# QLAR- Q LEARNING BASED ACOUSTIC ROUTING FOR UNDERWATER SENSOR NETWORKS

**PRATHIBA N[1], MALA C S[2]**

[1]Assitant Professor, Department of Telecommunication Engineering, BMS Institute of Technology and Management, Bangalore-560064, India

[2]Professor, Department of Telecommunication Engineering, BMS Institute of Technology and Management, Bangalore-560064, India

E-mail: [1]pratibha.yashas@bmsit.in, [2]csmala63@gmail.com

## ABSTRACT

Underwater sensor networks have gained huge attention due to their significant use in monitoring and exploring the ocean, lakes, seas and rivers. The UWSNs are different from the ground based static sensor networks. The offshore networks or the networks which are deployed on portion of land have been explored widely but implementing the traditional protocols is not considered as a feasible solution because these networks suffer from several challenges such as high- water pressure, low bandwidth, delay and error rate etc. Therefore, we focus on introducing a novel routing protocol for underwater sensor networks by using Q learning based feedback mechanism. The proposed Q learning approach uses path cost, packet delivery probability and link analysis to maximize the payoff of the network. Moreover, the proposed approach is based on the opportunistic routing where the relay node is selected based on the maximum payoff. Finally, we present a simulation study to show the improved performance by using proposed approach where average network lifetime, delay and packet delivery values are obtained as 3522 rounds, 0.39s and 95.33, respectively.

**Keywords:** *Acoustic Networks, Q Learning, Opportunistic Routing, Underwater WSN*

## 1. INTRODUCTION

The use of underwater sensor networks (UWSNs) to monitor and explore lakes, rivers, seas, and oceans has recently been recognized as a strong technology. Water covers 2/3 portion of earth's surface but only a small portion of water has been investigated [1-6]. UWSNs are a useful study area for tackling underwater applications as a result. UWSNs have gained a lot of attention over the past two decades due to their extensive range of applications in numerous fields, including water pollution monitoring, collecting the underwater data, development of early warning system based on the underwater data, alert generation for disaster management, surveillance, and exploring new resources in underwater scenarios [7–10].

Acoustic communication is a method of transmitting and receiving messages via sound propagation in an underwater environment. A number of vehicles and sensors are deployed in an area by underwater sensor networks to carry out cooperative monitoring and data collecting duties [1]. Underwater sensors are often installed for capturing data at a fixed point while monitoring the ocean floor, and they are recovered after the completion of monitoring task. Lack of interactive connection between the different ends of sensors, inability to accurately capturing the data, and destruction of recorded data in the event of failure are the main drawbacks of the traditional approach. Moreover, these networks have several constraints such as restricted bandwidth, long propagation times, 3D topologies, and power limitations. At any oceanic location, radio and optical waves are impractical for communication. Due to their inherent constraints, underwater sensor networks are only able to use acoustic signals, a method that has been used by nature since the beginning of the ocean [4, 5]. In these environments, the speed of sound is considered as constant. However, the undersea environment's temperature, depth, and salinity have an impact on sound speed. The sound speed varies in an underwater environment due to several reasons [6]. Diverse acoustic users in the underwater environment share a significant portion of the underwater acoustic channel frequencies spectrum, particularly on mid-frequencies. The acoustic spectrum is still insufficiently used in underwater environments both temporally and spatially [7]. Utilizing an acoustic channel has grown challenging due to the variable undersea environment. For instance, fading and phase variations are brought on by multipath propagation, and the Doppler Effect also affects the process of sender and receiver nodes. Similarly, speed of sound and underwater noises affect the efficiency of acoustic channel [8].

In contrast to ground-based sensor network nodes, underwater sensor network nodes are not static. Instead, they move as a result of various undersea activities and environmental conditions, often at a speed of 2-3 m/s with water currents. Due to their rapid absorption in sea water, radio waves, which are the main form of communication for terrestrial networks, are not a viable option for UWSNs [2]. Under water, optical waves may travel short distances at high data rates [3]. They are only suited for short-range communication because they are susceptible to dispersion and rapid attenuation [4]. On the other hand, because acoustic waves can travel a greater distance, they are considered to be the best communication medium for use underwater [5]. However, the underwater acoustic channel has its own difficulties, including slow and variable sound wave speeds (average of 1500 m/s), a limited bandwidth that depends on transmission range and frequency, multipath effects, high bit error rates [6, Doppler's shift, channel asymmetry caused by moving currents [6,] and a lack of underwater satellite positioning systems. The radio waves used by satellite positioning systems for communication cannot go deeper than a few meters under water, making them ineffective for underwater localization [7].

In general, factors like communication distance, energy consumption, and network lifetime have an impact on efficiency of these networks. In order to complete the communication across sensor nodes, clustering-based communication models are commonly used. Since they may balance energy usage, extend the life of the networks, and decrease communication interference, clustering algorithms provide a significant answer to the UWSNs problem [10]. Additionally, routing protocols are crucial for enhancing networks' communication efficiency. In general, multihop communication is the main emphasis of clustering-based routing schemes, and relay nodes are chosen to carry the data packets. Numerous research has demonstrated that cluster routing algorithms are effective in preventing collisions. The existing methods employ multi-hop methods to transmit data between clusters while balancing traffic loads [8]. A cluster routing technique divides the nodes into several groups, each of which contains a head node (CHN) and numerous other nodes known as "cluster member nodes." The CHN assigns channels to transport data across as soon as clusters are created. Based on a distribution that could prevent collisions, the CMN delivers data [9]. Aggregation is subsequently handled by CHNs, which may minimize data redundancy and the quantity of data packets that must be delivered to the SN, saving energy [10].

### 1.1 Research contribution

Numerous studies have shown that clustering routing algorithms are more effective in reducing data transfer and controlling traffic. Additionally, they can reduce the amount of packets lost, conserve energy, and prolong the network's lifespan. Energy efficiency is a big concern since underwater sensors are limited by energy and it is time-consuming to change or recharge batteries. In order to overcome these issues, clustering based approaches are widely adopted but the traditional clustering based approaches are not feasible in underwater communication scenarios due to several parameters such as high water pressure, low bandwidth, delay and error rate etc. To overcome the aforementioned issues, we focus on development of a novel routing algorithm by using a novel clustering based mechanism. The main contribution of proposed approach are as follows:

**Contribution 1:** First of all, we focus on studying the current trends and techniques in this domain of underwater communication and report the challenges. This section provides the detailed analysis of existing schemes and drawbacks to boost the research direction.

**Contribution 2:** We present a novel routing method based on the energy and QoS aware clustering mechanism. Therefore, the proposed approach is developed by using combination of opportunistic routing and Reinforcement learning.

**Contribution 3:** The proposed approach focus on minimizing the packet collision to ensure the efficient packet delivery. To ensure the minimized collision, we incorporated game theory based model for better packet delivery.

Rest of the article is arranged in following sections: section 2 describes the brief overview of existing routing schemes for underwater networks, section 3 presents the proposed QLAR approach which uses Q learning and opportunistic routing based concept to determine the best routing path, section 4 presents the outcome of proposed approach and comparative analysis, finally, section 5 presents the overall remarks about proposed approach and future scope of the work.

## 2 LITERATURE SURVEY

This section presents the description about recent techniques in this field of underwater wireless communication. Several methods have been developed to improve the overall performance.

Karim et al. [1] reported that the dynamic nature of water waves is one of the most challenging issues in this domain which affects the performance of these networks. Therefore, ensuring the packet delivery and minimizing the energy consumption for these networks becomes a challenging issue. Thus, authors have suggested to incorporate geographic and opportunistic routing mechanisms which can use the relay node to rout the data efficiently. Authors introduced a new routing protocol called Geographic and Cooperative Opportunistic Routing Protocol (GCORP) which uses the cooperative mechanism along with geographic and opportunistic routing. according to this approach, multiple sink based network is established and later relay node is estimated by considering the depth based fitness factor. Later, the weights are computed of the node whose fitness is evaluated to identify its suitability for selecting the relay node.

Coutinho et al. [2] reported the challenges in UWSNs such as harsh environment and characteristics of acoustic channel. Therefore, authors adopted the concept of multi-modal communication for underwater sensor networks. Authors proposed a stochastic model which uses opportunistic routing for UWSNs. Further, two methods are designed for candidate selection are also described which are OMUS-E and OMUS-D. This helps to identify the most suitable modem for data transmission.

Zhou et al. [3] reported that energy consumption and latency are the main issues which affect the performance of underwater sensor networks. However, the routing schemes can be used to improve the communication. Thus, authors focused on anypath routing protocol by adopting the Q-learning based mechanism. The objective function of Q-leaning is designed based on the residual energy and depth information of sensor nodes for efficient routing.

Similar to this, sun et al. [4] discussed that the harsh environment poses several challenges in these networks. due to the unattended environments replacement of power sources is considered as challenging task which affects the network lifetime. Thus, developing the energy –efficient routing can help to minimize the energy consumption and prolong the network lifetime. To overcome these issues, authors presented adaptive clustering approach which is based on multi-agent reinforcement learning scheme. This approach uses reinforcement learning model to select the optimal route cooperatively. Furthermore, a cluster head selection approach is also developed which doesn’t incur any extra communication overhead and doesn’t require consensus of neighbouring node to be selected as cluster head. Additionally, a reward function is also defined to feedback the impact of clustering mechanism. This helps to identify the best node for relay and minimizes the energy consumption.

Gul et al. [5] developed Energy-Efficient Regional Base Cooperative Routing (EERBCR) protocol with sink mobility to enhance the network lifetime. According to this approach, the network is divided into 12 regions which are covered by four mobile sink nodes. Until the sink node enters their zone, all sensor nodes remain in sleep mode. When the sink node enters a region, it broadcasts a “Hello” message. The message is received by every node in that area, and they all turn on. Another packet is broadcast by the sink just before it leaves the area, alerting the nodes of its impending departure and allowing them to resume sleep mode.

Lilhore et al. [6] reported that due to dynamic nature of UWSNs maintaining the location and adding new nodes also becomes tedious task. Therefore, maintaining the energy efficient communication also considered as challenging task to prolong the network lifetime. In order to deal with this issue, authors presented depth controlled and energy efficient routing protocol which helps in adjusting the depth of lower energy nodes. Moreover, it swaps the lower energy nodes with higher energy nodes to ensure the efficient energy utilization without any wastage of energy resources. This method uses genetic algorithm approach and back propagation neural network for data fusion method to improve the energy efficiency of the network.

Wang et al. [7] developed energy balanced and lifetime extended routing protocol (EBLE) to improve the energy efficiency of the network. This method considers load balancing and optimizes the data transmission by selecting the low-cost paths. The complete work is carried out into two phases as candidate selection phase and data transmission phase. In the first phase, the forwarder nodes are selected based on the node position and energy information whereas second phase focus on selecting the path with low cost and nodes with high residual energy in the path.

Alsalman et al. [8] discussed that radio frequency is used for wireless networks but it fails for underwater communication scenarios therefore authors described the advantages of acoustic, optical and magnetic induction based communication standards. However, the challenges faced in UWSNs have serious impact on the network performance. In order to deal with this, authors presented machine learning based solution by considering the power, and latency parameters. The performance of this system is improved by incorporating Q learning approach. Similar to this, in [10] authors used Q learning based approach called as QELAR. With the aim of optimising network lifespan for UWSNs, QELAR is a single-path routing protocol based on the Q learning method. Each node's residual energy as well as the energy distribution among its neighbours are taken into consideration by the reward function. To balance the burden among the sensor nodes and extend the network lifetime, QELAR's routing path is chosen. Additionally, QELAR uses a retransmission mechanism after transmission failures to increase the dependability of data transfer.

DBR [11] is the first routing system for underwater sensor networks that makes use of node depth information. According to the concept of DBR, it collects the data from sensor nodes and adopts a greedy mechanism to transmit the data packets towards the direction of water surface. Additionally, it uses a packet holding mechanism where collected packets are hold for certain duration. This packet holding mechanism helps to increase the cooperation among the forwarder nodes which is used to select the nearest forwarder node. The average end-to-end delay also can be decreased with DBR's greedy method.

Similarly, authors in [12] presented energy efficient depth based routing protocol known as EEDBR. The residual energy of the sensor nodes is taken into account when choosing the forwarder node in EEDBR. This approach also uses the concept of holding time which helps to plan the packet forwarding by considering the residual energy. In order to balance the energy consumption across sensor nodes, EEDBR is an energy balanced algorithm.

In [13] authors introduced distance vector based routing protocol called as DVOR. This approach considers the hop count of sensor nodes toward the destination to determine the shortest routing path. It establishes distance vectors for each node using a query technique. Data packets can be routed using the shortest way in terms of hop counts by using the distance vectors, which hold the fewest hop counts to the sink. By reducing diversions during packet transfers based on distance vectors, DVOR can lower energy usage and average end-to-end latency.

## 2.1. Problem statement

Underwater communication networks face unique challenges due to the characteristics of the underwater environment, including limited bandwidth, high propagation delays, and variable channel conditions. Designing an efficient and reliable routing protocol for underwater networks is essential to enable effective communication among underwater nodes and support various applications such as underwater surveillance, environmental monitoring, and underwater exploration. The routing protocol needs to address issues such as energy efficiency, robustness to node mobility, adaptability to changing network topology, and the ability to handle the unique characteristics of underwater channels, including acoustic signal propagation and interference. Furthermore, the protocol should consider the dynamic nature of the underwater environment, accounting for factors such as node mobility, varying link quality, and potential obstacles. Thus, the challenge lies in developing an underwater routing protocol that can optimize data delivery, minimize energy consumption, ensure reliable communication, and adapt to the specific constraints and characteristics of the underwater environment. The aforementioned studies have focused on the improving the communication performance for underwater sensor networks but these protocols suffer from various challenges such as slow speed of acoustic signals than radio waves, refraction, absorption and scattering through the water. Moreover, energy consumption and network lifetime also considered as challenging issues which affects the network lifetime and QoS.

## 2.1. Research question

Based on this problem statement, we formulate some research question which are as follows:

- How routing protocols impact on the network’s performance
- Can machine learning or intelligent learning approaches be beneficial for underwater communication scenarios.
- What are some important factors and parameters which need to be considered to

evaluate the performance of underwater routing approaches?

## 3. PROPOSED MODEL

In this work, the sensor network model is considered for underwater scenarios where multiple sinks are assumed to collect the final data from different nodes. Let us consider that $\mathcal{N} = \{n_1, n_2, \dots, n_{|\mathcal{N}|}\}$ denotes the total deployed sensor nodes in the 3D area where these nodes are randomly distributed by sonobuoys as $\mathcal{S} = \{s_1, s_2, \dots, s_{|\mathcal{S}|}\}$ which are called as sink nodes which are placed at the surface of ocean. However, the complete network is considered as a non-mobile network architecture where sink nodes remain at the deployed location.

This can be obtained with the help of buoys or anchors. The considered network topology can be represented using directed graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ denotes the set of nodes as$\mathcal{V} = \{\mathcal{N} \cup \mathcal{S}\}$ and $\mathcal{E}$ denotes the edges of network. The edge $e_{ij}$ exists only if the node $n_i$ has the probability of more than 0 to communicate with node $n_j$.

Each underwater sensor $i \in \mathcal{V}$ required some certain transmission power levels as $\mathcal{P} = \{p_t^1, p_t^2, \dots p_t^L\}$. Let us consider that neighboring node set $\mathcal{N}_i^{p_t^k}$of node $i$ consist of nodes that can observe $i$'s transmission when it is utilizing $k^{th}$ transmission power level. The communication link between sender $i$ and $j$ is denoted as $\mathcal{L}_{i\to j}^{p_t^k}$and this link has a packet delivery probability as $\mathbb{P}_{\mathcal{L}_{i\to j}^{p_t^k}} \in [0,1]$ which is based on the power consumption, distance and packet size. This link probability is further described to identify the relay node in opportunistic routing.

In this work, our main aim is to address the challenges faced by traditional methods and developing the new routing protocol to overcome the existing issues. As discussed in section 2, learning based schemes are widely adopted in this domain and these techniques have proven their significance to improve the communication performance. Moreover, opportunistic routing also plays important role in these networks where frequent link disruption is faced due to unstable network architecture. Thus, we focus on combining the opportunistic routing with machine learning paradigm to prolong the network lifetime and improve the Quality of Service (QoS) of the network.

### 3.1 Overview of basic routing modules

This section briefly describes the traditional opportunistic routing and reinforcement learning approach. The opportunistic routing approach is used for efficient data forwarding by using the relay node selection mechanism and reinforcement learning helps to identify the best suitable node as relay node to facilitate the communication which can improve the overall performance of the network.

The opportunistic routing is a routing process which is widely used in wireless sensor networks to ensure the packet delivery and increase the QoS. This approach doesn't rely on the single relay node for forwarding the data packets rather neighbouring node cooperation is also included to forward the data packet. This process of opportunistic routing helps to minimize the packet loss rate. Figure 1 shows a sample network where node $n_1$ is holding the packet and, generally, it transmits the packet to the neighbouring node $n_2, n_3$ and $n_4$ at the same time. during this process, each node can assign a timer based on their priority. The applied timer helps to estimate the time required to hold the data copies.

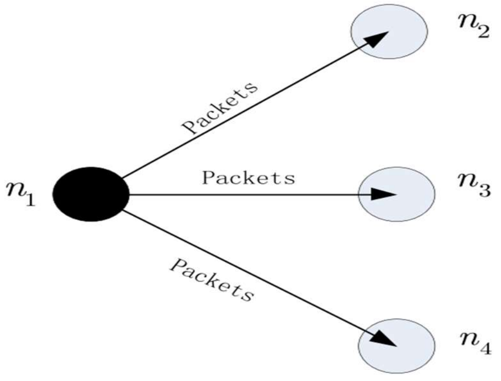

*Figure 1: Opportunistic Model*

Figure 2 demonstrates the complete overview of opportunistic routing where node $A$ is the source and $D$ is the destination. Nodes $B_1, B_2, B_3$ are closer to $A$ therefore selected as forwarder nodes. As the node A broadcasts the message, all neighbouring node $B_1, B_2, B_3$ receives the packets. Let us assume that node $B_1$ has the highest priority then it takes the task of packet forwarding to the neighbouring nodes. Similarly, $C_1, C_2$, and $C_3$ are the next neighbouring node where $C_1$ has the highest priority then $C_1$ is responsible for packet forwarding but let us assume that the node $C_1$ misses this process then the node $C_2$ which is having the second highest priority will be selected for packet forwarding to the destination node $D$

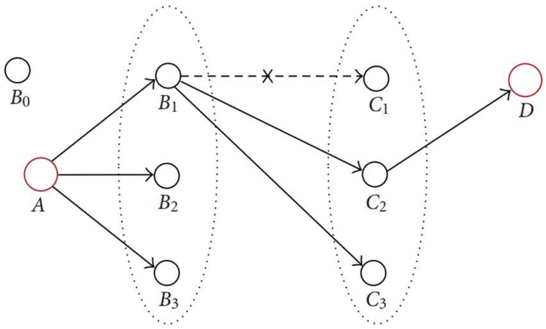

*Figure 2: Opportunistic Routing Forwarder Node Selection Process*

Currently, this approach is improved by incorporating intelligent machine learning schemes which analyze the behaviour of the nodes and select the best relay node. Reinforcement learning is considered as one of the most important approaches which doesn't require any prior knowledge of environment and labelled data to analyze the behaviour whereas it based on the reward mechanism. In general, the Reinforcement learning is a process which is used to map the environment behaviour to maximize the rewards. These rewards are assigned based on the actions taken by the agent in the given environment. The reward is used to analyse the action quality i.e., if the action taken leads to increase the performance of the system, then it is considered as a good action and positive reward is assigned to it otherwise the negative or poor reward is assigned for that action. This reward mechanism avoids considering the poor action repeatedly increase the use of action which have gained positive/good reward. Figure 3 depicts the general architectural module of reinforcement learning.

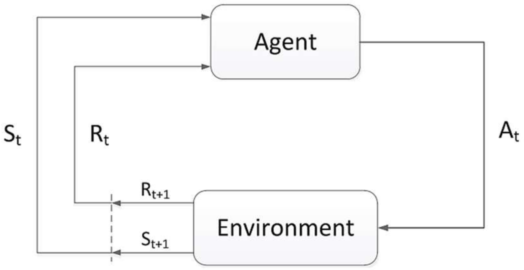

*Figure 3: Reinforcement Learning*

According to this process, the agents are responsible to interact with the environment which later generates the new states. The agent interacts with the surrounding environment to generate the new state. Based on this interaction, the agent receives a reward. In this process, environment and agents continually interact to generate fresh data. These agents learn from these strategies and adopt the strategy which is most suitable to accomplish the task in iterative manner.

### 3.2 Proposed Game Theory Based Reinforcement Learning Opportunistic Routing Model

This section presents the proposed solution for routing in underwater sensor networks. In this stage, each node makes individual decision by considering several parameters such as hop count, delay in packet delivery, and energy depletion etc. As discussed before, the learning approach focuses on maximizing the own profit based on the reward mechanism. In this work, we model the sensor network in such a way that these nodes focus on self-decision making to obtain the routing path. Based on this assumption, we model the complete problem in the form of routing task as $\mathbb{G} = \{\mathbb{N}, \mathbb{A}_{i,1\le i\le n}, U_{i,1\le i\le n}, \mathbb{U}_{i,1\le i\le n}[m], \mathbb{S}_{i,1\le i\le n}, T\}$ at considered time stamp $t$. The components of this model are described below:

- $\mathbb{N}$ represents the finite set of nodes in group which consist of sensor nodes as $\mathbb{N} = \{\delta_1, \delta_2, .., \delta_n\}$ where $\delta_i$ characterise the $i^{th}$ sensor node and $n$ is used to denote the total number of nodes in the group.
- The proposed model is based on the reinforcement learning which require some finite set of action given as $\mathbb{A}_i = \{a_i^1, a_i^2, .., a_i^m\}$ where $a_i^m$ denotes the selection of $m^{th}$ neighbouring node. This node is used to transmit the data packet where $m$ is used to denote the available number of neighbouring nodes.
- The reward mechanism of reinforcement learning depends on the payoff received by the node $\delta_i$ which is denoted as $U_{i,1\le i\le n}$, similarly, this node $\delta_i$ has the payoff history which is denoted as $\mathbb{U}_i[m] = [\mathcal{U}_i^1, .., \mathcal{U}_i^m]$ where$\mathcal{U}_i^m$ is the accumulated payoff of strategy $a_i^m$
- These nodes perform some set of action in the given state and identify new state. The state is represented as $\mathbb{S}_i = \{s_i^1, s_i^2, \ldots, s_i^m\}$. The state, and action are based on the employed routing approaches. If $\delta_i$ takes an action $a_i^m$ the next state of node $\delta_i$ is given as $s_i^m$
- Finally, $T$ denotes the time as $T = \{0,1, \ldots t, t+1, ..\}$

### (a) Reinforcement learning model

According to the reinforcement learning process, the agent nodes use this mechanism to maximize their final reward. The RL approach uses an environment model to perform certain task which is formulated with the help of Markov decision process (MDP). The MDP is considered the basic requirement, mathematical framework to which models the decision making process. Moreover, it is used to solve the optimization problem by using learning algorithms. In this stage, we have considered MDP which consists of states ($\mathbb{S}$), actions ($\mathbb{A}$), transition probabilities ($P$) which consist of probability of state transition from one state $s$ to another state $s'$ given as $P^{a}_{s\to s'}$ with a given action $a \in \mathbb{A}$ and rewards are denoted as $\mathcal{R}(s,a)$ for action $a$ from state $s$. Based on this, the transition probability $P^{a}_{s\to s'}$and $\mathcal{R}(s,a)$ is defined as follows:

$$\mathcal{R}(s,a) = \{\mathcal{R}_t(s_t,a_t)|s = s_t, a = a_t\} = \sum_{s_{t+1}\in S}\left(P^{a}_{s_t\to s_{t+1}} \times \mathcal{R}^{a_t}_{s_t\to s_{t+1}}\right) \quad (1)$$

where,

$$P^{a}_{s\to s'} = P_r\{S_{t+1} = S'|s = s_t\}, s.t., \sum_{s'\in S} P^{a}_{s\to s'} = 1 \quad (2)$$

where $s_t$ is the state and $a_t$ is the action at time $t$ whereas $\mathcal{R}^{a}_{s\to s'}$ is used to denote the obtained reward for any action which is performed at state $s$. According to the concept of reinforcement learning, a policy is designed which maps states $s \in \mathbb{S}$ and action $a \in \mathbb{A}$ to a probability function which denotes the policy of considered environment as $\psi(s,a)$. This policy denotes is obtained by taking any action $a$ in the given state $s$. According to the concept of this approach of Q learning, any policy is expected to receive a reward, in this case this is expressed as $\left(V^{\psi}(s)\right)$ which is received in state $s$ for policy $\psi$ at time $t$. The overall expected reward can be expressed as:

$$V^{\psi}(s) = E_{\psi}\{\mathcal{R}_t(s_t,a_t)\}$$
$$= E_{\psi}\{\sum_{k=0}^{\infty}(\gamma^k \times \mathcal{R}_{t+k}(s_{t+k},a_{t+k}))\}, s.t. \gamma \in [0,1) \quad (3)$$

Where $\gamma$ is the control factor which discounts the reward in future transitions. With the help of Bellman's principle, the expected reward can be rewritten as:

$$V^{*}(s) = \max_{\psi} V^{\psi}(s)$$
$$= \max_{\psi} E_{\psi}\left\{\sum_{k=0}^{\infty}(\gamma^k \times \mathcal{R}_{t+k}(s_{t+k},a_{t+k}))\right\}$$
$$= \max_{a}\left[\mathcal{R}_t(s_t,a_t) + \left(\gamma \times (\sum_{s_{t+1}\in S}(\left(P^{a_t}_{s_t\to s_{t+1}} \times V^{*}(S_{t+1})\right)\right)\right] \quad (3)$$

This optimality problem can be solved by applying reinforcement learning approach. In this work, we adopt the Q learning based mechanism which successively improves the evaluations of parameters where state action pair is denoted as $Q^{\psi}(s,a)$ which denotes that the action $a$is taken in state $s$ under the given policy $\psi$. The Q learning based state-action can be expressed as:

$$Q^{\psi}(s,a) = E_{\psi}\left\{\sum_{k=0}^{\infty}(\gamma^k \times \mathcal{R}_{t+k}(s_{t+k},a_{t+k}))\right\}$$
$$= \left[\mathcal{R}_t(s_t,a_t) + \left(\gamma \times (\sum_{s_{t+1}\in S}(\left(P^{a_t}_{s_t\to s_{t+1}} \times V^{*}(S_{t+1})\right)\right)\right] \quad (4)$$

Further, this $Q$ value is used for updating the rewards $V$

In proposed approach, each sensor node consists of several information parameters such as path cost, payoff history, and packet delivery probability. The complete information is used to formulate the routing decision. In order to compute the packet delivery probability, we compute the Bit Error Rate (BER) for underwater Rayleigh fading channel, denoted as:

$$Pe\left(SNR_d^{p_t^k}\right) = \frac{1}{2}\left(1 - \sqrt{\frac{SNR_d^{p_t^k}}{1+SNR_d^{p_t^k}}}\right) \quad (5)$$

where $SNR$ denotes the signal to noise ratio. The probability to transmit the $m$ bit data successfully over an underwater communication link $L^{p_t^k}_{i\to j}$ between node $i$ and $j$ for $d$ distance is given as:

$$p_{L^{p_t^k}_{i\to j},d,m} = \left(1 - Pe\left(SNR_d^{p_t^k}\right)\right)^m \quad (6)$$

The packet delivery probability is used for updating the $Q$ values

Further, we use Link cost between two nodes and the overall path cost is obtained by summing up all link obtained between source and destination node. The link cost $L_c$ can be expressed as:

$$L_c(i,j) = \left[(1-\alpha_i) \times \frac{d_{i,j}}{D_M}\right] + \left[\alpha_i \times \left(1 - \frac{\mathcal{E}_i}{E_M}\right)\right] \quad (7)$$

Where $d_{i,j}$ denotes the distance between node $i$ and $j$,$\mathcal{E}$ is the residual energy of node $j$, $E_M$ is the initial energy and $D_M$ is the maximum coverage of sensor node. During this process of topology establishment, each underwater node estimates its $L_c$ of its neighboring nodes. However, he nodes which are placed at one hop of the sink node consider their path cost as one-hop $L_c$, and remaining node compute their path cost as:

$$\mathcal{P}_c = \min_j \left\{P_{c_j} + L_c(i,j)\right\}, s.t. j \in \mathcal{N}_i \qquad (8)$$

Where $\mathcal{N}_i$ expresses the set of neighboring nodes. Further, path cost $\mathcal{P}_c$ is used to realize the payoff. Thus, the $\mathbb{U}[.]$ is adjusted as:

$$\mathbb{U}_i[j]_{j\in\mathcal{N}_i} = \mathcal{P}_{c_j} + L_c(i,j) + p_{L_{i\to j}^{p_t^k},d,m} \qquad (9)$$

### (b) Applying the reinforcement learning for routing

In order to apply the aforementioned model, we consider multiplayer Markov decision process where each underwater sensor node focuses on achieving a routing policy in such a way $\psi_i^* \in \mathbb{A}_i$ which can maximize the payoff, here $\mathbb{A}_i$ is the action and $\psi$ denotes the routing decision for underwater sensor nodes. Each node considers determines its states based on these routing decisions. Let us consider that $s_i(t) \in \mathbb{S}$ denotes the state and $a_i(t) \in \mathbb{A}$ is the corresponding action of UW node at time $t$. Similarly, the payoff for the considered state $s_i$ and action $a_i$ is denoted as $U_i\big(s_i(t), a_i(t)\big)$ where $U_i\big(s_i(t), a_i(t)\big) \to \mathbb{R}$. At this stage, $U_i(s,a)$ is interpreted as reward $\mathcal{R}$ in $Q^{\psi}(s,a)$.

Here, main aim of Q-learning is to increase the packet delivery to the sink node along with the maximized payoff which lead to improving the QoS. The proposed underwater routing approach considers relay node selection to ensure the packet delivery based on the concept of opportunistic routing. Let $m^{th}$ neighbouring node from the group is assigned as relay node as $a_i(t) = a_i^m$ then the payoff can be given as:

$$U_i(s_i(t), a_i(t) = a_i^m) = \mathcal{R}(s_i(t), a_i^m) = \left(\frac{1}{P_{c_m}^{t} + L_c(i,j) + p_{L_{i\to j}^{p_t^k},}}\right) \times (-\varphi)^{\mathcal{Q}(a_i^m)} \qquad (10)$$

Where value of $\mathcal{Q}(a_i^m) = 1$ if the packet delivery is successful by performing action $a_i$ otherwise it is considered as 0. $\varphi$ denotes the penalty factor for failure of packet delivery. Based on this $U_i$ the value of final payoff can be reorganised as:

$$\mathbb{U}_i[m] = \mathcal{U}_i^m(t+1) = \mathcal{J} + \left(\frac{1}{\eta} \times \{U_i(s_i(t), a_i(t) = a_i^m) - \mathcal{J}\}\right) \qquad (11)$$

$$\text{s.t. } \mathcal{J} = \frac{1}{t} \times \left(\sum_{e=`}^{t} U_i\big(s_i(e), a_i(e)\big)\right) \qquad (12)$$

Where $\eta$ denotes the weight parameter. In this process, the underwater node selects an action $a_i$ which increases the maximizing the payoff $U_i(.)$ resulting in maximizing the $\mathcal{V}$ as $\mathcal{V}^*(s) = \max_a Q^*(s,a)$. Thus, the complete routing approach is based on the maximizing the payoff according to the action performed by the node to select the next relay node. The relay node is selected based on the action which increases the $Q$ value of the current state. However, this process of Q learning considers link cost, packet delivery probability, and path cost to obtain the final relay node.

## 4. RESULTS AND COMPARATIVE ANALYSIS

This section presents the outcome of proposed QLAR approach and compared the obtained performance with existing schemes. We have considered a predefined sensor network region where sensor nodes are deployed randomly and stationary sink nodes is placed on the surface. The maximum transmission range is set to 150 meters, packet transmission energy is 2W, and energy consumed in packet reception is 0.5W. The detailed parameters are given in table 2. The performance of proposed QLAR is evaluated in terms of average end-to-end delay, network lifetime and packet delivery

*Table 1: Parameters considered*

| Parameter Name | Considered value |
|---|---|
| No. of Sensor Nodes | 100-500 |
| Power for packet transmission | 2W |
| Power for packet reception | 0.5 |
| Packet size | 150 bytes |
| Max. transmission range of sensor | 150m |
| Discount factor for Q learning | 0.8 |

The outcome of QLAR method is compared with existing protocols of underwater sensor networks such as QLFR [10], DBR [11], EEDBR [12], DVOR [13] and QELAR [14]. The QLFR uses the concept of anypath routing for underwater sensor networks, QELAER uses Q learning based model to balance the workloads. On the other hand, DBR, DVOR and EEDR are based on the localization free anypath routing approach.

First of all, we measure the average end-to-end delay performance for varied number of nodes. Figure 4 shows the obtained performance and comparative demonstration of existing and proposed QLAR approach. According to this experiment, the proposed approach achieves better performance by reducing the overall delay in packet transmission.

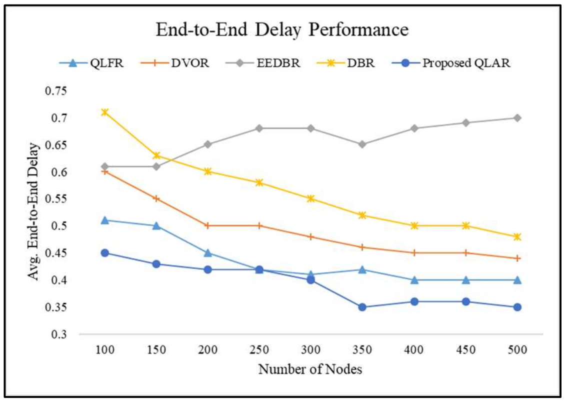

*Figure 4: Average End-To-End Delay Performance*

In this experiment, the average delay is obtained as 0.434, 0.492, 0.661, 0.563, and 0.393 by using QLFR, DVOR, EEDBR, DBR, and Proposed QLAR, respectively. As discussed before, the DBR and DVOR schemes are based on the approach where packets are hold for certain duration which leads to increase the latency and delay. Moreover, DBR, DVOR and EEDBR utilize one-hop routing information whereas the QLFR uses Q learning approach to formulate the decision. However, the proposed approach also uses Q learning model but we consider different parameters to ensure the maximum payoff for packet delivery.

Further, we measure the network lifetime performance for varied number of nodes. Figure 5 demonstrates the obtained network lifetime performance. In this experiment, the average performance is obtained as 3428, 895, 960, 838, and 3522 rounds by using QLFR, DVOR, EEDBR, DBR, and Proposed QLAR, respectively.

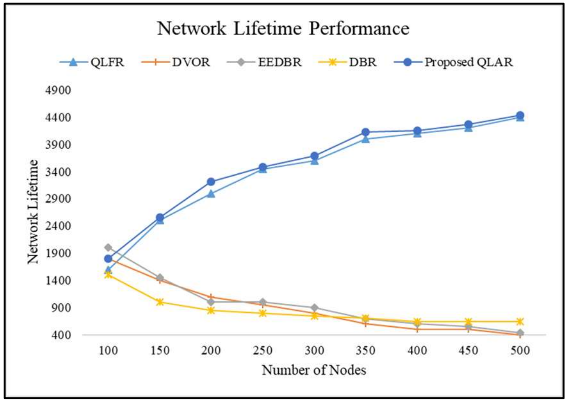

*Figure 5: Network Lifetime Performance for Varied Number of Nodes.*

In the case of DBR approach, increased number of sensor nodes leads to redundant data transmission resulting in increased energy consumption which degrades the network lifetime performance. On the other hand, the DVOR approach is based on the shortest path approach for packet transmission. This helps to reduce the overall energy consumption. However, nodes which lie in the computed shortest path, become overburdened therefore frequent switching occurs which also affects the network lifetime. On the other hand, EEDBR uses residual energy parameter which can prolong the network lifetime but this also suffers from the excessive packet redundancy. In contrast to these approaches, the proposed approach uses link cost and packet delivery probability computation which ensures the better path selection.

Finally, we measure the packet delivery performance and compared the obtained performance with existing schemes as depicted in figure 6.

According to this experiment, the average packet delivery performance is obtained as 0.93, 0.89, 0.88, 0.86, and 0.95 by using QLFR, DVOR, EEDBR, DBR, and Proposed QLAR, respectively.

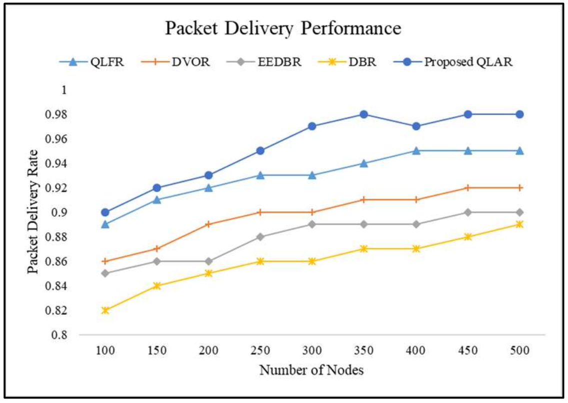

*Figure 6: Packet Delivery Performance*

## 5. CONCLUSION

Underwater sensor networks have gained huge attention in academia and industrial applications due their wide range of applications in acoustic monitoring and route discovery for underwater application scenarios. However, placement of sensor node and packet routing is one of the challenging tasks in UWSNs. Therefore, in this work we focus on developing a novel approach for routing in underwater sensor networks. In this work, we have introduced Q-learning based approach which is based on the feedback mechanism. This feedback mechanism helps to learn from environment and improve the overall performance. The proposed approach uses path cost, link cost and packet delivery probability to update the Q value by achieving maximum payoff. Moreover, the proposed approach is based on the opportunistic routing model where the suitable relay node is selected based on the

Q-learning approach. We have presented a simulation study to show the robustness of proposed approach and measured the performance in terms of average end-to-end delay, lifetime and packet delivery for varied nodes. However, mobility management, secure communication and computational complexities are still challenging issues for large underwater networks.

## REFERENCES:

[1] Karim, S., Shaikh, F. K., Chowdhry, B. S., Mehmood, Z., Tariq, U., Naqvi, R. A., & Ahmed, A. (2021). GCORP: Geographic and cooperative opportunistic routing protocol for underwater sensor networks. *IEEE Access*, *9*, 27650-27667.

[2] Coutinho, R. W., & Boukerche, A. (2021). OMUS: efficient opportunistic routing in multi-modal underwater sensor networks.IEEE Transactions on Wireless Communications, 20(9), 5642-5655.

[3] Zhou, Y., Cao, T., & Xiang, W. (2020). Anypath routing protocol design via Q-learning for underwater sensor networks. IEEE Internet of Things Journal, 8(10), 8173-8190.

[4] Sun, Y., Zheng, M., Han, X., Li, S., & Yin, J. (2022). Adaptive clustering routing protocol for underwater sensor networks. Ad Hoc Networks, 136, 102953.

[5] Gul, H., Ullah, G., Khan, M., & Khan, Y. (2021). EERBCR: Energy-efficient regional based cooperative routing protocol for underwater sensor networks with sink mobility. Journal of Ambient Intelligence and Humanized Computing, 1-13.

[6] Lilhore, U. K., Khalaf, O. I., Simaiya, S., Tavera Romero, C. A., Abdulsahib, G. M., & Kumar, D. (2022). A depth-controlled and energy-efficient routing protocol for underwater wireless sensor networks. *International Journal of Distributed Sensor Networks*, *18*(9), 15501329221117118.

[7] Wang, H., Wang, S., Zhang, E., & Lu, L. (2018). An energy balanced and lifetime extended routing protocol for underwater sensor networks. Sensors, 18(5), 1596.

[8] Alsalman, L., & Alotaibi, E. (2021). A Balanced Routing Protocol Based on Machine Learning for Underwater Sensor Networks. IEEE Access, 9, 152082-152097.

[9] Guan, Q., Ji, F., Liu, Y., Yu, H., & Chen, W. (2019). Distance-vector-based opportunistic routing for underwater acoustic sensor networks. *IEEE Internet of Things Journal*, *6*(2), 3831-3839.

[10] Zhou, Y., Cao, T., & Xiang, W. (2020). Anypath routing protocol design via Q-learning for underwater sensor networks. IEEE Internet of Things Journal, 8(10), 8173-8190.

[11] H. Yan, Z. J. Shi, and J.-H. Cui, “Dbr: depth-based routing for underwater sensor networks,” in international conference on research in networking. Springer, 2008, pp. 72–86.

[12] A. Wahid and D. Kim, “An energy efficient localization-free routing protocol for underwater wireless sensor networks,” International journal of distributed sensor networks, vol. 8, no. 4, p. 307246, 2012.

[13] Q. Guan, F. Ji, Y. Liu, H. Yu, and W. Chen, “Distance-vector-based opportunistic routing for underwater acoustic sensor networks,” IEEE Internet of Things Journal, vol. 6, no. 2, pp. 3831–3839, 2019.

[14] T. Hu and Y. Fei, “Qelar: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks,” IEEE Transactions on Mobile Computing, vol. 9, no. 6, pp. 796–809, 2010