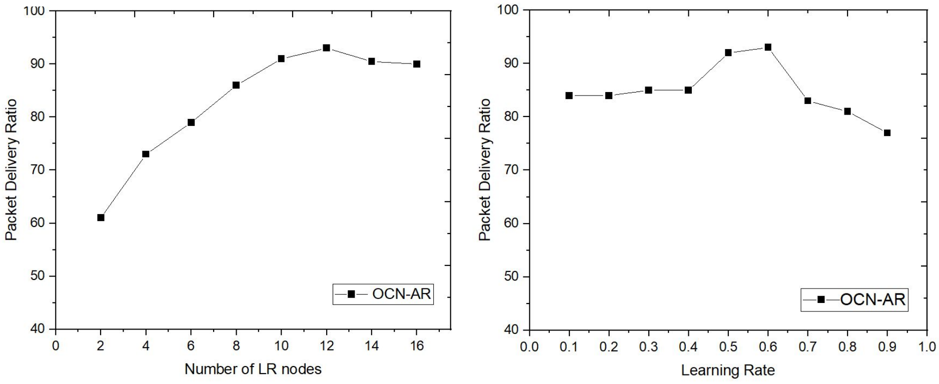
(a)

(b)

Fig. 6: (a) Variation of packet delivery ratio with data transmission rate. (b) Variation of packet delivery ratio with rocking movement vessels due to sea waves. The rocking degree indicates the wave state of the sea.

mobility levels increase, causing AODV and GPSR performance to decline. OCN-AR shows significant improvement over Q-Geo and other protocols, taking into account link availability length, direction of movement from the destination, and neighborhood connectivity level.

We varied the transmission rate at the source nodes from 2 to 64 packets per second. The results, shown in Figure 6a, reveal a significant drop in the packet delivery ratio when the transmission rate exceeds 16 packets per second across all protocols. Among the protocols, AODV experiences a faster degradation in performance compared to the other two adaptive protocols, primarily due to higher channel occupancy and increased packet loss from link failures. In contrast, OCN-AR performs better than Q-Geo as it considers the connectivity quality of the nodes.



(a)

(b)

Fig. 7: (a) Variation of packet delivery ratio with the density of LR nodes (b) Variation of packet delivery ratio with increase in learning rate.

One of the distinctive features of OCN is the rocking movement of the nodes, which often causes link breakages. We simulated this scenario using a random waypoint mobility model with varying velocities. We dropped a fixed percentage of packets from the source in each rocking stage to simulate the vessel movement factor. In this case, the packet delivery ratio of OCN-AR was compared to AODV and Q-Geo as shown in Figure 6b. The performance of AODV decreases considerably with increasing rocking degrees because of the large number of link failures. OCN-AR performs effectively in high rocking conditions compared to other protocols. This improvement is because it utilizes signal strength data and estimates of link availability. The simulation scenario includes both access and LR nodes.

We tested the impact of the density of LR nodes on the packet delivery ratio and the results demonstrated in Figure 7a. The number of LR nodes varied from 2 to 16 and we noticed an improvement in packet delivery with increased density. The learning rate is one of the critical hyper-parameters in OCN-AR that controls the rate of adaptation to the dynamic topology. Different fishing stages experience varying mobility degrees, and hence a context-dependent learning rate is used. To tune the parameter, we varied it from 0.1 to 0.9 and observed a better performance at value 0.7 in the fish searching context, as shown in Figure 7b. The optimal learning rate computed in other fishing stages is slightly different from this context.

6 Conclusion and Future Work

In this paper, we addressed the routing challenges in an offshore communication network of fishing vessels and introduced OCN-AR, a Q-learning-based routing protocol designed for effective message dissemination. OCN-AR utilized a reward function that incorporates important features- real-time connectivity quality forecasts, path probability, link availability duration, and distance to the destination of OCN. The effectiveness of this routing strategy was validated through simulations using data from sea trials, which demonstrated that the reinforcement learning strategy enhances routing performance under the challenging conditions of the ocean. The packet delivery ratio of OCN-AR was evaluated across various mobility scenarios, transmission rates, vessel rocking factors, and node densities, and compared with existing protocols. The results indicate that OCN-AR is well-suited for maritime networks of fishing vessels. However, the protocol faces challenges, particularly with the high computational overhead required for updating Q-values. In future, we plan to focus on exploring value function approximation techniques to address the computational efficiency.

References

[1] Rao, S.N., Ramesh, M.V., Rangan, V.: Mobile infrastructure for coastal region offshore communications and networks. In: Proc. of the IEEE Global Humanitarian Technology Conf. (GHTC), pp. 99–104 (2016). IEEE

[2] Yau, K.-L.A., Syed, A.R., Hashim, W., Qadir, J., Wu, C., Hassan, N.: Maritime networking: Bringing internet to the sea. IEEE Access **7**, 48236–48255 (2019)

[3] Shrivastava, P.K., Vishwamitra, L.: Comparative analysis of proactive and reactive routing protocols in vanet environment. Measurement: Sensors **16**, 100051 (2021)

[4] Kim, B.-S., Ullah, S., Kim, K.H., Roh, B., Ham, J.-H., Kim, K.-I.: An enhanced geographical routing protocol based on multi-criteria decision making method in mobile ad-hoc networks. Ad Hoc Networks **103**, 102157 (2020)

[5] Nazib, R.A., Moh, S.: Reinforcement learning-based routing protocols for vehicular ad hoc networks: A comparative survey. IEEE Access **9**, 27552–27587 (2021)

[6] Okine, A.A., Adam, N., Naeem, F., Kaddoum, G.: Multi-agent deep reinforcement learning for packet routing in tactical mobile sensor networks. IEEE Transactions on Network and Service Management (2024)

[7] Prabhu, D., Alageswaran, R., Miruna Joe Amali, S.: Multiple agent based reinforcement learning for energy efficient routing in wsn. Wireless Networks **29**(4), 1787–1797 (2023)

[8] Nandyala, C.S., Kim, H.-W., Cho, H.-S.: Qtar: A q-learning-based topology-aware routing protocol for underwater wireless sensor networks. Computer Networks **222**, 109562 (2023)

[9] Yang, X., Yan, J., Wang, D., Xu, Y., Hua, G.: Woad3qn-rp: An intelligent routing protocol in wireless sensor networks—a swarm intelligence and deep reinforcement learning based approach. Expert Systems with Applications **246**, 123089 (2024)

[10] Mammeri, Z.: Reinforcement learning based routing in networks: Review and

classification of approaches. IEEE Access **7**, 55916–55950 (2019)

[11] Boyan, J.A., Littman, M.L.: Packet routing in dynamically changing networks: A reinforcement learning approach. In: Advances in Neural Information Processing Systems, pp. 671–678 (1994). Citeseer

[12] Choi, S.P., Yeung, D.-Y.: Predictive q-routing: A memory-based reinforcement learning approach to adaptive traffic control. In: Advances in Neural Information Processing Systems, pp. 945–951 (1996)

[13] Dowling, J., Curran, E., Cunningham, R., Cahill, V.: Using feedback in collaborative reinforcement learning to adaptively optimize manet routing. IEEE Tran. Systems, Man, and Cybernetics-Part A: Systems and Humans **35**(3), 360–372 (2005)

[14] Liu, J., Wang, Q., He, C., Jaffrès-Runser, K., Xu, Y., Li, Z., Xu, Y.: Qmr: Q-learning based multi-objective optimization routing protocol for flying ad hoc networks. Computer Communications **150**, 304–316 (2020)

[15] Li, R., Li, F., Li, X., Wang, Y.: Qgrid: Q-learning based routing protocol for vehicular ad hoc networks. In: 2014 IEEE 33rd Int. Performance Computing and Communications Conf. (IPCCC), pp. 1–8 (2014). IEEE

[16] Förster, A., Murphy, A.L.: Froms: A failure tolerant and mobility enabled multicast routing paradigm with reinforcement learning for wsns. Ad Hoc Networks **9**(5), 940–965 (2011)

[17] Lu, Y., He, R., Chen, X., Lin, B., Yu, C.: Energy-efficient depth-based opportunistic routing with q-learning for underwater wireless sensor networks. Sensors **20**(4), 1025 (2020)

[18] Zhu, R., Jiang, Q., Huang, X., Li, D., Yang, Q.: A reinforcement-learning-based opportunistic routing protocol for energy-efficient and void-avoided uasns. IEEE Sensors Journal **22**(13), 13589–13601 (2022)

[19] Chai, Y., Zeng, X.-J.: A multi-objective dyna-q based routing in wireless mesh network. Applied Soft Computing **108**, 107486 (2021)

[20] Elwhishi, A., Ho, P.-H., Naik, K., Shihada, B.: Arbr: Adaptive reinforcement-based routing for dtn. In: 2010 IEEE 6th Int. Conf. on Wireless and Mobile Computing, Networking and Communications, pp. 376–385 (2010). IEEE

[21] Li, X., Hu, X., Zhang, R., Yang, L.: Routing protocol design for underwater optical wireless sensor networks: A multiagent reinforcement learning approach. IEEE Internet of Things Journal **7**(10), 9805–9818 (2020)

[22] Zhang, Y., Zhang, Z., Chen, L., Wang, X.: Reinforcement learning-based

opportunistic routing protocol for underwater acoustic sensor networks. IEEE Transactions on Vehicular Technology **70**(3), 2756–2770 (2021)

[23] Lin, S.-C., Akyildiz, I.F., Wang, P., Luo, M.: Qos-aware adaptive routing in multi-layer hierarchical software defined networks: A reinforcement learning approach. In: 2016 IEEE Int. Conf. on Services Computing (SCC), pp. 25–33 (2016). IEEE

[24] Chen, Y.-R., Rezapour, A., Tzeng, W.-G., Tsai, S.-C.: Rl-routing: An sdn routing algorithm based on deep reinforcement learning. IEEE Transactions on Network Science and Engineering **7**(4), 3185–3199 (2020)

[25] Ye, D., Zhang, M., Yang, Y.: A multi-agent framework for packet routing in wireless sensor networks. sensors **15**(5), 10026–10047 (2015)

[26] Kumar, S., Miikkulainen, R.: Dual reinforcement q-routing: An on-line adaptive routing algorithm. In: Proc. of the Artificial Neural Networks in Engineering Conf., pp. 231–238 (1997). Citeseer

[27] Macone, D., Oddi, G., Pietrabissa, A.: Mq-routing: Mobility-, gps-and energy-aware routing protocol in manets for disaster relief scenarios. Ad Hoc Networks **11**(3), 861–878 (2013)

[28] Hu, T., Fei, Y.: Qelar: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks. IEEE Transactions on Mobile Computing **9**(6), 796–809 (2010)

[29] Coutinho, N., Matos, R., Marques, C., Reis, A., Sargento, S., Chakareski, J., Kassler, A.: Dynamic dual-reinforcement-learning routing strategies for quality of experience-aware wireless mesh networking. Computer Networks **88**, 269–285 (2015)

[30] Jung, W.-S., Yim, J., Ko, Y.-B.: Qgeo: Q-learning-based geographic ad hoc routing protocol for unmanned robotic networks. IEEE Communications Letters **21**(10), 2258–2261 (2017)

[31] Wu, C., Kumekawa, K., Kato, T.: Distributed reinforcement learning approach for vehicular ad hoc networks. IEICE transactions on communications **93**(6), 1431–1442 (2010)

[32] Saleem, Y., Yau, K.-L.A., Mohamad, H., Ramli, N., Rehmani, M.H.: Smart: A spectrum-aware cluster-based routing scheme for distributed cognitive radio networks. Computer Networks **91**, 196–224 (2015)

[33] Costa, L.A.L., Kunst, R., Freitas, E.P.: Q-fanet: Improved q-learning based routing protocol for fanets. Computer Networks **198**, 108379 (2021)

[34] Arafat, M.Y., Moh, S.: A q-learning-based topology-aware routing protocol for

flying ad hoc networks. IEEE Internet of Things Journal (2021)

[35] Lahsen-Cherif, I., Zitoune, L., Vèque, V.: Energy efficient routing for wireless mesh networks with directional antennas: When q-learning meets ant systems. Ad Hoc Networks **121**, 102589 (2021)

[36] Lee, S., Yu, H., Lee, H.: Multi-agent q-learning based multi-uav wireless networks for maximizing energy efficiency: Deployment and power control strategy design. IEEE Internet of Things Journal (2021)

[37] Simi, S., Ramesh, M.V.: A reinforcement learning approach for improving internet connectivity in maritime network. Journal of Advanced Research in Dynamical and Control Systems **10**, 598–604 (2018)

[38] Rao, S.N., Raj, D., Aiswarya, S., Unni, S.: Realizing cost-effective marine internet for fishermen. In: 2016 14th Int. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), pp. 1–5 (2016). IEEE

[39] Unni, S., Raj, D., Sasidhar, K., Rao, S.: Performance measurement and analysis of long range wi-fi network for over-the-sea communication. In: 2015 13th Int. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), pp. 36–41 (2015). IEEE

[40] Dhivvya, J., Rao, S.N., Simi, S.: Towards maximizing throughput and coverage of a novel heterogeneous maritime communication network. In: Proc. of the 18th ACM Int. Symposium on Mobile Ad Hoc Networking and Computing, p. 39 (2017). ACM

[41] Surendran, S., Ramesh, M.V., Montresor, A., Montag, M.J.: Link characterization and edge-centric predictive modeling in an ocean network. IEEE Access **11**, 5031–5046 (2023)

[42] Watkins, C.J., Dayan, P.: Q-learning. Machine learning **8**(3), 279–292 (1992)

[43] Surendran, S., Ramesh, M.V., Montag, M.J., Montresor, A.: Modelling communication capability and node reorientation in offshore communication network. Computers & Electrical Engineering **87**, 106781 (2020)

Declarations

- **Funding** : We wish to confirm that there has been no significant financial support for this work that could have influenced its outcome.
- **Conflict of interest** : On behalf of all authors, the corresponding author states that there is no conflict of interest.