

## Article

# An On-Site-Based Opportunistic Routing Protocol for Scalable and Energy-Efficient Underwater Acoustic Sensor Networks

Rongxin Zhu <sup>1,2</sup>, Xiwen Huang <sup>3</sup>, Xiangdang Huang <sup>1,3</sup>, Deshun Li <sup>1,3</sup> and Qiuling Yang <sup>1,3,\*</sup> 

<sup>1</sup> School of Cyberspace Security, Hainan University, Haikou 570228, China

<sup>2</sup> School of Mathematics and Information Science, Nanjing Normal University of Special Education, Nanjing 210038, China

<sup>3</sup> School of Computer Science and Technology, Hainan University, Haikou 570228, China

\* Correspondence: yql0515@163.com

**Abstract:** With the advancements in wireless sensor networks and the Internet of Underwater Things (IoUT), underwater acoustic sensor networks (UASNs) have attracted much attention, which has also been widely used in marine engineering exploration and disaster prevention. However, UASNs still face many challenges, including high propagation latency, limited bandwidth, high energy consumption, and unreliable transmission, influencing the good quality of service (QoS). In this paper, we propose a routing protocol based on the on-site architecture (SROA) for UASNs to improve network scalability and energy efficiency. The on-site architecture adopted by SROA is different from most architectures in that the data center is deployed underwater, which makes the sink nodes closer to the data source. A clustering method is introduced in SROA, which makes the network adapt to the changes in the network scale and avoid single-point failure. Moreover, the Q-learning algorithm is applied to seek optimal routing policies, in which the characteristics of underwater acoustic communication such as residual energy, end-to-end delay, and link quality are considered jointly when constructing the reward function. Furthermore, the reduction of packet retransmissions and collisions is advocated using a waiting mechanism developed from opportunistic routing (OR). The SROA realizes opportunistic routing to choose candidate nodes and coordinate packet forwarding among candidate nodes. The scalability of the proposed routing protocols is also analyzed by varying the network size and transmission range. According to the evaluation results, with the network scale ranging from 100 to 500, the SROA outperforms the existing routing protocols, extensively decreasing energy consumption and end-to-end delay.

**Keywords:** routing protocol; underwater sensor networks; Q-Learning; clustering; Internet of Underwater Things (IoUT)



**Citation:** Zhu, R.; Huang, X.; Huang, X.; Li, D.; Yang, Q. An On-Site-Based Opportunistic Routing Protocol for Scalable and Energy-Efficient Underwater Acoustic Sensor Networks. *Appl. Sci.* **2022**, *12*, 12482. <https://doi.org/10.3390/app122312482>

Academic Editor: Subhas Mukhopadhyay

Received: 9 November 2022

Accepted: 1 December 2022

Published: 6 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

During recent years, Underwater Acoustic Sensor Networks (UASNs) have gained much attention for the potential to explore and monitor the underwater environment [1]. UASN is one of the fundamental techniques of the Internet of Underwater Things (IoUT), which was developed from the concept of terrestrial Internet of Things (IoT) [2]. Large-scale UASN enables the extension of IoT to ocean applications, considered to be a promising solution for exploring the oceans [3]. One of the key problems for these applications is how to collect and forward the sensed data from the source node to the sink node [4].

Owing to the unique features of the underwater acoustic environment, routing in UASNs confronts crucial challenges, such as signal propagation delay, limited bandwidth, and low energy efficiency [5]. Scaling up or down the network size according to the actual demand in the underwater environment is usually necessary, while maintaining the reliability of the network. Adaptive formation of the network is taken into consideration in this article, enabling nodes to independently join or leave the network. Moreover, the

propagation delay has a significant impact on energy consumption, resulting in node failure due to fast energy depletion. Hence, centralized topology should avoid the failure of a single node, which could make the overall network crash. Moreover, due to the serious loss of signals and multipath effect, the packet loss rate of the acoustic channel leads to the unreliable transmission of data packets [6]. In addition, the energy of battery is constrained, and the battery is expensive to recharge or replace owing to the harsh underwater environment. To make things worse, the communication energy cost is greater than radio communications. In this context, to overcome energy constraints in UASNs, energy efficiency should also raise great attention [7]. Therefore, a scalable and energy-efficient routing protocol urgently needs to be designed.

Recently, numerous routing techniques for UASNs have been proposed. To improve the network scalability, Hindu et al. [8] proposed a Self-Organizing and Scalable Routing Protocol. The proposed protocol makes use of a multi-hop communication method to send sensed data to the sink node, and each node creates its own routing table by utilizing control packet broadcasts during the startup and neighbor discovery phases. While the protocol makes the communication efficient and the network scalable, the energy consumption for control packets is relatively high. To decrease energy consumption, Nicolaou et al. [9] proposed the hop-by-hop vector-based forwarding (HHVBF) routing protocol to reduce energy consumption and network latency, in which the forwarders are selected within the virtual pipeline. The performance of HHVBF is highly related to the pipeline radius, and thereby short or large radius will result in the collisions and low delivery ratio, respectively. Anand et al. [10] introduces another calculation called compelling Energy Resourceful Routing utilizing cost work. This calculation chooses the course that fulfills the nature of the administration in vitality stable, postpone requirements and throughput, hub in a flexible, and reduces power usage which drags out the lifetime of the system. Moreover, intelligent algorithms are utilized by routing protocols in UASNs. These routing protocols design the reward function, taking remaining energy into account. For example, Hu et al. [11] presented a Q-learning-based routing protocol for energy-efficient and lifetime-extended underwater sensor networks (QELAR). The QELAR applied the Q-Learning approach to the underwater sensor network. Routing decisions were made using a reward function that took residual energy into account. However, the link quality is ignored, and propagation delay is not seriously considered, which may result in unreliable transmission when the distances between nodes are far from each other. Furthermore, routing algorithms that rely on instant rewards may become stuck in local optima instead of discovering global optima.

In addition to the scalability and energy efficiency, the reliability of the network is also critical to UASN. Opportunistic routing (OR) has been adopted by UASN to improve the reliability in wireless networks [12]. The opportunistic forwarding method can reduce packet loss while avoiding retransmission [13]. Nonetheless, the majority of existing OR protocols lack an alternation mechanism for sorting the priorities of relay set nodes, resulting in the frequent participation of dominating nodes in forwarding and nonuniform energy consumption over the network. Coutinho et al. [14] proposed the GEDAR, which is a geographic and opportunistic routing protocol. Each node in the protocol greedily chooses the forwarding node with the highest expected packet advance (EPA), and the EPA is proportional to the distance between the nodes. The authors additionally designed a recovery mode for the void node based on the depth adjustment to address void routing. However, the greedy criterion and depth adjustments consume a lot of energy in packet forwarding while energy is extremely important for acoustic signal propagation in UASNs.

Moreover, many researchers propose routing protocols to overcome the constraints of the undersea environment, which are based on routing strategies [15]. To some extent, the forwarding algorithm of protocols can be utilized to reduce latency and energy consumption, while the key problem remains unresolved. In terms of the direction of data forwarding, the traditional underwater sensor network usually works in the way that data packets are sent upward from the bottom to the ocean surface, namely the land data

center [11]. However, the vast majority of monitoring data originates from the deep sea, hence long paths caused by the traditional architecture result in more energy consumption and transmission delay. Tilak et al. [16] shows that the major source of energy consumption is bulk data and long transmission distances, particularly in the underwater environment. Based on these facts, an on-site architecture is taken into consideration in UASNs, which deploys data centers under the sea. Fortunately, recent studies have demonstrated the viability of locating servers underwater [17]. With the on-site architecture, energy consumption and transmission delay can be significantly reduced. The acoustic channel in the deep sea is less affected by seasonal changes and transmission quality is much better, when comparing with the sea surface. Moreover, the cost of deploying and maintaining large-scale servers will be significantly lowered [18].

Therefore, to enhance the performance of opportunistic routing and intelligent routing algorithms, based on the on-site architecture, we propose a scalable and energy-efficient routing protocol (SROA) that applies the Q-Learning technique to the OR paradigm in UASNs. The SROA protocol is a clustered-based protocol with four phases and finds the optimal routing paths to achieve scalable and energy-efficient transmission.

Three main contributions of this paper are summarized in the following.

- (1) We apply a novel on-site architecture to the proposed protocol, locating the data center near the data source on the seabed. The on-site architecture can effectively minimize the number of forwarding hops in routing by shortening the distance between the source and sink nodes, lowering the hop count in routing and enhancing transmission reliability.
- (2) We group the network into a number of clusters by an unsupervised learning algorithm. Besides, to improve the reliability of the network and to avoid the failure of a single-point, a mechanism for the selection of the cluster head and potential cluster head is designed, which both takes the residual energy and location of the nodes into account.
- (3) We introduce the Q-Learning algorithm to the OR paradigm and elaborately design the reward function for Q-Learning, which jointly considers the factors of residual energy, delay, and PDR. In addition, a waiting mechanism based on the computed Q-value is designed to improve transmission reliability and reduce packet conflicts via the OR broadcast features, making the routing protocols reliable and scalable.
- (4) To be more realistic, different communication ranges and network scales are set. The overall performance of SROA is evaluated and compared to existing routing protocols.

The remainder of this paper is organized as follows. Section 2 introduces the system model and fundamental concept of reinforcement-learning. The SROA protocol is then introduced in Section 3. Section 4 evaluates the performance of the proposed protocol. Lastly, Section 5 concludes this paper.

## 2. Network Model

We mainly introduce the system model, underwater acoustic model, and machine learning algorithm adopted by SROA in this section.

### 2.1. System Model

The network architecture of the proposed SROA is shown in Figure 1. Sensor nodes are randomly deployed underwater and divided into a number of clusters. Each cluster elects the cluster head and potential cluster head for data transmission between clusters. In order to transmit the sensed data through acoustic channels, sink nodes are deployed on the seabed that are integrated with the underwater data center, making it convenient for aggregating or processing sensed data, then forward the sensed data to the terrestrial base station for further data analysis. The deployed sensor nodes collect data from the surrounding environment and the sensed data will be sent to the sink node through multi-

hop forwarding. In a three-dimensional system, the distance between the two sensor nodes [9] is calculated as follows,

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (1)$$

where  $(x_i, y_i, z_i)$  are the location coordinates for the  $i$  node.

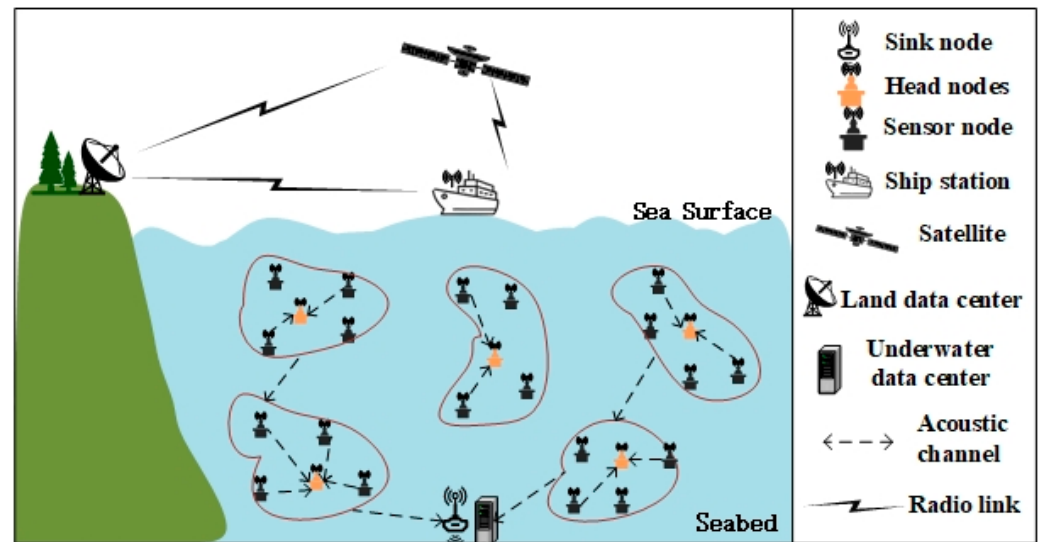


Figure 1. Network model.

Some assumptions:

- (1) The sink node and sensor nodes can obtain its location information. Sensor nodes can obtain the location of the sink using the existing services [19,20].
- (2) The initial energy of underwater sensor nodes is same; however, the sink node is not restricted by energy. Each node has the ability to keep its recent communication records in local storage [21].
- (3) Nodes in a particular layer send packets to that layer's cluster head, who then transmits packets to the cluster head. Sensor nodes have uniform transmission radius and are not impacted by water flow in a short period of time [22,23].

## 2.2. Underwater Acoustic Channel Model

Path loss caused by an ever-changing feature of acoustic channels is signal frequency dependent. Underwater ambient noise is the main factor that affects underwater acoustic transmission. In the underwater environment, signal attenuation is related to noise interferences, frequency, turbulence, and distance. In a propagation path without obstacles, a signal's attenuation factor at frequency  $f$  is [24]:

$$A(d, f) = d^S a(f)^d \quad (2)$$

where  $d$  and  $S$  represent the distance and spreading factor, respectively.  $S$  is set to one for shallow water cylindrical propagation; 1.5 for practical propagation; and two for deep water spherical propagation. The absorption coefficient happens as a result of a signal's frequency, and the absorption factor  $a(f)$  shown by the Thorp equation is:

$$10 \log a(f) = \frac{0.11 f^2}{1 + f^2} + \frac{44 f^2}{4100 + f} + 2.75 \times 10^{-4} f^2 + 0.003 \quad (3)$$

Turbulence, ships, wind, and thermal noise are the key factors of underwater ambient noise  $N(f)$ , represented as  $N_t(f)$ ,  $N_s(f)$ ,  $N_w(f)$  and  $N_{th}(f)$  [21]. Considering the application scenarios of UASNs, shipping noise and sea surface noise are the main factors affecting

transmission frequency. Therefore, the influence of uncertain noise must be taken into account in the prediction of underwater acoustic transmission quality. Since most ambient noise sources can be described by Gaussian statistics, the following empirical formula gives the estimations of the four noise components [25]:

$$\begin{aligned} 10 \log N_t(f) &= 17 - 30 \log f \\ 10 \log N_s(f) &= 40 + 20 * (s - 0.5) + 26 \log f - 60 \log(f + 0.3) \\ 10 \log N_w(f) &= 50 + 7.5\sqrt{w} + 20 \log f - 40 \log(f + 0.4) \\ 10 \log N_{th}(f) &= -15 + 20 \log f \end{aligned} \quad (4)$$

where  $s$  is the shipping activity factor, and the value of  $s$  is between zero and one;  $w$  is wind speed in m/s. The effective noise level at frequency  $f$  is the sum of the contributions of the above factors:

$$N(f) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f) \quad (5)$$

Based on the noise model of the underwater environment, the average signal-to-noise ratio (SNR) is expressed as follows:

$$\Gamma(d, f) = \frac{P_{tp} / A(d, f)}{N(f)B} \quad (6)$$

where  $B$  denotes bandwidth and  $P_{tp}$  is the power for transmission. The bit error rate between nodes for distance  $d$  is [25]:

$$p_e(d) \approx \frac{1}{4SNR} \quad (7)$$

Therefore, for the successful packet transmission,  $p(d, m)$  represents the probability that  $m$  bits are transmitted between two nodes across the distance  $d$ :

$$p(d, m) = (1 - p_e(d))^m \quad (8)$$

### 2.3. Q-Learning Technique

Machine learning is very popular and applied to many fields. As a subset of machine learning, reinforcement learning obtains specific objectives by interacting with the environment [26]. Q-Learning is one of the reinforcement-learning techniques and it does not need to know prior information of environment [27]. It eventually converges on the optimal strategy by iteratively learning the information gained from environmental feedback. In this context, it is suitable for the dynamic underwater environment.

The node is described with a tuple  $(s, a, r)$ , denoting the state of sensor nodes, action taken by nodes, and direct reward, respectively.

In UASNs, when node  $i$  processes a data packet, the state of it is changed to busy; otherwise,  $s_i$  is idle. The neighbors selected as the next hop are the actions made by the node. The agent performs action  $a_i$  from strategy  $\pi$  before proceeding to state  $s_j$  from state  $s_i$ . Reward is the evaluation of an agent's actions.

$Q_\pi(s_i, a_i)$  is the reward that constitutes the direct reward and discounted future rewards, as defined below:

$$Q^\pi(s_i, a_i) = r_i + \gamma \sum_{s_j \in X} P_{s_i s_j}^{a_i} Q^\pi(s_j, a) \quad (9)$$

The first part  $r_i$  is the direct reward and the second part is the future reward.  $\gamma \in (0, 1)$  is the discount factor of the future reward. The probability of an agent in state  $s_i$  entering state  $s_j$  is given by  $P_{s_i s_j}^{a_i}$ . The optimal value for a state can be derived after the execution of

optimal policy. Furthermore, the Bellman equation can be used to determine at least one optimal strategy  $\pi^*$  [28]. Under the policy, the optimal value is defined as:

$$\begin{aligned} V^*(s) &= \max_a (Q^*(s, a)) \\ Q^*(s_i, a_j) &= r_i + \gamma \sum_{s_j \in X} P_{s_i s_j}^{a_j} V^*(s_j) \end{aligned} \quad (10)$$

$Q^*(s_i, a_j)$  is the expected reward obtained by performing action  $a_j$  in accordance with the optimal policy at state  $s_i$ . Therefore, the optimal action  $a_i^*$  can be described as:

$$a_i^* = \operatorname{argmax}_{a_i \in A(s_i)} Q(s_i, a_i) \quad (11)$$

The design of reward function in the SROA will be introduced in Section 3.4.

### 3. Design of SROA

The details of SROA are described in this section, including the packet format, the mechanism of SROA, and the reward function.

#### 3.1. The SROA Overview

The SROA protocol is proposed to find routing paths for achieving scalability, energy efficiency, and reliability based on on-site architecture in UASNs. The proposed protocol maintains stable with the increasing network size, selecting efficient routes for transmission and avoiding single-point failure through the decentralized mechanism in the network. In addition, a machine learning method is adopted to select optimal routes. The design of the proposed routing protocol is depicted in Figure 2.

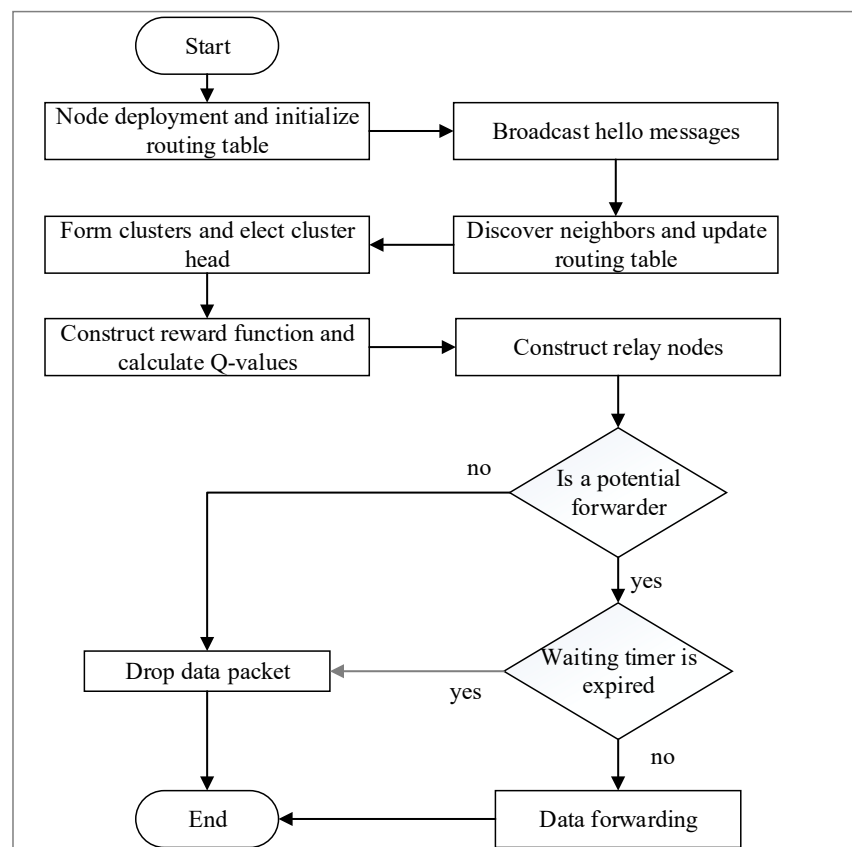


Figure 2. Flowchart of the SROA protocol.



By monitoring channel conditions, sensor nodes deployed in the underwater environment collect information and their local information tables are kept up to date. Through broadcasting messages, all the sensor nodes are grouped into multiple clusters and elect the cluster heads for each cluster. Sensor nodes in the SROA are divided into three types: Cluster heads (CHs), the potential cluster node (PCHs), and ordinary nodes (ONs). The cluster head node is responsible for aggregating data from ONs and transferring the data packet to the sink node through a multi-hop communication; PCH is used as an assistant for CH and similar to the CH in the basic function. The rest of the sensor nodes are the ONs that collect data and forward packets to cluster head within a single hop. Afterwards, Q-Learning is applied to select relay forwarders. The Q-values of qualified neighboring nodes are calculated by the sender node, employing the Q-Learning technique. Moreover, it should be noted that data packets in SROA are transmitted to sink node through multiple-hop communication using the OR strategy. The candidate forwarding set is constructed by taking energy, latency, and connection quality into account. However, it is not suitable for all the nodes participating in the same packet forwarding, which will result in energy consumption and collisions of packets. Therefore, a waiting mechanism is designed to the coordinate candidate set. The obtained Q-value determines the waiting time of candidate nodes. The greater Q-value implies the higher priority for data transmission; hence the waiting time of that node is shorter. When receiving a packet, the sensor node will first check the packet header. If the receiver is not in the candidate list, it will drop the packet. Otherwise, the packet will be kept by the receiver for the waiting time. Furthermore, if the node overhears other nodes transmitting this packet during the waiting period, it will cease forwarding the packet. Table 1 lists the notations in the SROA protocol.

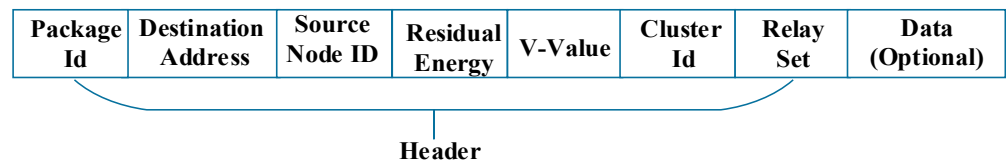
**Table 1.** List of symbols.

Symbol	Meaning
$R_0$	the constant cost
$T_{delay}$	The predefined maximum delay
$C$	The set of clusters
$n_i$	the $i$ th node
$Neighbor_i$	neighbors of $n_i$
$neighbor_{ij}$	The $j$ th neighbor of $n_i$
$E_{n_i}^{res}$	$n_i$ 's remaining energy
$E_{n_i}^{ini}$	$n_i$ 's initial energy
$T_i$	The waiting time of $n_i$ before forwarding
$P_b^{n_j}$	Buffered packets of $n_j$
$E_r, E_s$	Energy for packet reception and transmission
$Cl_j$	The $j$ th cluster
CH, PCH	Cluster head, potential cluster head
CR	Communication Range

### 3.2. Packet Structure of SROA Protocol

Figure 3 shows the data packet structure of SROA. The packet format is used to convey information across nodes and to coordinate the clustering and routing processes. There are two parts in SROA: The header and the data. The header contains packet-related fields and routing-related fields. The first two fields denote the source and destination address. Other header fields are the routing decision-related fields, including source node ID, residual energy, V value, cluster ID, and relay set.

Once an ordinary node receives the data packet, it updates the local neighbor table through the received packet header and then forwards the data packet to its CH or PCH. If the data packet is received by a CH or PCH, it retrieves the packet header and neighbor table for the information updates. If it is in the relay set, the candidate set will be constructed by the calculated the Q-values of the Q-Learning, based on related factors and the packet header is wrapped with the relevant fields. A waiting timer then starts. Otherwise, the current node is not in the relay set and the reception will be dropped.



**Figure 3.** Packet structure of the SROA protocol.

The payload data field is not required. When no payload data is present, data from the upper level will be relayed to the sink node. Otherwise, the data packet serves only to exchange information.

### 3.3. SROA Protocol Description

The SROA protocol includes four phases: Initialization, clustering, relay set construction, and packet transmission.

**(1) Initialization:** In this phase, initializations such as neighbor tables, routing tables, and initial energy of nodes are established. The sensor nodes communicate with their neighboring nodes through the exchange of data packets, and then update their local tables. Each node maintains a local neighbor table that stores neighboring node information and clustering information for routing determinations. By this way, each node may learn about the whole network, not just the information of its own neighbors.

**(2) Clustering:** At this stage, the sensor nodes are grouped into clusters, and each of these clusters has a CH and PCH, respectively. Ordinary sensor nodes simply communicate with the CH. The CH then sends the fused data to the sink node through the multi-hop communication path. The clustering phase enables nodes in the same group to have similar characteristics. In the underwater environment, the expenditure for resizing the network is known to be high, thus clustering makes the SROA adapt to the scaling of different network sizes by clustering and residing tables in each sensor node. Moreover, the energy distribution of nodes in the cluster is more uniform, extending the overall network lifetime.

Before clustering analysis, preliminary exploratory analysis of the sensor nodes is required, the core of which is to determine the number of categories for clustering analysis, which is helpful for the identification of abnormal nodes in the later stage. The silhouette coefficient method has been widely recognized in the evaluation of the clustering effect, and it is a better evaluation method. The silhouette coefficient can evaluate the quality of the clustering model. Its main basis is the degree of cohesion and separation [29]. The contour coefficient is calculated using the following formula:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (12)$$

where  $b(i)$  is the average distance of the nearby clusters, and  $a(i)$  is the average distance for each node in the cluster. The silhouette coefficient is then computed for each of the  $k$  random values. As a result, a  $k$  with a greater coefficient is the better value [30].

The  $k$ -means clustering algorithm is an algorithm that finds  $k$  clusters of a dataset, each cluster described by its centroid. However, the initial seed of  $k$ -means is randomly selected, so the convergence speed of the algorithm is very closely related to the initial value. Therefore, we adopt the  $k$ -means++ algorithm in a three-dimensional underwater network, which can improve the selection of the initial value. For the 3D underwater context, we use  $k$ -means++ with modifications. The silhouette coefficient is used to compute the value of  $k$ , which is required by the method.

The proposed algorithm randomly selects the first centroid, and the subsequent centroids are selected by calculating the distance from other nodes to the previous centroid. Then, the node with the farther distance replaces the randomly selected centroid as the new centroid. Then, the above process is repeated until all  $k$  centroids have been chosen.



At last, the conventional  $k$ -means processes are used to assign each data point to the closest centroid.

After clustering the sensor nodes, the selection of CH and PCH is performed. Similar to initialization phase, the broadcast message is also sent to other nodes containing the clustering information. This raises communication costs, but it can synchronize the state of the network and avoid unnecessary packet transmission among nodes for dynamic topologies by exchanging information.

When selecting CHs, the residual energy and location of the node are taken into account. The nodes that are with more residuals and are closer to other nodes in the cluster are more likely to be CHs. The probability for node  $i$  to be selected as CH is expressed as:

$$P_i = \rho E_i + (1 - \rho) \sum_{j \in N} L_{ij} \quad (13)$$

where  $\rho$  is the coefficient and can be tuned for a specific scenario.  $L_{ij}$  is the distance between the current node  $i$  and other nodes in the cluster. Generally, there is one CH and the majority are ONs for each cluster. The CH collects the sensed data and aggregates data from its members, then the data will be forwarded to the sink through multi-hop communication. Besides, in our proposed protocol, a PCH node is elected to alleviate the burden on the CHs, wherein data packets can be forwarded by either CHs or PCHs. PCH can be considered as a replication or backup for CH, which improves the high availability and avoids single-point failure. Hence, the selection of PCH is almost similar to CH. When the CH crashes and is not able to be connected any more, PCH will take the place of CH and become the new CH. Meanwhile, the cluster is triggered to start a new round of CH elections. After completing the clustering, sensor nodes broadcast messages indicating the cluster information. Usually, the CH consumes more energy than other sensor nodes, resulting in a shorter CH lifespan. To prolong the network lifetime, the proposed protocol relies on the periodic reselection approach, where CHs and PCHs change periodically, namely, when the remaining energy of CHs and PCHs becomes less than a certain threshold, reselection is performed automatically based on the previously described factors. In SROA, we apply the average energy of the cluster member nodes  $E_a$  as the threshold value  $E_t$ . The procedure of clustering is described in Algorithm 1.

**(3) Relay set construction:** At this stage, when a node  $n_i$  receives data packet, it first checks whether itself is a CH or PCH, if not, the packet will be forwarded to the CH in the cluster. If it is true, to improve the packet delivery rate and reduce energy consumption, the sender constructs the relay set  $R_i$ , which are in the transmission range of  $n_i$  and calculates the Q-value of them. If the nodes in  $R_i$  receive data packet, and all the candidate nodes forward the reception without suppression, this will deplete the energy as well as occupy the channel bandwidth. Therefore, the forwarding priority list should be determined and packaged into packet header after constructing the  $R_i$ .

In SROA, the sender computes the Q-values of qualified neighbors through the received packet header and local neighbor table. Afterwards, the Q-value-based waiting time for candidate nodes, namely the forwarding priority, is computed. It is necessary to set the waiting time of each node properly. If the waiting time is set too long, it will result in long delay during transmission. Otherwise, too short a waiting time cannot suppress the low-priority nodes and the packet has already been transmitted before the expiration of the waiting time, leading to packet redundancy. As a result, the sender computes the waiting time for each candidate node using Q-values, local neighbor table, and packet header. The greater the Q-value, the higher the priority of that node, thus the node with the shorter waiting time participating in the forwarding. Based on the calculated Q-value, the waiting time is:

$$T_i = k \cdot \left(1 - \frac{Q_i}{Q_{\max}}\right), \quad k = 2 \cdot \frac{R}{v_d} \quad (14)$$

where parameter  $k$  is equal to the maximal delay, during which candidate nodes hear the packet delivery from other high-priority nodes before forwarding. Taking the worst condition into account,  $k$  is set to  $2 \cdot \frac{R}{v_a}$ , which is twice the propagation delay between the two nodes.  $Q_i$  is the Q-value of  $n_i$ , while  $Q_{\max}$  is the maximum Q-value among the nodes. The waiting time  $T$  is zero when the Q-value of the candidate node is just the maximum. As a result, the end-to-end delay can be reduced. Before data forwarding, the sender node will wrap the collection of candidate forwarders and the calculated suppressing time into the header.

---

**Algorithm 1:** The procedure of Clustering.

---

```

1.  Procedure Clustering(all nodes)
2.  Get all nodes  $N = \{n_1, \dots, n_m\}$  where  $m$  is the number of nodes;
3.  Get all the locations of  $N$ ;
4.  //Calculate the optimal  $k$ 
5.  Calculate the silhouette coefficient for  $N$ ;
6.   $k$  is the highest silhouette score;
7.  Select centroid  $c_1$  randomly where  $c_1 \in N$ ;
8.  //cluster the network, apply  $k$ -means++
9.  For  $j = 1; j \leq k; j++$ 
10. For  $i = 1; i \leq m; i++$ 
11. Calculate distance between  $n_i$  and previously  $c_j$ 
12. New centroid  $c_{i+1} \in N$  is selected with longer distance;
13. End for
14. End for
15. Assign  $n_i \in N$  to the nearest  $c_j \in C$  by  $k$ -means++;
16. //select CH and PCH for each cluster
17. For  $j = 1; j < k; j++$ 
18. Calculate the average energy  $E_a$  for  $Cl_j$ ;
19.   For  $n_i \in Cl_j$  and  $E_i > E_a$ 
20.   Select CH and PCH by distance and residual energy;
21.   Update the cluster status;
22.   End for
23. End Procedure

```

---

**(4) Packet transmission:** When a node receives the packet, it will first check the package header. If the candidate set contains itself, it starts the waiting timer according to the fields in packet header. Then, the data forwarding repeats the above steps until the data packet is received by the sink node. Therefore, a complete routing path has been built. Then, the successive packets from the same source node are sent directly along the calculated routing path. When a transmission failure occurs, Q-Learning will run again and converge to alternative routes. The routing procedure of SROA protocol is described in Algorithm 2.

---

**Algorithm 2:** Routing Process.

---

```

1.  Procedure routing(node  $n_i$ )
2.  Initialize  $V(s)$ ;
3.  Get  $E_{res}^{n_{ij}}, P_b^{n_{ij}}$ , and locations of  $Neighbor_i$ ;
4.  If ( $n_i \neq$  sink node)
5.  If ( $n_i$  is not a CH or PCH)
6.  find the CH(head node) of  $n_i$ ;
7.  transmit data packet to CH;
8.  Return;

```

---

**Algorithm 2:** *Cont.*


---

```

9.   Else
10.    For  $n_{ij}$  in  $Neighbor_i$  do
11.     Compute direct reward  $r_{ij}$ ;
12.     Compute  $Q^*(s_i, a_i)$ ;
13.     Calculate waiting time  $T_{ij}$ ;
14.     Start a timer with the waiting time  $T_{ij}$ ;
15.     While current time < expired time
16.      If the packet is already transmitted
17.       Update local table and drop the packet;
18.      End if
19.     End while
20.     Update the packet loss rate  $P_{s_i s_i}^{a_i}$ ;
21.     Update  $V(s)$  based on  $\max Q^*(s_i, a_i)$ ;
22.     Data forwarding;
23.   End for
24. End Procedure

```

---

**3.4. Design of Reward Function**

The reward function is a critical part of Q-Learning, so we go over the reward function in depth. The SROA adopts three performance indicators in the reward function to assess the action interacting with the environment, including remaining energy, network latency, and link quality, to make the protocol more energy-efficient and reliable. The Q-value calculated represents the quality of routing decisions. When taking action  $a_j$  successfully in the transmission, the reward is denoted by  $R_{s_i s_j}^{a_j}$ , which is defined as follows:

$$R_{s_i s_j}^{a_j} = -R_0 - \alpha_1 * [c(en) + \alpha_2 * (c(delay) - c(pdr))] \quad (15)$$

where  $\alpha_1, \alpha_2 \in (0, 1)$

The reward function takes constant cost, energy cost, delay cost, and link quality cost into account. Due to the occupation of the channel bandwidth during communication,  $R_0$  represents the constant cost. Hence, the constant cost increase with the number of relay hops. If the reward function only contains the constant cost, it will lead to selecting just the shortest path. Nevertheless, the shortest path is not always the best path owing to the imbalanced energy use and transmission reliability. As a result, additional factors, such as remaining energy, network latency, and package delivery, must be addressed. In addition, network latency and package delivery ratio are the indicators of transmission, so a link sensitivity factor denoting  $\alpha_2$  is introduced for balancing energy and to link the quality of the path. When the link sensitivity factor is set to zero, the selected path takes only the energy into account. As a result, the sensitivity factor in the formula is the weight assigned to the link cost.

$c(en)$  denotes the energy-related cost. When packet transmission is successful, it is defined as:

$$c(en) = (1 - \frac{E_{res}^{n_j} - E_r}{E_{ini}}) + (1 - \frac{E_{res}^i - E_s}{E_{ini}}) \quad (16)$$

In Equation (16),  $E_r$  and  $E_s$  represent the energy consumption to receive and transmit packets. The sensor nodes with higher residual energy have lower energy-related costs, thereby balancing energy distribution and increasing network lifespan in UASNs.

$c(delay)$  is a reflection of the congestion in the underwater sensor network. The nodes with many packets in their buffers will have long network latency. It is defined as:

$$c(delay) = 1 - \frac{1}{p_b^{n_j} + 1} \quad (17)$$

where  $P_b^{n_j}$  is the number of buffered packets of the neighboring node. With more packets in a neighbor node's buffer, the waiting time for the packet to be successfully forwarded from that node will be longer, causing data packets to wait in the queue for a longer period of time. As a result,  $c(delay)$  is comparatively greater.

The packet delivery-related cost, denoted by  $c(pdr)$ , represents the transmission quality in UASNs. The SROA calculates the PDR using the acoustic signal attenuation model, which is defined in Section 2.2 and indicated as  $p(d_j, m)$ :

$$c(pdr) = p(d_j, m) \quad (18)$$

The packet delivery ratio is a crucial metric for assessing transmission reliability. The node with the highest delivery ratio is thought to be more trustworthy in packet advancement, hence it is more likely to be chosen as the forwarder.

Since the SROA mainly aims to improve transmission reliability and energy efficiency,  $c(en)$ ,  $c(delay)$ , and  $c(pdr)$  are in the range of (0, 1) by definition, which is enabled to balance  $\alpha_1$  and  $\alpha_2$  in Equation (15) by further tuning the weights for various demands. They are, by definition, in the range (0, 1), allowing us to balance  $\alpha_1$  and  $\alpha_2$  in Equation (15) by fine-tuning the weights for various demands.

However, the failure of the transmission also occurs in the real environment. If the packet retransmission approaches the limit and the receiver still does not receive the packet, significant energy and time will be consumed. Retransmission of data packets results in extra delay and energy consumption, raising the cost of unsuccessful transmission. The failure reward function is described as:

$$R_{s_i s_i}^{a_j} = -R_0 - \alpha_1 * [c(en) + \alpha_2 * c(delay)] \quad (19)$$

where

$$\begin{aligned} \alpha_1, \alpha_2 &\in (0, 1) \\ c(en) &= c_i(en) = 1 - \frac{E_{res}^i - E_s * N_{max}}{E_{ini}} \\ c(delay) &= 1 - \frac{1}{p_b^{n_j} + 1} \end{aligned}$$

According to the definition of the reward function, the direct reward for successful and failed transmissions is defined as follows:

$$r_i(a_j) = P_{s_i s_j}^{a_j} R_{s_i s_j}^{a_j} + P_{s_i s_i}^{a_j} R_{s_i s_i}^{a_j} \quad (20)$$

In order to estimate the acoustic channel state and state transition probability, each node records recent packet transmissions locally. The lost packets are indicated as  $\lambda$  and  $n$  is the total number of packet transfers. Therefore, the loss rate  $P_{s_i s_i}^{a_j}$  and the successful transmission rate  $P_{s_i s_j}^{a_j}$  are stated as follows:

$$\begin{aligned} P_{s_i s_i}^{a_j} &= \frac{\lambda}{n} \\ P_{s_i s_j}^{a_j} &= 1 - \frac{\lambda}{n} \end{aligned} \quad (21)$$

Therefore, substituting  $P_{s_i s_i}^{a_j}$  and  $P_{s_i s_j}^{a_j}$  into the reward function, the reward function can be updated:

$$Q(s_i, a_j) = r_i(a_j) + \gamma((1 - \frac{\lambda}{n})Q^*(s_j) + (\frac{\lambda}{n}) * Q^*(s_i)) \quad (22)$$

The Q-value is related to the actions taken in the underwater environment and information exchange of the network. Initially, the Q-value of each node is set to zero, except the sink node. When a node delivers a packet, it updates its Q-value based on the information from the forwarder. In SROA, since the Q-value of sensor node is less than zero after packet forwarding, the Q-value of sink node is fixed at zero to ensure that the protocol converges.

#### 4. Simulations and Analysis

In this section, our proposed protocol SROA is evaluated for the performance based on Matlab R2021a and NS 3.26 [31] from three aspects. First, simulation settings are introduced before evaluations. Afterwards, we assess the impact of various parameters on the SROA. The performance of on-site architecture is also evaluated with the same protocols. Finally, we evaluate the performance of SROA and compare it with the other three routing protocols for different metrics.

##### 4.1. Simulation Setting

Sensor nodes are randomly deployed in a  $5000\text{ m} \times 5000\text{ m} \times 5000\text{ m}$  three-dimensional space in our simulations. Any sensor node is the same in functional features, and each node near to the seafloor can generate the data packets independently as a source node. The sink node is deployed on the seabed, which is considered to be difficult to reposition once deployed. For analysis, we select a source node from the seabed. The propagation loss model for underwater acoustic channels is Thorp [32]. The acoustic transmission speed is set at  $v_0 = 1500\text{ m/s}$ , and the network size ranges between 100 and 500. Table 2 displays the simulation parameters [9].

**Table 2.** Simulation settings.

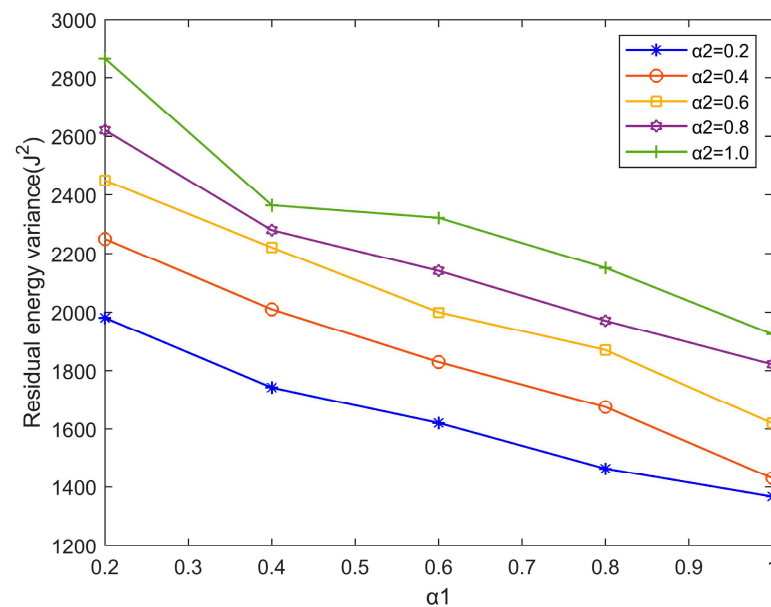
Parameter	Value
Network size	100 to 500
Transmission power	10 W
Receiving power	3 W
Transmission rate	1 kbps
Data packet size	50 Bytes
Simulation rounds	200
Communication range	1000 m, 1500 m
Idle power	30 mW
Initial energy	1000 J

To evaluate the SROA, we employ the Carrier Sense Multiple Access (CSMA) as the underlying MAC protocol. Specifically, when the channel is not occupied, the forwarding node is able to broadcast the data packet; otherwise, it backs off and discards the packet after five times of backing off [22]. We mainly evaluate the SROA protocol in several quantitative metrics and scenarios against the two different parameters: Network size and transmission range. Network size is different for various demands and environments in reality, hence the routing protocol is significant to have the ability which scales up the network with stability. Hence, the test for network scalability, irrespective of variation in the number of nodes, is essential. However, the transmission range also impacts the metrics of the protocol. The larger the transmission range of the sensor nodes, the more energy is needed for communicating. Based on this, two different transmission ranges are tested to evaluate the effects on the performance of the network. Furthermore, the performance of SROA is evaluated using the following metrics: Average End-to-End Delay indicates the network latency, namely the average time consumption for forwarding a data packet to the sink node, including the waiting time, packet propagation time, and processing time; the Packet Delivery Ratio represents the ratio of delivered data packets; Energy Consumption is defined as the total energy consumed by all of the nodes for transmission, which includes the packet transmission and reception consumption [33]; Average Hop Count of Delivered Packets means the average number of hops from the source to the destination on the routing path.

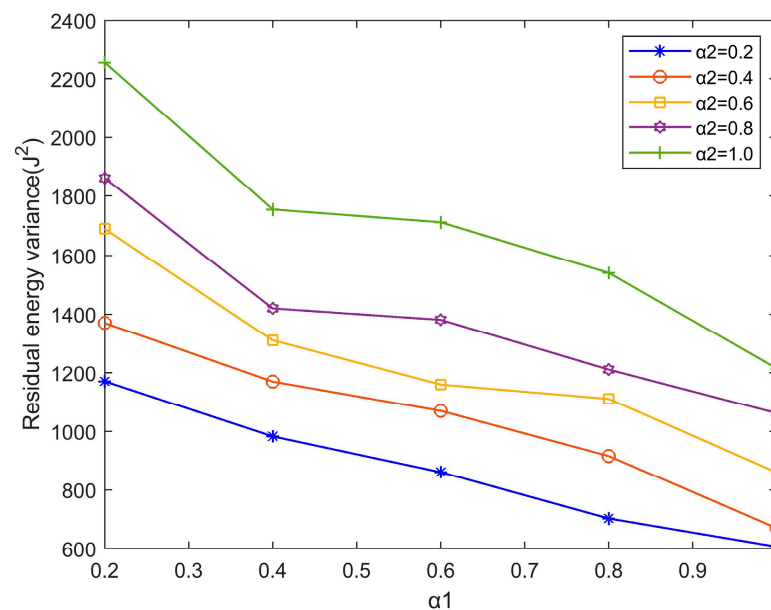
## 4.2. Numerical Results

### 4.2.1. Parameter Analysis

The simulation experiments of different coefficients are conducted in the network with 300 nodes under different communication ranges of 1000 m and 1500 m. We evaluate the performance metrics of residual energy variance and average end-to-end delay in the network. The effect of  $\alpha_1$  (total cost weight) and  $\alpha_2$  (delay and link quality weight) on the residual energy variance with two different CRs are shown in Figures 4 and 5, respectively. The reward function is influenced by the coefficients, with  $\alpha_1$  varying between 0.2 and 1.0 and  $\alpha_2$  varying between 0.2 and 1.0.



**Figure 4.** Residual energy variance of  $\alpha_1$  and  $\alpha_2$  (CR = 1000 m).



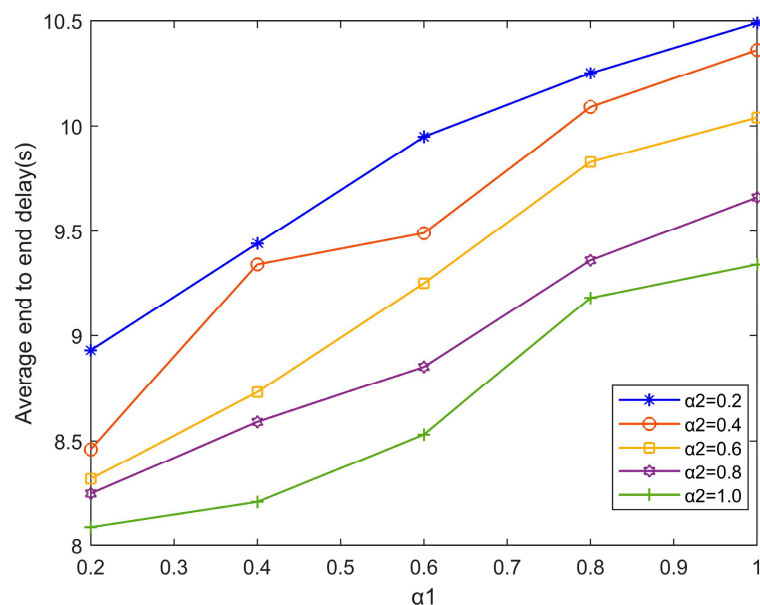
**Figure 5.** Residual energy variance of  $\alpha_1$  and  $\alpha_2$  (CR = 1500 m).

Figures 4 and 5 shows that the residual energy variance decreases while expanding the CR from 1000 m to 1500 m. This is because the CR increasing makes fewer nodes participate in data packet forwarding, resulting in less energy consumption. It is also evidently



observed that the residual energy variance increases with the value of  $\alpha_2$  increasing, because link quality and end-to-end delay account for a greater portion in selecting forwarding nodes. Similarly, taking merely the globally optimum path into account cannot assure the uniform distribution of remaining energy. Therefore, a delay-limited routing approach cannot provide uniform distributions of energy. Furthermore, the residual energy variance diminishes as  $\alpha_1$  grows. In Figure 5, for example, when  $\alpha_2$  is 0.2 and communication range is 1500 m, the residual energy variance for  $\alpha_1 = 0.8$  is 40% smaller than at  $\alpha_1 = 0.2$ . The reason is that energy has a stronger impact on the reward function, hence influencing the routing decisions. It is apparent that the greater the value of  $\alpha_1$ , the more probable it is that a node with more remaining energy will be selected as a forwarder. This is because the energy of the sensor nodes is well-distributed and the network lifetime may be prolonged.

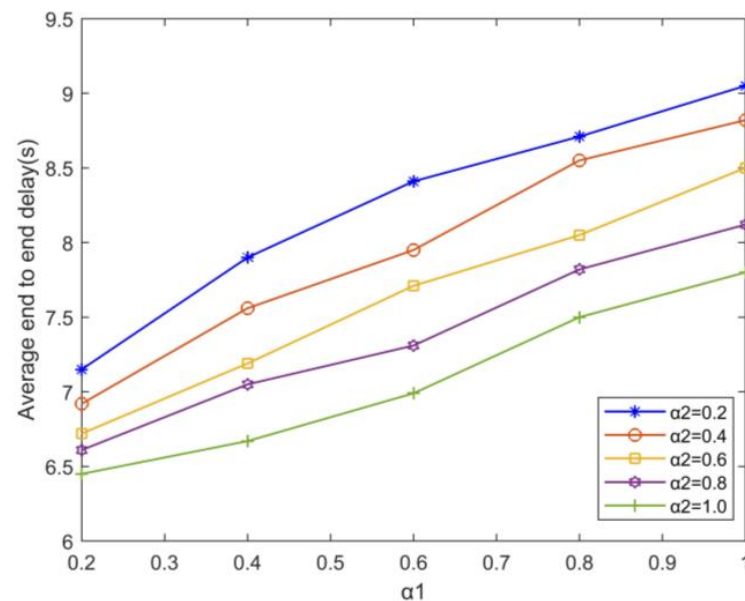
Correspondingly, Figures 6 and 7 depict the influence of  $\alpha_1$  and  $\alpha_2$  of SROA on the average end-to-end delay in a 300 node network with CRs of 1000 m and 1500 m. Comparing the two figures, we can find that the average end-to-end delay decreases with the expanding of CR. The reason is that the greater coverage of transmission makes less sensor nodes involved in the data forwarding, hence packets are routed by a shorter path which reduces the average end-to-end delay. It can be witnessed that an increase of  $\alpha_2$  promotes the protocol to choose the node that best balances the factors of residual energy and link quality as the forwarder. As a result, the protocol can converge on the path with the fewest hops, which not only improves energy efficiency but also minimizes end-to-end delay. It is also shown from each individual figure that increasing  $\alpha_1$  results in the larger end-to-end delay when  $\alpha_2$  is set. This is due to the even energy distribution, indicating that the SROA cannot select the shortest routing path. Specifically, the CR is 1500 m in Figure 7. When  $\alpha_2$  is set to 0.8 and  $\alpha_1$  is set to 0.2, the average delay is 6.61 s, which is approximately 24% less than that of  $\alpha_2 = 0.2$  and  $\alpha_1 = 0.8$ .



**Figure 6.** Average end-to-end delay of  $\alpha_1$  and  $\alpha_2$  (CR = 1000 m).

By comparing Figures 4–7, it can be seen that the increasing of the CR of sensor nodes decreases both the residual energy variance and the average end-to-end delay. While increasing  $\alpha_2$  reduces the end-to-end latency, it also increases the residual energy variance, resulting in a shorter network lifespan. As a result, it can be summarized that a greater value of  $\alpha_1$  makes the distributions of energy more uniform, nonetheless, this increases the average network latency. A higher value of  $\alpha_2$  indicates less end-to-end delay and greater residual energy variation. The values of  $\alpha_1$  and  $\alpha_2$  are weighted according to the scenario,

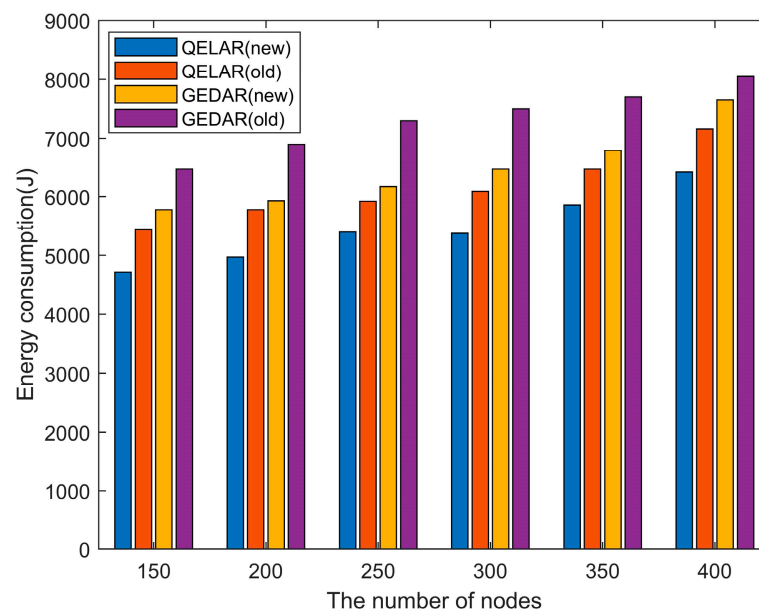
and different values are used to meet varied network needs. As a result, for the successive assessments,  $\alpha_1$  and  $\alpha_2$  are set to 0.5.



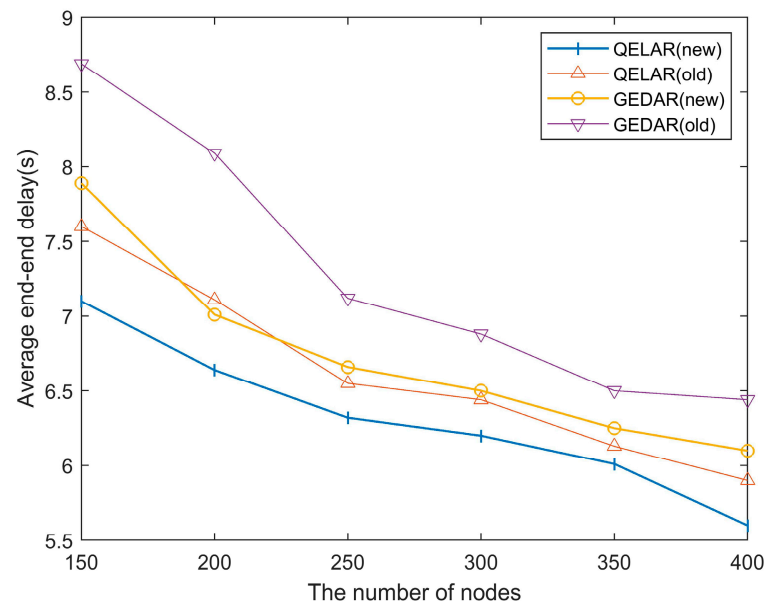
**Figure 7.** Average end-to-end delay of  $\alpha_1$  and  $\alpha_2$  (CR = 1500 m).

#### 4.2.2. Architecture Analysis

To assess the performance of on-site architecture, we apply the architecture to QELAR and GEDAR, respectively, comparing the metrics of end-to-end delay and energy consumption. Figures 8 and 9 show the result of comparison on the total energy consumption and average end-to-end delay for different architectures with the CR of 1000 m, varying network size from 150 to 400. Clearly, it can be seen from the results that the performances of the same protocols with on-site architecture are better than that of the original architecture in terms of total energy consumption and average end-to-end delay.



**Figure 8.** Comparison of total energy consumption for different architectures.



**Figure 9.** Comparison of average end-to-end delay for different architectures.

We can observe from Figure 8 that QELAR and GEDAR with the on-site architecture consumes less energy than the original architecture, which reduces the total energy consumption by 25%. One of the reasons is that the data center deployed near the data source shortens the transmission distance thus reducing energy consumption. Besides, with the same architecture, the energy consumption of the QL-based QELAR protocol is lower. Due to its constant cost, it tends to choose the shortest path to forward data packets. In addition, it can also be seen that the Q-Learning-based QELAR protocol consumes less energy with the same architecture, because it usually chooses the more optimal path to forward data packets in a global view.

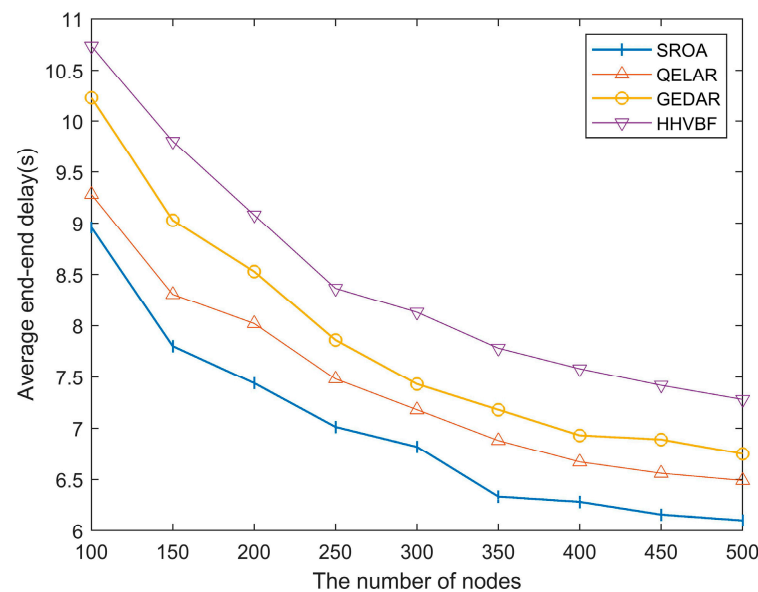
The average end-to-end delay of the protocols with different architectures is shown in Figure 9. Corresponding to Figure 8, the end-to-end delay with the on-site architecture is significantly reduced. The reason is similar to Figure 8, with the on-site architecture, data packets are forwarded with fewer hops and there are fewer nodes participating in the route, thereby reducing the average end-to-end delay.

As a result of the on-site deployment, the protocols with new architecture show apparent improvements in terms of total energy consumption and average end-to-end delay. Furthermore, the Q-Learning-based protocols also outperform classical protocols.

#### 4.2.3. Average End-to-End Delay

The average end-to-end delay for different protocols with the same CR of 1500 m is shown in Figure 10. We can observe that, as the network size increases, the average end-to-end delay decreases. In general, with the increasing of sensor nodes deployed underwater, it means the deployment of the network is denser and all four protocols can route packets along shorter paths from the source node to the sink node. Therefore, when the network size is 500, the delay of the network is minimal. Furthermore, the SROA protocol appears to have a lower average end-to-end delay than other protocols. The average delay of SROA is 7.4 s when there are 200 nodes in the network and the CR is set to 1500 m, whereas QELAR, GEDAR, and HHVBF are 7.88 s, 8.56 s, and 9.1 s, respectively. This is due to the fact that SROA uses a Q-value-based waiting mechanism to coordinate relay nodes, thus reducing retransmissions and collisions. Moreover, the routing paths to the sink node are shorter because of the on-site architecture where the data center is deployed close to the data source. The average delay of HHVBF is the highest among the four protocols. Owing to the hidden terminal problem, the happening of collisions results in the increasing of the average end-to-end delay. The average latency for GEDAR is the second as it utilizes

opportunistic routing to enhance PDR but transferring void sensor nodes to other areas still takes time, bringing about extra time consumption.



**Figure 10.** Comparison of average end-to-end delay for the four protocols.

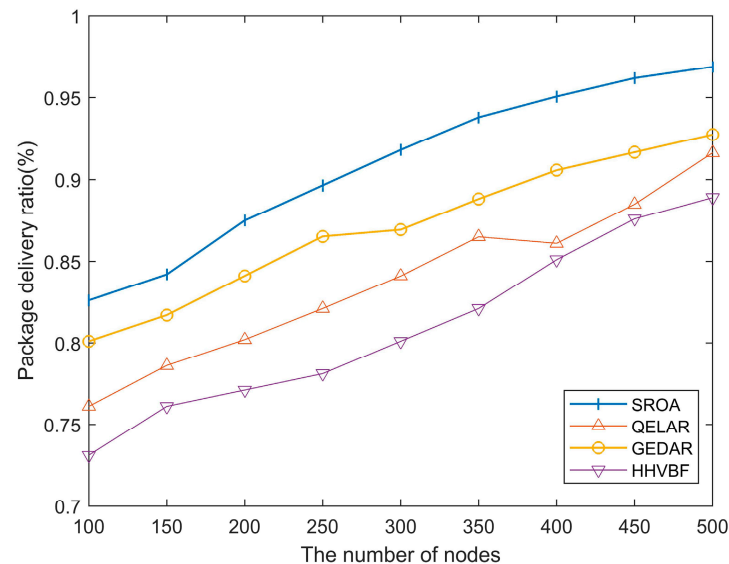
#### 4.2.4. Packet Delivery Rate

Figure 11 compares the packet delivery rate of SROA to that of QELAR, HHVBF, and GEDAR. It can be observed that the PDR of all four protocols rises as the network size increases. Since relay nodes have more available neighboring nodes for relaying data packets, the packet delivery rate improves. We can also see that SROA has a greater PDR than the other methods. For example, SROA's PDR reaches 96.6% when the network size is 500, which is greater than that of GEDAR, QELAR, and HHVBF. One of the reasons is that the reward function of the SROA protocol considers not only link quality while determining routing decisions, but also related factors such as residual energy and end-to-end latency, ensuring high PDR globally. However, grouping the sensor nodes into several clusters and the replication mechanism of cluster head makes the transmission more reliable. Moreover, based on the on-site architecture, the data packet travels fewer hops to the sink node, improving the packet delivery ratio. For the QELAR, since the sender may choose a path with fewer hops to improve the packet delivery rate, the PDR of it increases. As GEDAR considers the expected packet advance, more than one node participates in the packet forwarding. Therefore, in the figure, GEDAR's PDR is substantially larger. The PDR of HHVBF is the lowest, as it does not take the packet error rate into account, resulting in unnecessary retransmissions and low PDR.

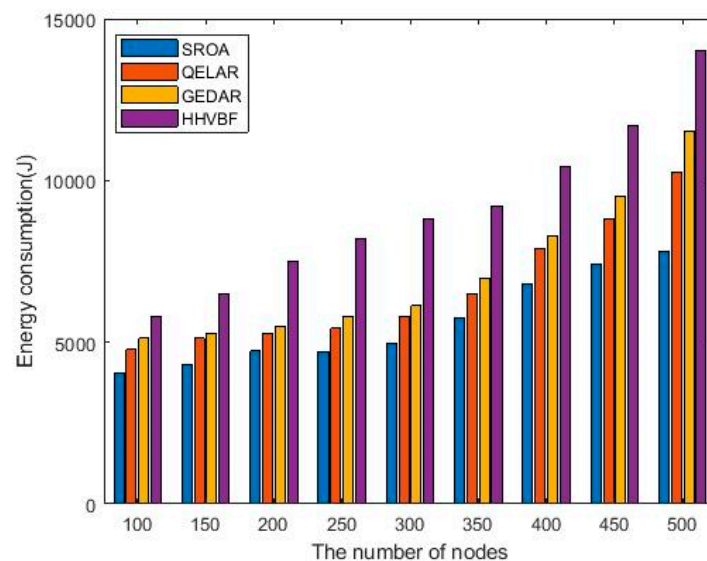
#### 4.2.5. Energy Consumption

Energy is very precious and important in the underwater environment; each routing protocol should consider the efficiency of energy consumption seriously. The comparison of energy consumption with a network size ranging from 100 to 500 is shown in Figure 12. From the figure, we can find that the total energy consumption of these protocols increases as the network size grows. Generally, it also can be seen clearly that the energy consumed by the SROA is less than other protocols. Specifically, when network size is 500, the SROA consumes 23.8%, 32.1%, and 44.2% less energy than QELAR, GEDAR, and HHVBF. Since the SROA is a cluster-based protocol, the energy distribution of nodes in each cluster is more uniform which extends the network lifetime. Moreover, the SROA uses waiting mechanism based on opportunistic routing to forward packets which reduces the retransmissions, thus SROA consumes less energy in comparison with QELAR and GEDAR. Besides, among the

four protocols, the energy consumption of HHVBF is the maximum and grows faster with the increase of nodes. The HHVBF consumes more energy, because neighbors of the source build their own routing pipes and more nodes participating in data forwarding, which results in significant energy consumption.



**Figure 11.** Comparison of the packet delivery ratio for the four protocols.

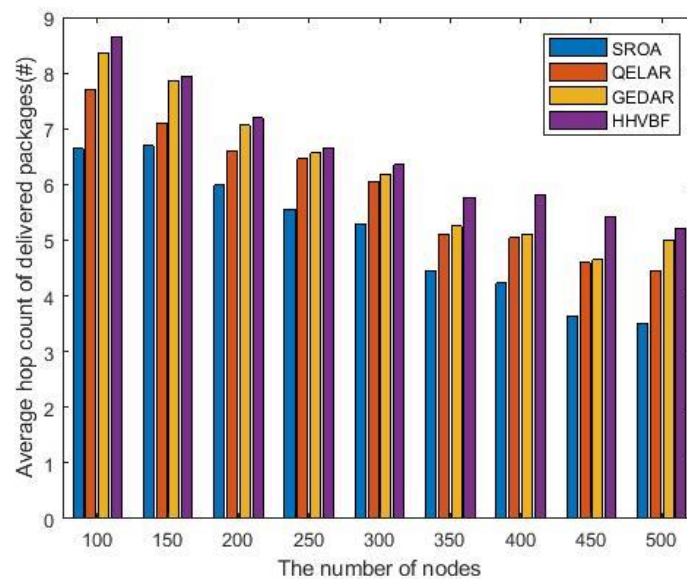


**Figure 12.** Comparison of energy consumption for the four protocols.

#### 4.2.6. Average Hop Count

Figure 13 depicts the average hop count of packets delivered from the source to the sink node. In some extreme cases, sensor nodes may be unable to entirely cover the shortest routing path; consequently, the average hop count and packet delivery ratio must be balanced. Figure 13 illustrates that, as the network size rises, so does the average hop count, and the results are consistent with all four protocols. The reason for this is that as node density increases, packets will be routed along optimum routing paths, thus fewer nodes participating in routing. In particular, when the network scale is 400, the average hop count is 4.23, whereas it is 4.95, 5.16, and 5.81 for QELAR, GEDAR, and HHVBF, respectively. Among these protocols, the SROA and QELAR, which are machine-learning-based protocols, takes less hop counts than others because they use the intelligent

algorithms to choose the best forwarders. In addition, a global view of the network architecture is enabled by the Q-Learning technique. This not only reduces the average hop-count but also adapts to various network sizes. Furthermore, the SROA outperforms the QELAR because the sink node is deployed close to the source node with the on-site architecture, thus shortening the forwarding path. HHVBF is restricted to the sensor nodes within the pipe radius, making it inflexible in terms of finding the routes with less hops to the destination. For the GEDAR, it applies the greedy forwarding method for advancing packets and does not take hop count into consideration, causing GEDAR to choose routing pathways with larger hop counts than SROA.



**Figure 13.** Comparison of the average hop count for the four protocols.

In UASNs, it is expensive to increase the number of nodes and this makes a new deployment. Therefore, the scalability of network is very important. Considering the above experiments, the performance metric of network scalability can also be observed. Figures 8–13 demonstrate the comparison between SROA, QELAR, GEDAR and HHVBF in terms of average end-to-end delay, PDR, energy consumption, and average hop count with different scales of nodes varying from 100 to 500. The simulation results show that the performance metrics of the proposed SROA are excellent among the four protocols and remain stable irrespective of the increasing network size. By adding new nodes in the network, the four evaluation indexes are improved.

## 5. Conclusions

In this paper, a scalable and energy-efficient routing protocol based on the on-site architecture for UASN is proposed. The SROA groups the sensor nodes into clusters, which enables better resource allocation and easily adapts to the changes in the network scale. By deploying the data center close to the data source, the on-site architecture can shorten the routing path and greatly reduce transmission delay and energy consumption. The SROA follows a decentralized mechanism where the failure of a single node does not interrupt the connectivity in the network. Moreover, a reward function for Q-Learning is applied for routing decisions, which trades off multiple factors of the network. Considering both the instant rewards and the discounted long-term rewards, SROA is more likely to select the optimal candidate forwarders globally. Furthermore, in order to coordinate the forwarding among the candidate nodes, the SROA designs the waiting mechanism, which is developed from opportunistic routing. Different from the traditional OR, this mechanism picks a group of qualified forwarders and sets a waiting time based on the computed Q-values for each candidate node. The simulation results show that the on-site architecture enables



QELAR and GEDAR with new architecture that outperforms traditional architecture in terms of the energy consumption and end-to-end delay obviously. In addition, SROA performs better than other routing protocols (QELAR, GEDAR, HHVBF) when considering performance metrics, such as energy consumption, end-to-end delay, PDR and average hop count of delivered packages. For future work, we will try to deploy the proposed SROA in a real UWSN hardware platform since it is only evaluated in simulation software at current. Additionally, a multi-sink and AUV-aided architecture will be considered to coordinate packet transmission, aiming to improve the packet delivery ratio, avoid routing holes, and reduce end-to-end delay.

**Author Contributions:** Conceptualization, R.Z. and Q.Y.; investigation X.H. (Xiwen Huang); Resources, X.H. (Xiangdang Huang); writing, R.Z. and Q.Y.; validation, D.L.; methodology, R.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 61862020 and Grant 62162021; in part by the Key Project of Hainan Province under Grant ZDYF2020199; in part by Scientific Research Set Up Fund of Hainan University under Grant KYQD(ZR)1877; in part by the Scientific Research Project of Hainan Province under Grant Hnky2022-4.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors would be grateful to the reviewers for their comments to improve the quality of this paper and would also like to thank the editors for their help with this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Khisa, S.; Moh, S. Survey on Recent Advancements in Energy-Efficient Routing Protocols for Underwater Wireless Sensor Networks. *IEEE Access* **2021**, *9*, 55045–55062. [\[CrossRef\]](#)
2. Wei, X.H.; Guo, H.; Wang, X.W.; Wang, X.N.; Qiu, M.K. Reliable Data Collection Techniques in Underwater Wireless Sensor Networks: A Survey. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 404–431. [\[CrossRef\]](#)
3. Qiu, T.; Zhao, Z.; Zhang, T.; Chen, C.; Chen, C.L.P. Underwater Internet of Things in Smart Ocean: System Architecture and Open Issues. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4297–4307. [\[CrossRef\]](#)
4. Jin, Z.; Zhao, Q.; Su, Y. RCAR: A Reinforcement-Learning-Based Routing Protocol for Congestion-Avoided Underwater Acoustic Sensor Networks. *IEEE Sens. J.* **2019**, *19*, 10881–10891. [\[CrossRef\]](#)
5. Xiao, X.; Huang, H.; Wang, W. Underwater Wireless Sensor Networks: An Energy-Efficient Clustering Routing Protocol Based on Data Fusion and Genetic Algorithms. *Appl. Sci.* **2021**, *11*, 312. [\[CrossRef\]](#)
6. Alfouzan, F.A. Energy-Efficient Collision Avoidance MAC Protocols for Underwater Sensor Networks: Survey and Challenges. *J. Mar. Sci. Eng.* **2021**, *9*, 741. [\[CrossRef\]](#)
7. Chen, Y.G.; Zhu, J.Y.; Wan, L.; Fang, X.; Tong, F.; Xu, X.M. Routing failure prediction and repairing for AUV-assisted underwater acoustic sensor networks in uncertain ocean environments. *Appl. Acoust.* **2022**, *186*, 108479. [\[CrossRef\]](#)
8. Hindu, S.K.; Hyder, W.; Luque-Nieto, M.A.; Poncela, J.; Otero, P. Self-Organizing and Scalable Routing Protocol (SOSRP) for Underwater Acoustic Sensor Networks. *Sensors* **2019**, *19*, 3130. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Nicolaou, N.; See, A.; Xie, P.; Cui, J.-H.; Maggiorini, D. Improving the robustness of location-based routing for underwater sensor networks. In Proceedings of the OCEANS 2007-Europe, Aberdeen, UK, 18–21 June 2007.
10. Anand, M.; Antonidoss, A.; Balamanigandan, R.; Rahmath Nisha, S.; Gurunathan, K.; Bharathiraja, N. Resourceful Routing Algorithm for Mobile Ad-Hoc Network to Enhance Energy Utilization. *Wirel. Pers. Commun.* **2021**. [\[CrossRef\]](#)
11. Hu, T.S.; Fei, Y.S. QELAR: A Machine-Learning-Based Adaptive Routing Protocol for Energy-Efficient and Lifetime-Extended Underwater Sensor Networks. *IEEE Trans. Mob. Comput.* **2010**, *9*, 796–809.
12. Hao, K.; Shen, H.F.; Liu, Y.L.; Wang, B.B.; Du, X.J. Integrating Localization and Energy-Awareness: A Novel Geographic Routing Protocol for Underwater Wireless Sensor Networks. *Mob. Netw. Appl.* **2018**, *23*, 1427–1435. [\[CrossRef\]](#)
13. Ge, L.; Jiang, S. An Efficient Opportunistic Routing Based on Prediction for Nautical Wireless Ad Hoc Networks. *J. Mar. Sci. Eng.* **2022**, *10*, 789. [\[CrossRef\]](#)
14. Coutinho, R.W.L.; Boukerche, A.; Vieira, L.F.M.; Loureiro, A.A.F. Geographic and Opportunistic Routing for Underwater Sensor Networks. *IEEE Trans. Comput.* **2016**, *65*, 548–561. [\[CrossRef\]](#)

15. Wang, T.; Zhao, D.; Cai, S.; Jia, W.; Liu, A. Bidirectional Prediction-Based Underwater Data Collection Protocol for End-Edge-Cloud Orchestrated System. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4791–4799. [[CrossRef](#)]
16. Tilak, S.; Abu-Ghazaleh, N.B.; Heinzelman, W. A taxonomy of wireless micro-sensor network models. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **2002**, *6*, 28–36. [[CrossRef](#)]
17. Cutler, B.; Fowers, S.; Kramer, J.; Peterson, E.; Wang, D.L. Dunking the Data Center. *IEEE Spectrum* **2017**, *54*, 26–31. [[CrossRef](#)]
18. Jin, Z.; Duan, C.; Yang, Q.; Su, Y. Q-learning-Based Opportunistic Routing with an on-site architecture in UASNs. *Ad Hoc Netw.* **2021**, *119*, 102553. [[CrossRef](#)]
19. Bharathiraja, N.; Padmaja, P.; Rajeshwari, S.B.; Kallimani, J.S.; Buttar, A.M.; Lingaiah, T.B. Elite Oppositional Farmland Fertility Optimization Based Node Localization Technique for Wireless Networks. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 5290028. [[CrossRef](#)]
20. Teymorian, A.Y.; Cheng, W.; Ma, L.R.; Cheng, X.Z.; Lu, X.C.; Lu, Z.X. 3D Underwater Sensor Network Localization. *IEEE Trans. Mob. Comput.* **2009**, *8*, 1610–1621. [[CrossRef](#)]
21. Chen, K.; Ma, M.; Cheng, E.; Yuan, F.; Su, W. A Survey on MAC Protocols for Underwater Wireless Sensor Networks. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1433–1447. [[CrossRef](#)]
22. Zhang, J.; Cai, M.; Han, G.; Qian, Y.; Shu, L. Cellular Clustering-Based Interference-Aware Data Transmission Protocol for Underwater Acoustic Sensor Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 3217–3230. [[CrossRef](#)]
23. Song, Y. Underwater Acoustic Sensor Networks With Cost Efficiency for Internet of Underwater Things. *IEEE Trans. Ind. Electron.* **2021**, *68*, 1707–1716. [[CrossRef](#)]
24. Liu, J.; Wang, Z.H.; Cui, J.H.; Zhou, S.L.; Yang, B. A Joint Time Synchronization and Localization Design for Mobile Underwater Sensor Networks. *IEEE Trans. Mob. Comput.* **2016**, *15*, 530–543. [[CrossRef](#)]
25. Coutinho, R.W.L.; Boukerche, A.; Loureiro, A.A.F. Modeling power control and anypath routing in underwater wireless sensor networks. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 15–18 April 2018; pp. 1–6.
26. Li, Y. Reinforcement learning in practice: Opportunities and challenges. *arXiv* **2022**, arXiv:2202.11296.
27. Naeem, M.; Rizvi, S.T.H.; Coronato, A. A Gentle Introduction to Reinforcement Learning and its Application in Different Fields. *IEEE Access* **2020**, *8*, 209320–209344. [[CrossRef](#)]
28. Su, Y.S.; Fan, R.; Fu, X.M.; Jin, Z.G. DQELR: An Adaptive Deep Q-Network-Based Energy- and Latency-Aware Routing Protocol Design for Underwater Acoustic Sensor Networks. *IEEE Access* **2019**, *7*, 9091–9104. [[CrossRef](#)]
29. Le, T.K.; Le, V.S.; Duc, D.D.; Ngoc, T.B.; Phuong, T.N.T. iK-means: An improvement of the iterative *k*-means partitioning algorithm. In Proceedings of the 12th International Conference on Knowledge and Systems Engineering (KSE), Can Tho City, Vietnam, 12–14 November 2020; pp. 300–305.
30. Alsaman, L.; Alotaibi, E. A Balanced Routing Protocol Based on Machine Learning for Underwater Sensor Networks. *IEEE Access* **2021**, *9*, 152082–152097. [[CrossRef](#)]
31. The Network Simulator-ns-3. Available online: <http://www.nsnam.org> (accessed on 10 January 2020).
32. Gao, C.X.; Hu, W.W.; Chen, K.Y. Research on Multi-AUVs Data Acquisition System of Underwater Acoustic Communication Network. *Sensors* **2022**, *22*, 5090. [[CrossRef](#)]
33. Kumar, P.; Chaturvedi, A. Fuzzy-interval based probabilistic query generation models and fusion strategy for energy efficient wireless sensor networks. *Comput. Commun.* **2018**, *117*, 46–57.