

Article

Underwater Wireless Sensor Networks: An Energy-Efficient Clustering Routing Protocol Based on Data Fusion and Genetic Algorithms

Xingxing Xiao ^{1,2,3} , Haining Huang ^{1,2,3,*} and Wei Wang ^{1,2}

¹ Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China; xiaoxingxing15@mails.ucas.edu.cn (X.X.); eway20@sina.cn (W.W.)

² Key Laboratory of Science and Technology on Advanced Underwater Acoustic Signal Processing, Chinese Academy of Sciences, Beijing 100190, China

³ School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: hhn@mail.ioa.ac.cn

Abstract: Due to the limited battery energy of underwater wireless sensor nodes and the difficulty in replacing or recharging the battery underwater, it is of great significance to improve the energy efficiency of underwater wireless sensor networks (UWSNs). We propose a novel energy-efficient clustering routing protocol based on data fusion and genetic algorithms (GAs) for UWSNs. In the clustering routing protocol, the cluster head node (CHN) gathers the data from cluster member nodes (CMNs), aggregates the data through an improved back propagation neural network (BPNN), and transmits the aggregated data to a sink node (SN) through a multi-hop scheme. The effective multi-hop transmission path between the CHN and the SN is determined through the enhanced GA, thereby improving transmission efficiency and reducing energy consumption. This paper presents the GA based on a specific encoding scheme, a particular crossover operation, and an enhanced mutation operation. Additionally, the BPNN employed for data fusion is improved by adopting an optimized momentum method, which can reduce energy consumption through the elimination of data redundancy and the decrease of the amount of transferred data. Moreover, we introduce an optimized CHN selecting scheme considering residual energy and positions of nodes. The experiments demonstrate that our proposed protocol outperforms its competitors in terms of the energy expenditure, the network lifespan, and the packet loss rate.

Keywords: data fusion; underwater wireless sensor network; back propagation neural network; clustering routing protocol; genetic algorithm; network lifespan



Citation: 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. <https://doi.org/10.3390/app11010312>

Received: 29 November 2020

Accepted: 28 December 2020

Published: 30 December 2020

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



Copyright: © 2020 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

Underwater wireless sensor networks (UWSNs) consist of many underwater wireless sensor nodes distributed within the marine environment, which support a wide variety of applications such as surveillance, navigation, data acquisition, resource exploration, and disaster prevention [1–3]. Each sensor node of UWSNs is equipped with an acoustic modem because it uses acoustic signals to communicate with each other [4]. These nodes are capable of forming a network without any infrastructure. The responsibility of the sensor nodes is to monitor the underwater environment such as the temperature, and send the collected data to a sink node (SN) through a single hop or multiple hops [5]. The SN, located on the sea surface, has the ability to receive the data from underwater sensor nodes through acoustic signals and send the received data to terrestrial network devices through radio signals [6]. In the underwater environment, the radio signals face the absorption problem and attenuate quickly [7]. Hence, they are not suitable for long-distance underwater communications. The sound wave is adopted during underwater

communications because it is less affected by attenuation, scattering, as well as absorption loss [6]. The acoustic signals propagate slowly, which causes the high propagation delay [8]. What is more, there are other drawbacks of the underwater acoustic channel, such as low bandwidth, as well as high error rate [9]. Therefore, it consumes lots of energy to successfully transmit data packets in UWSNs and keep the good performance of UWSNs. Furthermore, the sensor nodes have restrained energy, and it is not easy to recharge or redistribute them [10]. Therefore, the energy consumption and the network lifetime are major concerns in UWSNs [11]. There are some energy-efficient routing approaches in terrestrial wireless sensor networks (TWSNs), which could not be directly adopted by UWSNs because UWSNs employ acoustic signals while TWSNs use radio signals to send data [12]. Moreover, two-dimensional network models are usually employed in TWSNs, whereas UWSNs often use three-dimensional network models, which is very challenging for researchers [13]. As a result, how to choose effective transmission paths for complicated three-dimensional UWSNs becomes crucial.

A number of works have shown that clustering routing protocols are effective and efficient in finding optimal routing paths for data transmission, which can save energy and extend the network lifespan [14,15]. The clustering routing protocols divide the whole network into lots of clusters. Every cluster is composed of a cluster head node (CHN) and some cluster member nodes (CMNs). After completion of the cluster formation, the CHN allocates channel resources for the CMNs in the same cluster, and the CMNs send data to the CHN based on the allocation, thereby decreasing collisions [16]. In every cluster, after the CHN receives the data sent by the CMNs, it fuses the received data, which can eliminate the redundant data and reduce the amount of data to be transmitted to the SN, thus contributing to energy conservation [4]. Moreover, the decreased data size reduces collisions during data transmissions. In addition, clustering routing protocols employ a CHN rotation mechanism and select CHNs in every round, which helps avoid the excessive energy dissipation of the selected CHNs, balance the energy consumption, and prolong the network lifecycle [17].

The data fusion technique [18,19], which is used by CHNs to fuse data in this paper, is a common and effective way to eliminate the data redundancy, reduce the data size, and decrease the energy consumption. This paper uses an improved back propagation neural network (BPNN) to implement the data fusion. In UWSNs, sensor nodes may collect the data with high redundancy. When the redundant data are sent to the SN, unnecessary energy consumption arises, leading to a premature death of the node and a shortened lifespan of the network. In contrast, if the CHNs fuse the data and transmit them towards the SN, it could greatly save energy [20]. The four advantages of data fusion technique are concluded in [21]: decreasing energy consumption, enhancing data security, improving transmission efficiency, and optimizing network resources.

The data transmission by the multi-hop mechanism, has been proven in [22] more effective in energy conserving in long-distance transmissions compared to the single-hop mechanism. Thus, we need to find the optimal multi-hop paths to achieve minimized energy consumption, enhanced transmission efficiency, and reduced packet loss ratio during data transmission. In the paper, the SN is the destination node, and the CHN that has data packets to send becomes the source node. The relay node is chosen from CHNs rather than CMNs. Genetic algorithms (GAs), which imitate the natural evolution process to search for optimal solutions, have the potential to resolve the optimization problem and they are effective in finding the optimal multi-hop routing paths [23,24]. The population initialization of the GA is usually based on a random scheme. The fitness function evaluates the individuals and the better ones are more likely to be chosen to produce the next generation [25]. Through repeated crossover, mutation, and selection operations, the population can be improved and the optimal solution can be found [26].

To our knowledge, no existing routing protocol in UWSNs combines the advantages of the clustering scheme, the data fusion technique and the GA. However, it is vital to propose an energy-efficient routing protocol that is capable of minimizing the energy dissipation

and ultimately maximizing the lifecycle of UWSNs. Hence, we present an energy-efficient clustering routing protocol on the basis of a modified GA and an improved BPNN for UWSNs, which could greatly enhance network performance. The proposed protocol has three main phases: the CHN selection, the cluster formation, and the data transmission. In the CHN selection phase, the protocol introduces an optimized CHN selecting scheme. In the second phase, CMNs choose to join the CHNs according to depths of nodes and distances between nodes, and the clusters are thus formed. In the third phase, the CMNs transmit data to the CHNs through a single-hop mechanism. Once receiving the data, the CHNs fuse them by using the improved BPNN algorithm, and forward the fused data to the SN through a multi-hop scheme. Each effective multi-hop transmission path is identified through the enhanced GA.

The innovations of our work are as follows:

1. Based on a new encoding scheme, which encodes routing paths as chromosomes and sensor nodes as genes, this paper presents a modified GA to search for optimal multi-hop routing paths for CHNs to transmit data packets to the SN.
2. This paper proposes a scaling function to reallocate the range of the fitness value in the selection operator of the GA, which helps keep the population diversity and improve the convergence of the GA.
3. This paper introduces a particular crossover operator and an improved mutation operator in the GA, and also adopts an adaptive mutation probability scheme instead of using the fixed mutation probability, which helps avoid the local convergence of the GA.
4. This paper presents an improved BPNN by adopting an optimized momentum method, which is employed by CHNs to fuse data in order to reduce the energy consumption through the elimination of data redundancy and the decrease of the amount of data.
5. This paper introduces an optimized CHN selecting scheme, and improves the cluster formation process by taking into account the depth of the nodes and the distance between nodes.
6. This paper combines the clustering routing protocols, the GA, and the data fusion technique, which is an innovative application in UWSNs. Simulation results verified its effectiveness in improving network performance.

The remainder of this paper is as follows. Section 2 introduces the related work. Section 3 describes the network model and the energy consumption model. The modified GA is presented in Section 4. Section 5 focuses on the improved BPNN. Section 6 presents the proposed clustering routing protocol. The experiments are analyzed in Section 7. The conclusion is drawn in Section 8.

2. Related Work

To reduce the energy consumption and prolong the network lifetime, many studies have been done. This section presents related works concerning the clustering routing protocol, the data fusion technique, and the GA. In Section 2.1, some clustering routing protocols are presented and the difference between these protocols and the proposed underwater clustering routing protocol in this paper is provided. In Sections 2.2 and 2.3, some researches about the data fusion technique and the GA are reviewed. Because the proposed underwater clustering routing protocol cannot be comparable to the data fusion technique or the GA, we just summarize the advantages of the data fusion technique and the GA, and present that they can reduce energy dissipation in UWSNs.

2.1. The Clustering Routing Protocol

This section presents related works on clustering routing protocols and discusses the difference between them and our proposed protocol. The earliest one is the low-energy adaptive clustering hierarchy (LEACH) protocol that uses a probabilistic method to select CHNs, but the remaining energy of nodes is not considered [27]. This makes

some selected CHNs die too early, which affects the balance and efficiency of the network energy. Moreover, the LEACH does not support the multi-hop transmission mechanism. Therefore, researchers proposed the improved clustering routing protocols based on the LEACH. Lee et al. optimized the LEACH based on expected residual energy (LEACH-ERE), which adopts an improved CHN selection scheme based on the LEACH protocol, employs energy predication, and distributes the network load evenly in order to extend the network lifetime [28]. Mohapatra et al. presented a partitioned-based and energy-efficient LEACH (PE-LEACH) protocol that divides the whole network into quadrants, which is energy-efficient and fault-tolerant [29]. In addition, the CHN selection scheme and the data transmission process are improved in PE-LEACH protocol. However, the protocols in [28] and [29] are designed for TWSNs, and they should be modified for UWSNs. Wang et al. adopted an energy-efficient grid routing based on 3D cubes (EGRCs) for UWSNs, where the network is divided into lots of small cubes and each cube is regarded as a cluster [30]. What is more, the EGRC protocol optimizes the CHN selection and improves the search process for the next-hop node. However, the EGRC does not present the detail of the data fusion mechanism as the data redundancy may exist and should be reduced. In [31], an underwater clustering protocol on the basis of the fuzzy c means and the moth-flame optimization (FCMMFO) was proposed to enhance the performance of UWSNs. In the FCMMFO, the optimal number of clusters is determined by using the fuzzy c means and the appropriate CHNs are selected by the moth-flame optimization. Nevertheless, a multi-hop mechanism is not provided in the FCMMFO. Krishnaswamy et al. presented an energy-efficient underwater clustering protocol based on the fuzzy scheme and particle swarm optimization (FBCPSO) in [32], where the fuzzy scheme and particle swarm optimization are used to form clusters and select CHNs respectively, which can balance and reduce the energy dissipation. However, the multi-hop routing mechanism and the data fusion method are not considered by the authors. Wang et al. put forward an underwater clustering scheme based on the magnetic induction for UWSNs [33], where the Voronoi diagram is employed to form clusters and the jellyfish breathing process is used for CHN selection. This scheme can achieve the high energy-efficiency and prolong the network lifetime. However, the multi-hop routing path has not been optimized in [33]. Ahmed et al. introduced an underwater clustering protocol according to redundant transmission control (RTC), which eliminates the data redundancy at the CHN level and at the region head level [6]. Moreover, the authors presented a dynamic CHN rotation method, which can balance the energy consumption and improve the reliability of the network. However, this scheme relies on a mobile SN that moves from the surface to the bottom to collect data. Islan et al. presented an underwater clustering-based localization protocol [34], where the CHNs perform localization procedure rather than the whole cluster. Furthermore, the retransmission control mechanism is carried out to control unnecessary transmission, which can reduce energy dissipation. Nevertheless, this protocol does not consider the data redundancy that affects the energy consumption. Wan et al. presented an underwater adaptive clustering routing scheme [35], where the CHNs perform the data fusion in order to decrease the energy loss. Moreover, the competition radius of nodes is decided based on the distance factor and the residual energy of nodes. The selection of CHNs and the routing rules are in the light of node energy, but the influence of the distance has not been considered. Bansal et al. provided a multilevel underwater clustering protocol [36], where the cluster and the logical level are formed based on the remaining energy of nodes rather than the geography. The nodes with the similar level of energy are thought to be in the same level and only the nodes with the highest level of energy communicate with the SN. Moreover, the protocol employs the multi-hop transmission mechanism and the data fusion technique, but the details are not given. Zou et al. proposed an underwater cluster-based adaptive routing (CBAR) protocol [37], which optimizes the network architecture and establishes the routing path based on the focus beam routing. Moreover, the CBAR employs a dynamic routing update mechanism and a power control mechanism. However, the

multi-hop routing path is not optimized and the detail of the data redundancy elimination scheme is not given in the CBAR.

2.2. The Data Fusion Technique

This section presents related works on the data fusion technique and describes that it can be used to eliminate the data redundancy, thus reducing the energy consumption during data transmissions. Sun et al. presented a data fusion method based on BPNNs, and they put the input layer of the BPNNs in CMNs, and put the hidden and output layers in CHNs [20]. Only the fused data representing the features of the input data are sent to the SN in order to improve energy efficiency. Cao et al. developed a clustering protocol in the light of data fusion scheme by using BPNNs for TWSNs, which adopts a stable election protocol model based on the LEACH protocol to select appropriate CHNs [38]. The selected CHNs fuse the data after receiving them and send the fused data to a destination node. Yue et al. proposed a data fusion scheme by employing an improved radial basis function neural network in mobile TWSNs, which improves the data fusing model so as to reduce the energy consumption [39]. Nevertheless, the underwater environment has not been taken into account in [20,38,39]. Goyal et al. introduced a fuzzy-based clustering routing protocol combined with the data fusion technique for UWSNs, where the residual energy, the distance, the node density, the load, and the link quality are considered as inputs to the fuzzy logic as a way to select CHNs and determine the cluster size [40]. The selected CHNs fuse the received data and transmit them to a destination node, reducing the energy dissipation and enhancing the network lifespan. The clustering routing protocol with a two-tier data fusion for UWSNs is described in [41], where CMNs reduce the data redundancy before transmitting the data to CHNs. The CHNs adopt a developed K-means method based on an ANOVA model to aggregate the received data, and send the aggregated data to the SN, thereby minimizing the energy consumption. Wang et al. introduced a data fusion technique based on the BPNN, which combines with clustering routing protocols [42]. The scheme optimizes the selection of CHNs, and the selected CHNs extract features from the data and send them to the SN, which can save energy. Gang et al. described a data fusion method on the basis of the rough set theory and the BPNN [43], where the rough set theory is used to reduce redundant data and the reduced useful data are used to train the BPNN. It has been validated that this protocol can enhance the performance of the data fusion system and improve the training speed of the BPNN. Lin et al. introduced a data collection and fusion mechanism that uses a mobile SN to collect the data from collection points [44]. The collection points are selected periodically and the collection points are the places where the data fusion is performed, which is capable of reducing the energy consumption and extending the network lifetime.

2.3. The GA

This section presents the related works concerning the GA, which demonstrate its effectiveness in finding the optimal multi-hop transmission paths, thereby saving energy, extending the network lifetime, and decreasing the transmission delay. Lorenzo et al. proposed an improved GA to optimize the routing paths, which encodes the paths as chromosomes and presents special crossover and mutation operations for realizing the optimal topology [45]. Moreover, they developed the fitness function by considering power consumption, time delay, and throughput of the network. The GA they proposed possesses the merits of fast convergence and robustness. Lu et al. presented an improved GA to optimize the multicast routing by using a simplified encoding operation, a special crossover operation, and a modified mutation operation [46]. In addition, they defined the fitness function based on the energy cost and time delay, which can decrease the energy consumption and extend the life expectancy of the network. Silva et al. put forward a routing protocol based on GAs that are used to look for suitable routing paths to satisfy the requirements of anycast sessions, which can improve the efficiency of the delay tolerant network [47]. An optimal multi-hop path finding method (OMPFM) was proposed in [48],

where an enhanced GA is adopted to find the optimal paths through the proposal of a fitness function. Furthermore, the performance of the GA is improved in the execution time and the chromosome quality. Results show that the OMPFM can find an optimal multi-hop path, thereby saving energy and prolonging the network lifetime. In [49], ThamaraiKannan et al. introduced a compact GA to select the optimal path for mobile ad-hoc networks, which can reduce the path cost, improve the packet delivery rate and decrease the energy consumption. Xin et al. presented a modified GA through the increase in the number of offspring and the conduction of the second fitness assessment that can remove the undesirable offspring and keep the dominant individuals [50]. Moreover, this enhanced GA was used in the navigation, as well as the control system of unmanned surface vehicles, and simulation results indicate the GA performs well in the convergence speed, the robustness, and the optimal path searching.

Therefore, combining the clustering routing protocol, the data fusion technique, and the GA in UWSNs could greatly reduce the energy dissipation and prolong the network lifetime. In our proposed underwater clustering routing protocol, the data fusion technique is used by CHNs to eliminate the data redundancy and the GA is employed to find the optimal multi-hop transmission paths when CHNs transmit the fused data to the SN.

3. Model Assumptions

3.1. Network Model

In the section, we present a three-dimensional network model, which is shown in Figure 1. Underwater acoustic sensors are distributed at random within the marine environment, and other details are:

1. There are two kinds of nodes: underwater sensor nodes, which are immobile and divided into CHNs and CMNs after cluster formation, and an SN, which is located on the surface of the monitoring area.
2. There is only one SN in the network, which is the destination node and has energy supplies. Nevertheless, underwater sensor nodes have limited energy and they do not have energy supplies.
3. The ordinary underwater nodes have the equal initial energy and the unique IDs.
4. The locations of nodes could be acquired through the localization algorithm [51].
5. We could control transmitting power based on the different distances to receiving nodes.
6. CMNs gather data and transmit them to CHNs through a single hop. Once the CHN receives the data, the CHN fuses them and forward them towards the SN through multiple hops. If one CHN is close to the SN, it sends data towards the SN through one hop.

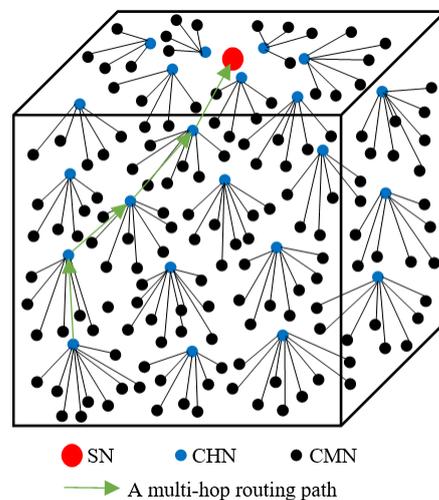


Figure 1. The schematic diagram of the network model.

3.2. Energy Consumption Model

The underwater energy consumption model provided in [52] is employed in the paper. This paper assumes P_0 is minimal power that a node needs to receive packets, and minimal transmitting power should reach $P_0A(l)$, where $A(l)$ denotes an attenuation function. The energy consumption for transmitting and receiving can be calculated by:

$$E_t(l) = T_t P_0 A(l) \tag{1}$$

$$E_r = T_r P_0 \tag{2}$$

$$A(l) = l^{1.5} a^l \tag{3}$$

$$a = 10^{\alpha(f_c)/10} \tag{4}$$

$$\alpha(f_c) = 0.11 \frac{f_c^2}{1 + f_c^2} + 44 \frac{f_c^2}{4100 + f_c^2} + 2.75 \times 10^{-4} f_c^2 + 0.003 \tag{5}$$

where $E_t(l)$ is the energy consumption for transmitting, and E_r is the energy consumption for receiving. T_t is the time for nodes to send packets, and T_r is the time to receive packets. l is the distance between transmitting nodes and receiving nodes. $\alpha(f_c)$ is the absorption coefficient in dB/km and f_c is the frequency in kHz.

4. The Improved GA

This section presents the improved GA that is used to find the optimal multi-hop paths between the CHNs and the SN, where the novel encoding scheme, as well as the specific selection, crossover, and mutation operators is proposed. The optimal paths can improve transmission efficiency, reduce packet loss ratio, and minimize energy consumption, thereby prolonging the network lifetime and improving the network performance.

4.1. The Problem Description

We assume that there are $N-1$ CHNs and 1 SN when implementing the GA to search for the optimal paths. The SN is the destination node. The CHN that needs to transmit data becomes the source node. The relay node is chosen from CHNs. Let x_{ij} , c_{ij} , d_{ij} , and l_{ij} denote the link indicator, the link energy cost, the link delay, and the link length between node i and node j , respectively. T_{ti} is the time duration for the node i to transmit packets and T_{rj} is the time duration for the node j to receive packets. D_{tmax} presents the maximum delay of the path. The value of x_{ij} is 1 when a link exists between node i and node j . Otherwise, the value of x_{ij} is 0. We regard the search process of multi-hop paths as a combinatorial optimization problem, finding the optimal path with the minimum cost. The objective function is given by:

$$\text{minimize : } F_{obj} = \sum_{i=1}^N \sum_{j=1}^N c_{ij} x_{ij} \tag{6}$$

$$\text{where } c_{ij} = T_{ti} P_0 A(l_{ij}) + T_{rj} P_0 \tag{7}$$

$$\text{subject to : } \sum_{i=1}^N \sum_{j=1}^N d_{ij} x_{ij} < D_{tmax} \tag{8}$$

The constraint Equation (8) makes sure that the total transmission delay is limited to a certain value so that it will not be too high.

4.2. The Encoding Scheme

This paper encodes routing paths as chromosomes and nodes as genes. The first gene of the chromosome presents the source node and the last gene of the chromosome denotes the destination node. The number of genes in one chromosome is not an invariant, which means that different routing paths could consist of different number of nodes. Moreover, one gene cannot appear at the different locations of one chromosome, which means that

one node can only appear once in one routing path so as to prevent the loops and improve the efficiency of the path. However, if it happens, this paper adopts the repair mechanism to solve it as described in Section 4.6. Figure 2 demonstrates the encoding process of a routing path from the source node to the destination node.

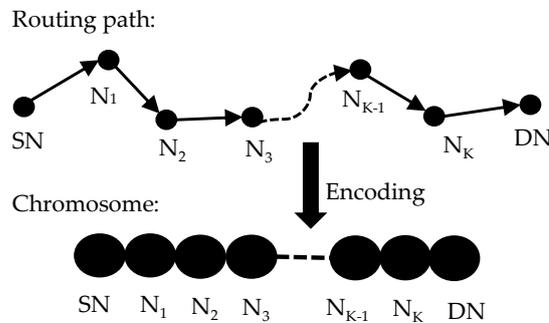


Figure 2. The process of encoding.

4.3. The Initialization

The initialization of population size, which is the number of the chromosomes, and the initialization of the chromosome formation should be taken into account before implementing the operation of the GA. The population size is vital to the GA and should be decided by the specific circumstance. It is more likely for the GA having more initial chromosomes to search for optimal solutions. However, it takes more time for the algorithm to converge and it is also a waste of resources. A small number of chromosomes may save network resources, but may lead to an undesired outcome. The initialization of the chromosome formation is based on the random selection. In this paper, the first gene represents the source node. The second gene is chosen randomly from the neighboring nodes of the source node and the third gene is picked randomly from the neighboring nodes of the second node. The procedure does not stop until the destination node is found. Additionally, one node should not be chosen repeatedly on one path in order to avert loops in paths.

4.4. The Fitness Function

In the GA, it is more likely for the individual with higher fitness value to be selected to generate the next generation. Hence, it is indispensable to design the fitness function, which demonstrates the characteristics of chromosomes so as to find the optimal chromosome that is the optimal routing path with the minimal cost. Accordingly, we define the fitness function:

$$F_m = \frac{\Phi(z)}{F_{obj}} = \frac{\Phi(\sum_{i=1}^N \sum_{j=1}^N d_{ij}x_{ij} - D_{tmax})}{\sum_{i=1}^N \sum_{j=1}^N c_{ij}x_{ij}} \tag{9}$$

$$\Phi(z) = \begin{cases} 1, & \text{if } z \leq 0 \\ \lambda, & \text{if } z > 0 \end{cases} \tag{10}$$

where F_m represents the fitness function of the m th chromosome, $\Phi(z)$ denotes the penalty function, and λ ranging from 0 to 1 decides the level of penalty. When the total path delay exceeds the maximum value D_{tmax} , the penalty function will affect the value of fitness function and always decrease the value, which means the penalty function could reduce the chance of a chromosome being selected for the next generation. If the value of λ is high, the level of penalty will be low. Otherwise, the level of penalty will be high.

4.5. The Selection Operator

One chromosome represents one routing path from the source node to the destination node. However, some paths may cost too much energy and it is better not to choose the

corresponding chromosomes to produce the next generation. Therefore, we adopt the roulette wheel selection as the selection operator to choose the chromosomes with high quality. The probability of choosing one chromosome to perform the crossover operation is presented by:

$$P_m = \frac{F_m}{\sum_{m=1}^M F_m} \quad (11)$$

where P_m is the probability for choosing the m th chromosome as a parent, which is higher when the chromosome has a higher fitness value. M represents the population size. However, this operator may result in the loss of population diversity because it is sensitive to the probability. To alleviate this problem, we propose a scaling function to reallocate the range of the fitness value. By referring to the simulated annealing algorithm, the scaling function is given as follows:

$$Q_m = \exp(-100\beta^{g-1}F_m) \quad (12)$$

where Q_m denotes the scaled fitness function of the m th chromosome. β is adjustment coefficient ranging from 0 to 1. g represents the number of generations. As shown in Equation (12), in early generations, it can narrow the gap between the fitness values of different chromosomes so that the potential chromosomes can be selected, thereby settling the local optimum problem. In late generations, it can amplify the difference between the chromosomes that have the close fitness values so as to highlight the advantages of the good-quality chromosomes, which renders the superior chromosomes selected to pass on to the next generation for the purpose of accelerating the convergence of the algorithm.

4.6. The Crossover Operator

Using the selection operator, the chromosomes are picked for the crossover operation to produce the offspring according to the crossover probability. In this process, two chromosomes generate two new chromosomes by exchanging some parts of them, but it is noted that these two chromosomes (paths) should have one or more same genes (nodes) besides the source node and the destination node because it may produce infeasible routing paths easily otherwise. The places of the same genes in two chromosomes are where the crossing points lie. Two crossover methods are adopted in this paper: single-point crossover and two-point crossover, which differ from the traditional ones. The single-point crossover is carried out when there is only one common gene in the two chromosomes and they exchange the latter parts of themselves, which start from the crossing point to the destination node. Two new chromosomes are thus formed as demonstrated in Figure 3. Additionally, as shown in Figure 4, the two-point crossover is used when two common genes exist in the two chromosomes and they exchange the parts that are between the two same genes so as to form two new chromosomes. If three or more same genes exist in two chromosomes, the paper still adopts the two-point crossover method and the crossing points are selected randomly from the same genes.

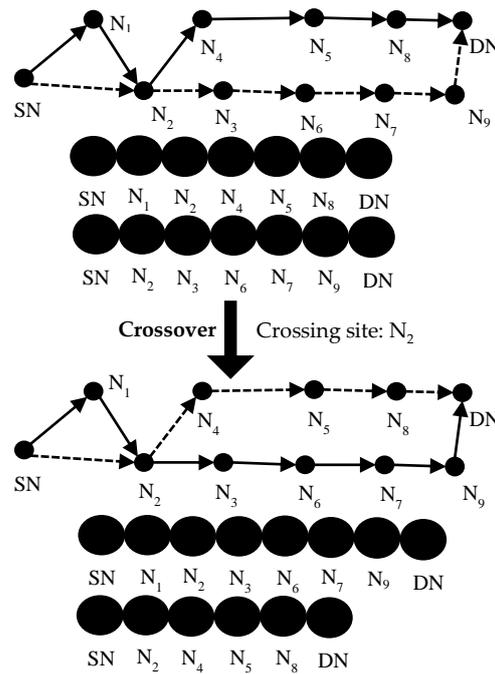


Figure 3. The single-point crossover procedure.

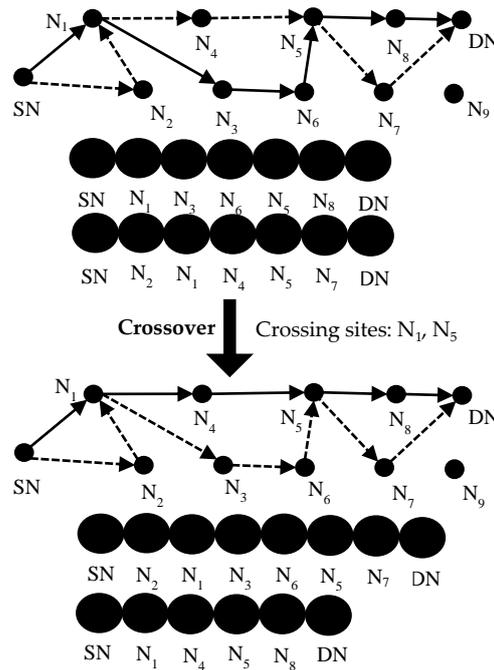


Figure 4. The two-point crossover procedure.

As illustrated in Figures 3 and 4, some of the new produced chromosomes are better than the original ones, which ensures that the preferable paths can be found. Therefore, the crossover operation can improve the ability of the path search, thus accelerating the algorithm convergence and finding the optimal path. However, sometimes the crossover operation may cause path loops, which is not desirable in the path search. Therefore, the repair mechanism is adopted to look for the loops and then wipe them out. The key point

is to find out whether one node exists in the different locations of one path. For example, there are two chromosomes representing the two paths:

$$\begin{aligned} \text{Path 1 : } & SN \rightarrow N_1 \rightarrow N_3 \rightarrow N_2 \rightarrow N_5 \rightarrow \dots \rightarrow DN \\ \text{Path 2 : } & SN \rightarrow N_2 \rightarrow N_4 \rightarrow N_3 \rightarrow N_5 \rightarrow \dots \rightarrow DN \end{aligned}$$

The two crossing points are N_3 and N_5 . After crossover operation, the produced paths are:

$$\begin{aligned} \text{Path 3 : } & SN \rightarrow N_1 \rightarrow N_3 \rightarrow N_5 \rightarrow \dots \rightarrow DN \\ \text{Path 4 : } & SN \rightarrow N_2 \rightarrow N_4 \rightarrow N_3 \rightarrow N_2 \rightarrow N_5 \rightarrow \dots \rightarrow DN \end{aligned}$$

There exists a loop $N_2 \rightarrow N_4 \rightarrow N_3 \rightarrow N_2$ in path 4. After the loop is wiped out, the path becomes a feasible one: $SN \rightarrow N_2 \rightarrow N_5 \rightarrow \dots \rightarrow DN$.

4.7. The Mutation Operator

The mutation randomly happens to chromosomes and changes the genes according to the mutation probability, which could provide the genes that do not exist in the population or those that are lost in the early operation, thereby retaining the diversity of the population and avoiding the local convergence. The mutation operator starts a new path search from the mutation gene (node) to the destination node at random, and this process of the partial path search is the same as the process of the initialization of the path (chromosome) as described in Section 4.3. In addition, the partial path between the source node and the mutation node stays the same as shown in Figure 5. What calls for special attention is that the nodes that already exist in the previous path extending from the source node to the mutation node should not be added to the path during the new partial path searching process in order to prevent loops.

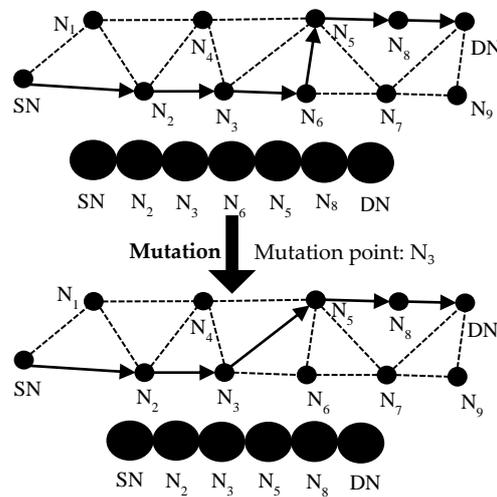


Figure 5. The mutation procedure.

To avoid the local convergence, this paper adopts an improved mutation operator by adjusting mutation probability adaptively instead of using the fixed mutation probability applied in the conventional algorithm. The proposed one is given by:

$$P_{mmut} = \begin{cases} \frac{P_{mutmax} - P_{mutmin}}{1 + \exp(\mu(\frac{2(Q_m - Q_{avg})}{Q_{max} - Q_{min}}))} + P_{mutmin}, & Q_m \geq Q_{avg} \\ P_{mutmax}, & Q_m < Q_{avg} \end{cases} \quad (13)$$

where P_{mmut} denotes the mutation probability of the m th chromosome. P_{mutmax} and P_{mutmin} are the maximum mutation probability and the minimum mutation probability. Q_{avg} ,

Q_{max} , and Q_{min} denote the average, maximum, and minimum scaled fitness values in the population, respectively. As displayed in Equation (13), the mutation probability of chromosomes is related to its scaled fitness value. The individual with a smaller fitness value has a higher chance to mutate so as to help remain the good-quality chromosomes, as well as keep the diversity of the population, which prevents the premature convergence of the algorithm.

4.8. The Termination Mechanism

When the mutation operation finishes, the next generation is produced. After the maximum number of iterations, one optimal multi-hop routing path from the source node to the destination node can be determined by selecting the chromosome with the largest fitness value in the population. That means one path between one CHN and the SN is determined. However, it is noted that there are $N-1$ CHNs in the network. Hence, the improved GA ends when all the CHNs find their paths to the SN.

5. The Improved BPNN

This section presents an improved BPNN that is used by the CHNs to perform data fusion after they receive data sent by CMNs, which can eliminate the redundant data and reduce the amount of transmitted data, thus saving the network energy and extending the network lifespan.

5.1. The BPNN Description

The three-layer neural network consisting of one input layer, one hidden layer, and one output layer is adopted in this paper, which is competent for most of the complicated problems. Figure 6 illustrates the structure of the BPNN. We assume that the input signal and the output signal for the structure are $\mathbf{U} = [u_1, u_2, \dots, u_U]$ and $\mathbf{Y} = [y_1, y_2, \dots, y_Y]$, respectively. U , R , and Y denote the number of neurons of the input layer, hidden layer, and output layer, respectively. Then the outputs of the hidden layer and the output layer can be calculated by:

$$h_j = f_v\left(\sum_{i=1}^U w_{ij}u_i + b_j\right) \quad (14)$$

$$y_k = f_v\left(\sum_{j=1}^R w_{jk}h_j + b_k\right) \quad (15)$$

$$f_v(v) = \frac{1}{1 + \exp(-v)} \quad (16)$$

where h_j and y_k represent the outputs of the j th neuron in the hidden layer and the k th neuron in the output layer, respectively. w_{ij} denotes the weight value connecting the i th neuron in the input layer and the j th neuron in the hidden layer, and w_{jk} indicates the weight value connecting the j th neuron in the hidden layer and k th neuron in the output layer. b_j and b_k are the biases of the j th neuron in the hidden layer and the k th neuron in the output layer, respectively. $f_v(v)$ is the activation function of the hidden layer and the output layer. The overall output is usually different from the expected output and the error function is thus employed, which is to be minimized and is given by:

$$e_{err} = \frac{1}{2} \sum_{k=1}^Y e_k^2 = \frac{1}{2} \sum_{k=1}^Y (y_k - y'_k)^2 \quad (17)$$

where y'_k represents the expected output of the k th neuron in the output layer. By propagating the error backward, the weights and biases can be adjusted based on the gradient

descent method. Hence, the error can be reduced gradually. The adjustments for the weights and the biases can be obtained by:

$$w_{ij}(t + 1) = w_{ij}(t) - \eta \frac{\partial e_{err}}{\partial w_{ij}(t)} \tag{18}$$

$$w_{jk}(t + 1) = w_{jk}(t) - \eta \frac{\partial e_{err}}{\partial w_{jk}(t)} \tag{19}$$

$$b_j(t + 1) = b_j(t) - \eta \frac{\partial e_{err}}{\partial b_j(t)} \tag{20}$$

$$b_k(t + 1) = b_k(t) - \eta \frac{\partial e_{err}}{\partial b_k(t)} \tag{21}$$

where η is the learning rate that should be set appropriately so as to speed up the training process, and t denotes the number of training times. The training does not cease until the error is decreased to a certain value or the preset number of training times is reached. However, the fixed learning rate sometimes cannot achieve high efficiency during the training. Accordingly, this paper employs an adaptive adjustment method for η as described in next section.

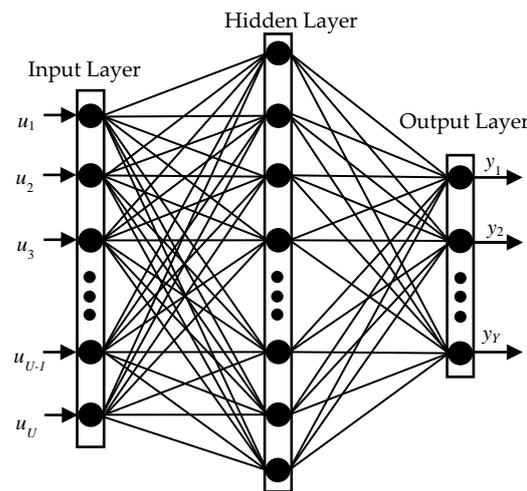


Figure 6. The structure of BPNN.

5.2. The Improved Momentum Method

The standard BPNN algorithm has the problem of slow convergence and is easy to run into a local minimum as a result of the adoption of the gradient descent method. This paper brings in the momentum method to adjust the weights and the biases as shown below:

$$\Delta w(t + 1) = -\eta(1 - \gamma) \frac{\partial e_{err}}{\partial w(t)} + \gamma \Delta w(t) \tag{22}$$

$$\Delta b(t + 1) = -\eta(1 - \gamma) \frac{\partial e_{err}}{\partial b(t)} + \gamma \Delta b(t) \tag{23}$$

$$\Delta w(t + 1) = w(t + 1) - w(t) \tag{24}$$

$$\Delta b(t + 1) = b(t + 1) - b(t) \tag{25}$$

where $\Delta w(t + 1)$ and $\Delta b(t + 1)$ are the increments of the weights and the bias, respectively. γ ranging from 0 to 1 denotes the momentum factor. As shown in (22), the added momentum $\gamma \Delta w(t)$ can reduce the oscillation of the training process and thus, the convergence can

be improved. To further enhance the performance of the method, we propose an improved momentum method as follows:

$$\Delta w(t+1) = -\eta(1-\gamma)\frac{\partial e_{err}}{\partial w(t)} + \gamma\Delta w(t) + \sigma\Delta w(t-1) \quad (26)$$

$$\Delta b(t+1) = -\eta(1-\gamma)\frac{\partial e_{err}}{\partial b(t)} + \gamma\Delta b(t) + \sigma\Delta b(t-1) \quad (27)$$

where σ is a constant that should be smaller than γ . However, the learning rate η is a fixed value in the method, which can be improved because during the training process, the learning rate should be higher when the learning process needs to be accelerated and it should be lower when the algorithm stability is the priority. Therefore, this paper employs an adaptive adjustment method for the learning rate, which is presented by:

$$\eta_w(t+1) = \eta_w(t) \times 2^{f_{sig1}} \quad (28)$$

$$\eta_b(t+1) = \eta_b(t) \times 2^{f_{sig2}} \quad (29)$$

$$f_{sig1} = \text{sign}\left[\left(-\frac{\partial e_{err}}{\partial w(t)}\right) \times \left(-\frac{\partial e_{err}}{\partial w(t-1)}\right)\right] \quad (30)$$

$$f_{sig2} = \text{sign}\left[\left(-\frac{\partial e_{err}}{\partial b(t)}\right) \times \left(-\frac{\partial e_{err}}{\partial b(t-1)}\right)\right] \quad (31)$$

$$\text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases} \quad (32)$$

where $\eta_w(t+1)$ and $\eta_b(t+1)$ denote the adaptive learning rate for the weights and the bias, respectively. The adaptive adjustment method can coordinate the training speed and the algorithm stability, thereby improving the convergence performance and finding the optimal solution.

6. The Proposed Clustering Routing Protocol

This section presents our proposed energy-efficient clustering routing protocol (EECRP) based on the modified GA and the improved BPNN that is used for data fusion. Referring to the LEACH protocol [27], the EECRP has three phases: CHN selection, cluster formation, and data transmission. In every cluster, CMNs transmit data to the CHN through one hop. Once the CHNs receive the data, they perform data fusion by using the improved BPNN algorithm and transmit the processed data to the SN through multiple hops. The relay nodes are other CHNs and the optimal multi-hop transmission paths are determined through the improved GA.

6.1. CHN Selection Phase

Selecting appropriate CHNs is of great importance to reduce and balance energy consumption. The CHNs receive the data from the CMNs, fuse the data, and transmit the fused data to the SN. The original LEACH generates CHNs through a probabilistic selection and the residual energy of nodes has not been taken into account. These selected nodes may die too early as a result of their insufficient remaining energy, which affects the balance and efficiency of the network energy. Hence, by taking the residual energy of nodes into consideration, we propose an improved CHN selection scheme as follows:

$$H_{th} = \begin{cases} \frac{P_{CHN}}{1-P_{CHN} \times (r \bmod \frac{1}{P_{CHN}})} \times \frac{E_{res}}{E_{av}}, & \text{node} \in G \\ 0, & \text{otherwise} \end{cases} \quad (33)$$

where H_{th} is the threshold for the node, and P_{CHN} denotes the percent of CHNs in the network (e.g., $P_{CHN} = 10\%$), and r is the current round, and E_{res} represents the residual energy of the node, and E_{av} is the average residual energy of all nodes, and G is the node set where the nodes do not become CHNs in the last $1/P_{CHN}$ rounds. In this process, every node produces a random number ranging from 0 to 1 and if the number of one node is less than its threshold H_{th} , it turns into a CHN candidate. Then the CHN candidate broadcasts candidate-messages with the residual energy to its neighbor nodes. If one neighbor node that receives the candidate-message is also a CHN candidate, the one with higher residual energy becomes a CHN, which can prevent the geographically close nodes from being CHNs. If one CHN candidate does not receive any candidate-messages from its neighbor nodes for a certain time, it becomes the CHN.

6.2. Cluster Formation Phase

When CHNs are successfully selected, every CHN broadcasts a CHN-message to invite non-CHNs to join it, which carries information such as the node ID, the node energy, and the node location. When a non-CHN receives the broadcast message, it judges whether the CHN is deeper than it is because the non-CHN only chooses to join the CHN in a shallower position. If a non-CHN receives two (or more) CHN-messages from CHNs in shallower positions, it selects the nearer (or nearest) CHN to join and replies with an acknowledgement message. If a non-CHN receives only one CHN-message, it directly replies to the CHN with an acknowledgement message. If a non-CHN does not receive any CHN-message, it will wait for a period of time until it receives one. After non-CHNs join CHNs, they become CMNs and clusters are hence formed.

6.3. Data Transmission Phase

After the clusters are formed, data transmission phase could start. In every cluster, the CHN allocates time slots through the time division multiple access mechanism for its CMNs, and the CMNs transmit data to the CHN based on the time slots through a single hop, thereby decreasing collisions. After transmitting the data for this round, the CMNs go into sleep mode so as to save energy. Once the CHNs receive the data, they fuse the data by using the improved BPNN algorithm and forward the processed data to the SN by employing the carrier sense multiple access with collision detection scheme. Each effective multi-hop transmission path to the SN is identified through the enhanced GA. If one CHN is close to the SN, it sends data towards the SN through one hop. It is noted that the training processes of the BPNN algorithm are conducted by the SN due to its energy supplies. Once clusters are formed, the SN transmits the trained weights and biases of the BPNN to the CHNs. Based on the trained model, the CHNs fuse the data, eliminate the redundancy, extract the features, and then send the processed data to the SN. In addition, the searching process of the multi-hop transmission paths is also completed in the SN. After the CHN selection phase, all the CHNs transmit the message packets with information such as the node ID, the node energy, and the node location to the SN. Then the SN figures out the optimal multi-hop transmission paths by employing the GA and sends the routing path information to the CHNs.

After the SN receives the data from all the CHNs, one round ends. If the remaining energy of every CHN is over half of the average residual energy of all the nodes, the CHNs of the next round remain unchanged, thereby saving time and energy. Hence, the next round directly begins with the data transmission phase. Otherwise, the next round starts with the CHN selection phase.

7. Simulation Results and Performance Analyses

In this section, some existing underwater clustering routing protocols: EGRC [30], LEACH-ERE [28], LEACH [27], FCMMFO [31], and FBCPSO [32] were selected as the references to verify the proposed EECRP. The used metrics for evaluating the performance were the network lifetime, the energy consumption, and the packet loss rate. We used

MATLAB to conduct the experiments. MATLAB is a simulation software, which can be applied to sensor networks, data analysis, deep learning, image processing, computer vision, risk management, control systems, communications, signal processing and so on. It is an abbreviation of matrix and laboratory and it is developed by MathWorks. The MATLAB settles the high-tech computing problems such as scientific computing, visualization, and interactive programming. It integrates many powerful functions like numerical analysis, matrix calculation, scientific data visualization, and nonlinear dynamic system modeling and simulation in an easy-to-use software environment. It provides a comprehensive solution for scientific research, engineering design, and many scientific problems that require effective numerical calculations. Moreover, it gets rid of the editing mode of traditional non-interactive programming languages, such as C and Fortran, to a large extent, and it provides many feature-rich practical toolboxes such as signal processing toolboxes and communication toolboxes. Figure 7 displays the MATLAB workspace and Figure 8 illustrates some code of calculating the nodes alive.

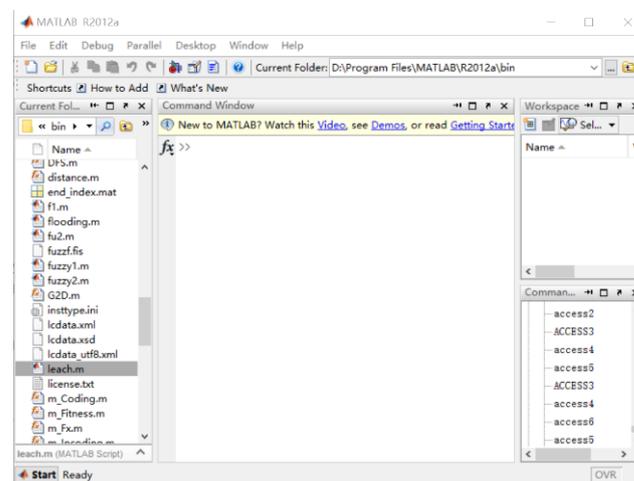


Figure 7. The MATLAB workspace.

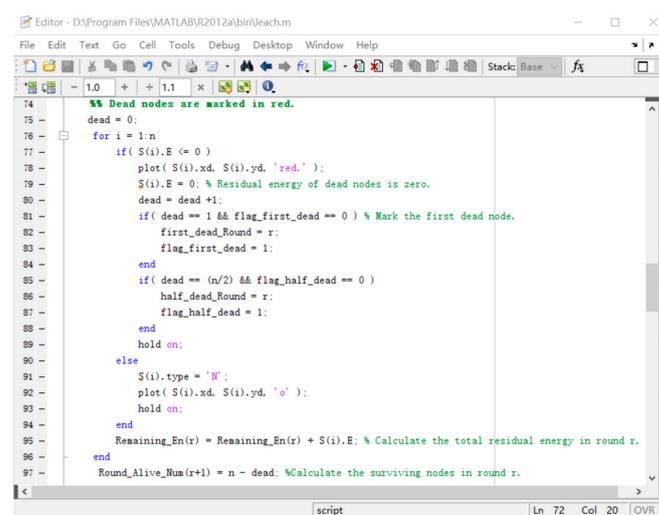


Figure 8. The code of calculating the nodes alive.

The simulation parameters are shown in Table 1:

Table 1. The simulation parameters.

Simulation Parameters	Values
Network size	5 km × 5 km × 1 km
The number of nodes	200, 250, 300
The percent of CHNs	10%
SN coordinate	(2500, 2500, 0)
Data packet size	1024 bits
Other packets size	64 bits
Sound velocity	1500 m/s
Transmission rate	2048 bps
Receiving power	50 μW
Energy initialization of nodes	100 J
Energy consumption for data fusion	50 nJ/bit
Frequency (f_c)	10 kHz

7.1. The Network Lifetime

This section compares the six protocols and analyzes the network lifetime of them by the number of surviving nodes in deferent rounds when 300 nodes are considered in the network. As illustrated in Figure 9, regardless of which protocol we use, the number of surviving nodes decreases as the number of rounds increases. Nevertheless, our proposed EECRP outperforms its competitors in the number of nodes alive. For better evaluation of the EECRP, we bring in indicators namely FND (first node dead), HND (half of the nodes dead), and LND (last node dead). As shown in Figure 10, the first node of the LEACH, LEACH-ERE, EGRC, FBCPSO, FCMMFO, and EECRP dies in about the 353rd, 451st, 505th, 548th, 574th, and 623rd round, respectively, which means that in terms of the FND indicator, the efficiency of the EECRP is 8.5%, 13.7%, 23.4%, 38.1%, and 76.5% higher than that of the FCMMFO, FBCPSO, EGRC, LEACH-ERE, and LEACH, respectively. In terms of the HND and the LND, the EECRP outperforms the LEACH protocol by 57.3% and 46.5%, respectively. To conclude, the proposed EECRP is the most effective in extending the network lifetime as it uses the enhanced CHN selecting scheme, which distributes the network load equally. In addition, the EECRP uses the BPNN to fuse data and adopts the GA to identify the optimal multi-hop transmission paths, reducing and balancing the energy consumption. The LEACH performs the worst among these protocols because it does not take the residual energy of nodes into account when selecting CHNs, which makes some selected nodes with low energy die too early. Additionally, the multi-hop transmission paths between the CHNs and the SN have not been considered in the LEACH. The FCMMFO and the FBCPSO outperform the LEACH, LEACH-ERE and EGRC. Nevertheless, they are both inferior to the EECRP, which is because they do not optimize the multi-hop routing paths between the CHNs and the SN.

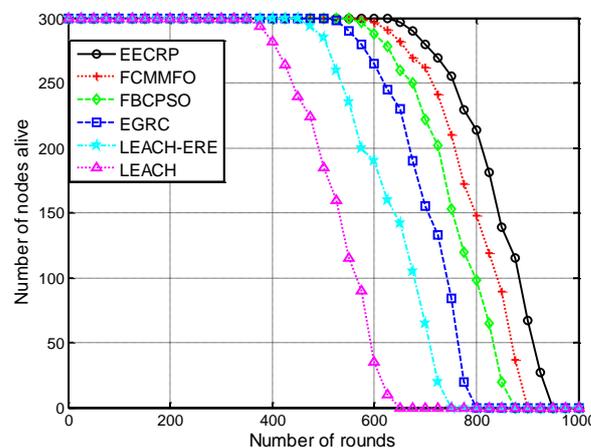


Figure 9. The number of nodes alive versus the number of rounds for different protocols.

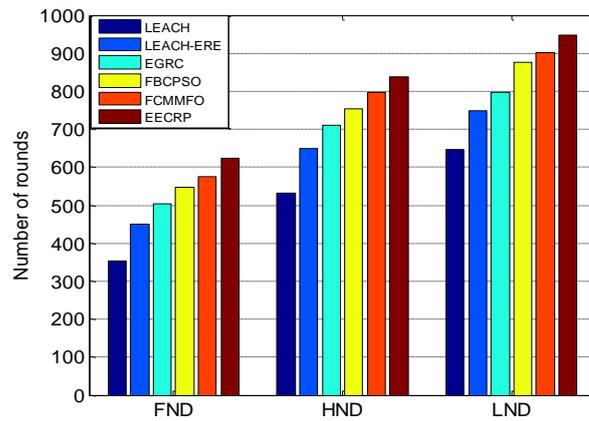


Figure 10. The number of rounds when FND (first node dead), HND (half of the nodes dead), and LND (last node dead) arise for different protocols.

7.2. The Energy Consumption

This section compares and analyzes the six protocols by the energy consumption. As displayed in Figure 11, when 300 network nodes are considered, the total energy consumption of the network increases as the number of rounds rises no matter which protocol is used. Nevertheless, the proposed EECRP has the best performance in energy consumption. For instance, in round 400, the total energy consumption of our proposed EECRP, FCMMFO, FBCPSO, EGRC, LEACH-ERE, and LEACH accounts for 26.5%, 28.6%, 33.4%, 36.2%, 51.2%, and 63.6% of the initial energy of the whole network, respectively. With respect to the situation where the network energy is exhausted, the EECRP has improved the energy efficiency by 46.5%, 8.2%, 18.8%, 26.7%, and 5.1% compared to the LEACH, FBCPSO, EGRC, LEACH-ERE, and FCMMFO, respectively. That is because the EECRP employs the BPNN to fuse the data and uses the optimal multi-hop paths for data transmission, thus minimizing the energy consumption.

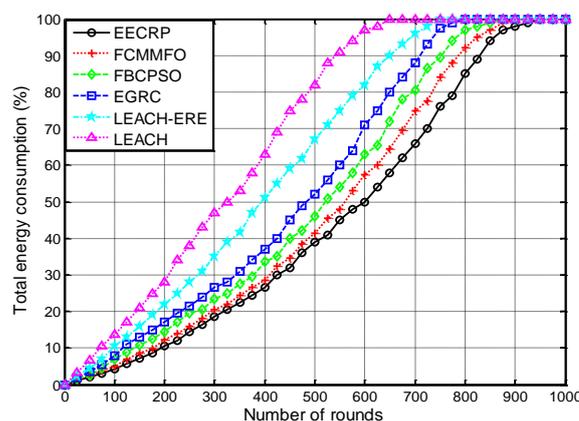


Figure 11. The total energy consumption of the network versus the number of rounds.

Moreover, Figure 12 illustrates the number of rounds when the energy of the whole network is completely consumed under the different number of network nodes, which verifies the influence of the different number of nodes on energy dissipation. With the decrease in the number of nodes, the distances between nodes increase, which consumes more energy for nodes to transmit data and thus shortens the network lifetime. However, the EECRP has the best performance among these protocols in all situations. For instance, when 250 nodes are considered in the network, the proposed EECRP protocol is 9.5%, 16.6%, 23.5%, 32.4%, and 67.1% more efficient than the FCMMFO, FBCPSO, EGRC, LEACH-ERE, and LEACH, respectively.

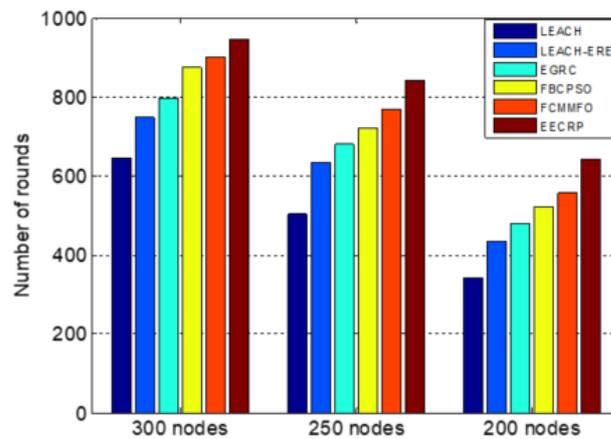


Figure 12. The number of rounds when the energy is exhausted under the different number of nodes.

7.3. The Packet Loss Rate

This section analyzes the performance of the network with 300 nodes and compares the six protocols by the packet loss rate, which is defined as the rate of the number of data packets sent by CHNs to the number of data packets received by the SN. The network load is defined as the number of data packets sent by every CHN per minute. Figure 13 illustrates the packet loss rate versus the network load for these six protocols, from which we conclude that the packet loss rate rises as the network load increases for these six protocols. However, the EECRP always has the lowest packet loss ratio. For example, when the network load is 3 packets per minute, the packet loss ratio of the EECRP, FCMMFO, FBCPSO, EGRC, LEACH-ERE, and LEACH is 16.8%, 18.1%, 19.8%, 21.6%, 24.6%, and 30.8%, respectively. The LEACH has approximately a 1.8 times higher packet loss rate than the EECRP does, which is because the EECRP employs the BPNN to fuse the data. Furthermore, it uses the improved GA to find the optimal multi-hop transmission paths, which is capable of reducing the risk of packet loss. Figure 14 displays the number of the packets that the SN receives versus the number of rounds for different protocols. The protocol is more effective when more packets are received by the SN. Apparently, the EECRP protocol is the most effective one, the efficiency of which is 86.7%, 18.1%, 31.3%, 46.9%, and 10.1% higher than that of the LEACH, FBCPSO, EGRC, LEACH-ERE, and FCMMFO, respectively, in round 1000.

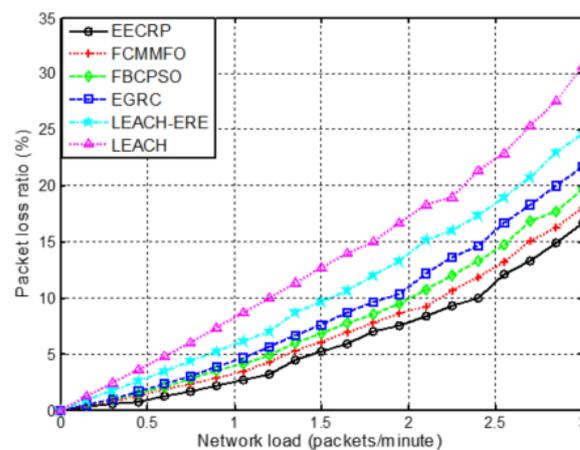


Figure 13. The packet loss ratio versus network load.

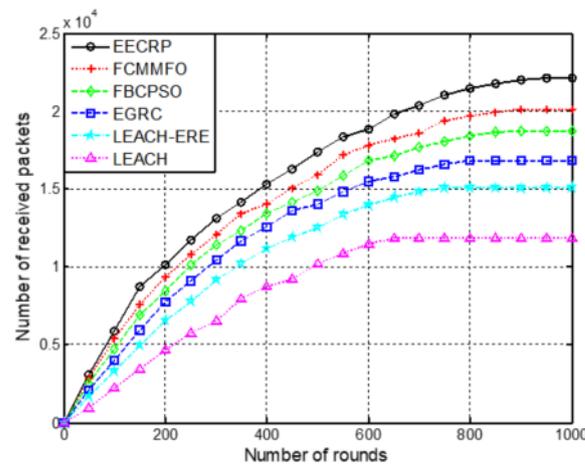


Figure 14. The number of received packets by the sink node (SN) versus the number of rounds for different protocols.

7.4. The Time Complexity

In this section, we analyze the time complexity of our proposed EECRP and its competitors, which is shown in Table 2. We can see that the time complexity of the LEACH and the LEACH-ERE is lower compared to other four algorithms. This is because they are older and more basic algorithms, and they are simpler and easier to be implemented. However, their performances in energy consumption are not as good as the newer algorithms that have been improved on the basis of the classic clustering approaches. The improved algorithms such as the EGRC, the FBCPSO, and the FCMMFO have the time complexity of $O(n^2)$. The time complexity of our proposed EECRP is the same as these three algorithms, but the EECRP has the best performance in reducing the energy consumption, prolonging the network lifecycle, and decreasing the packet loss rate. Moreover, in the EECRP, the training process of the BPNN and the process of the GA are accomplished by the SN as the SN has energy supplies. That can save the energy of nodes and extend the lifecycle of UWSNs. Therefore, our proposed EECRP possesses a high value and a wide prospect of applications in UWSNs. In addition, in the future research, we plan to lower the computational complexity of our protocol while keeping the energy-efficiency in UWSNs.

Table 2. The time complexity of six algorithms.

Algorithms	The Time Complexity
LEACH	$O(n)$
LEACH-ERE	$O(n)$
EGRC	$O(n^2)$
FBCPSO	$O(n^2)$
FCMMFO	$O(n^2)$
EECRP	$O(n^2)$

8. Conclusions

Due to the energy limitation of the underwater sensor nodes, we introduced an energy-efficient clustering routing protocol on the basis of the GA and the data fusion for UWSNs. The contributions were as follows. Firstly, this paper proposed the modified GA by proposing the new encoding scheme, the particular crossover operation, as well as the improved mutation operation. Secondly, this paper provided the improved BPNN by the developed momentum method to adjust the weights and biases, which is used by the CHNs to fuse the data in order to reduce energy consumption during data transmissions. Thirdly, the CHN selection operation was optimized, and the cluster formation process was improved. Finally, the experiments verified the effectiveness of our proposed EECRP in improving the network performance, and especially, the EECRP has improved the energy

efficiency by 46.5%, 26.7%, 18.8%, 8.2%, and 5.1% compared to the LEACH, LEACH-ERE, EGRC, FBCPSO, and FCMMFO, respectively.

However, this work focuses on the simulation experiment rather than the real implementation. The explanation is that the simulation experiment is our first step of the evaluation of our proposed EECRP. The sea experiment, which is extremely complicated and expensive to perform, is our following work. We have already done some small-scale sea experiments, which are the solid foundations of large-scale sea experiments where the EECRP can be conducted. In the real implementation, lots of underwater sensor nodes and a ship on the sea surface are needed. The nodes are equipped with the sensors to sense and acquire information, the battery to provide energy, the memory device to store data, the processor to achieve controlling and processing functions, the acoustic modem to achieve underwater wireless acoustic communications, the power amplifier, the waterproof device and so on. In terms of processing, the nodes should be high-speed, stable, and energy-saving. In memory, they need to have the large storage capacity and ensure that no data are lost after the death of nodes. As for the underwater wireless communication technology, we are trying to achieve low latency, low error rate, and long-distance communications. In addition, the nodes can provide functions like data acquisition, data storage, data processing, and data transmission and reception through underwater wireless acoustic communications. The ship acts as the SN and gathers information from the nodes. What is more, because the data transmissions between CHNs and the SN consume lots of energy, we plan to utilize autonomous underwater vehicles to get close to the CHNs and gather data from them, which further saves the energy of nodes and prolongs the lifecycle of UWSNs.

Author Contributions: Conceptualization, X.X. and H.H.; methodology, W.W.; software, X.X.; validation, X.X., H.H. and W.W.; formal analysis, W.W.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X. and W.W.; visualization, X.X.; supervision, H.H.; funding acquisition, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China, grant number 2018YFC1405904.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to privacy.

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

References

1. 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](#)]
2. Jouhari, M.; Ibrahim, K.; Tembine, H.; Ben-Othman, J. Underwater wireless sensor networks: A survey on enabling technologies, localization protocols, and internet of underwater things. *IEEE Access* **2019**, *7*, 96879–96899. [[CrossRef](#)]
3. Zanj, E.; Gambi, E.; Zanj, B.; Disha, D.; Kola, N. Underwater wireless sensor networks: Estimation of acoustic channel in shallow water. *Appl. Sci.* **2020**, *10*, 6393. [[CrossRef](#)]
4. Yu, W.; Chen, Y.; Wan, L.; Zhang, X.; Zhu, P.; Xu, X. An energy optimization clustering scheme for multi-hop underwater acoustic cooperative sensor networks. *IEEE Access* **2020**, *8*, 89171–89184. [[CrossRef](#)]
5. Hou, R.; He, L.; Hu, S.; Luo, J. Energy-balanced unequal layering clustering in underwater acoustic sensor networks. *IEEE Access* **2018**, *6*, 39685–39691. [[CrossRef](#)]
6. Ahmed, G.; Zhao, X.; Fareed, M.M.S.; Fareed, M.Z. An energy-efficient redundant transmission control clustering approach for underwater acoustic networks. *Sensors* **2019**, *19*, 4241. [[CrossRef](#)]
7. Durrani, M.Y.; Tariq, R.; Aadil, F.; Maqsood, M.; Nam, Y.; Muhammad, K. Adaptive node clustering technique for smart ocean under water sensor network (SOSNET). *Sensors* **2019**, *19*, 1145. [[CrossRef](#)]
8. Zhang, W.; Wang, J.; Han, G.; Zhang, X.; Feng, Y. A cluster sleep-wake scheduling algorithm based on 3D topology control in underwater sensor networks. *Sensors* **2019**, *19*, 156. [[CrossRef](#)]
9. Yahya, A.; Islam, S.U.; Zahid, M.; Ahmed, G.; Raza, M.; Pervaiz, H.; Yang, F. Cooperative routing for energy efficient underwater wireless sensor networks. *IEEE Access* **2019**, *7*, 141888–141899. [[CrossRef](#)]

10. Ayaz, M.; Baig, I.; Abdullah, A.; Faye, I. A survey on routing techniques in underwater wireless sensor networks. *J. Netw. Comput. Appl.* **2011**, *34*, 1908–1927. [[CrossRef](#)]
11. Bouabdallah, F.; Zidi, C.; Boutaba, R. Joint routing and energy management in underwater acoustic sensor networks. *IEEE Trans. Netw. Serv. Manag.* **2017**, *14*, 456–471. [[CrossRef](#)]
12. Zhou, Y.; Yang, H.; Hu, Y.; Kung, S. Cross-layer network lifetime maximization in underwater wireless sensor networks. *IEEE Syst. J.* **2020**, *14*, 220–231. [[CrossRef](#)]
13. Wang, Z.; Han, G.; Qin, H.; Zhang, S.; Sui, Y. An energy-aware and void-avoidable routing protocol for underwater sensor networks. *IEEE Access* **2018**, *6*, 7792–7801. [[CrossRef](#)]
14. Xu, Y.; Yue, Z.; Lv, L. Clustering routing algorithm and simulation of internet of things perception layer based on energy balance. *IEEE Access* **2019**, *7*, 145667–145676. [[CrossRef](#)]
15. Wang, C.; Zhang, Y.; Wang, X.; Zhang, Z. Hybrid multihop partition-based clustering routing protocol for WSNs. *IEEE Sens. Lett.* **2018**, *2*, 1–4. [[CrossRef](#)]
16. Lee, J.S.; Teng, C.L. An enhanced hierarchical clustering approach for mobile sensor networks using fuzzy inference systems. *IEEE Internet Things J.* **2017**, *4*, 1095–1103. [[CrossRef](#)]
17. He, W. Energy-saving algorithm and simulation of wireless sensor networks based on clustering routing protocol. *IEEE Access* **2019**, *7*, 172505–172514. [[CrossRef](#)]
18. Yin, Q.; Liu, M.; Cheng, J.; Ke, Y.; Chen, X. Mapping paddy rice planting area in northeastern china using spatiotemporal data fusion and phenology-based method. *Remote Sens.* **2019**, *11*, 1699. [[CrossRef](#)]
19. Xiao, L.; Xu, M.; Chen, Y.; Chen, Y. Hybrid grey wolf optimization nonlinear model predictive control for aircraft engines based on an elastic BP neural network. *Appl. Sci.* **2019**, *9*, 1254. [[CrossRef](#)]
20. Sun, L.; Cai, W.; Huang, X. Data Aggregation Scheme Using Neural Networks in Wireless Sensor Networks. In Proceedings of the 2010 2nd International Conference on Future Computer and Communication, Wuhan, China, 21–24 May 2010; pp. V1-725–V1-729. [[CrossRef](#)]
21. Cao, L.; Cai, Y.; Yue, Y.; Cai, S.; Hang, B. A novel data fusion strategy based on extreme learning machine optimized by bat algorithm for mobile heterogeneous wireless sensor networks. *IEEE Access* **2020**, *8*, 16057–16072. [[CrossRef](#)]
22. Xing, G.; Chen, Y.; He, L.; Su, W.; Hou, R.; Li, W.; Zhang, C.; Chen, X. Energy consumption in relay underwater acoustic sensor networks for NDN. *IEEE Access* **2019**, *7*, 42694–42702. [[CrossRef](#)]
23. Lin, N.; Shi, Y.; Zhang, T.; Wang, X. An effective order-aware hybrid genetic algorithm for capacitated vehicle routing problems in internet of things. *IEEE Access* **2019**, *7*, 86102–86114. [[CrossRef](#)]
24. Petres, C.; Pailhas, Y.; Patron, P.; Petillot, Y.; Evans, J.; Lane, D. Path planning for autonomous underwater vehicles. *IEEE Trans. Robot.* **2007**, *23*, 331–341. [[CrossRef](#)]
25. Wang, S.; Wu, Y. A Genetic Algorithm for Energy Minimization Vehicle Routing Problem. In Proceedings of the 2017 International Conference on Service Systems and Service Management, Dalian, China, 16–18 June 2017; pp. 1–5. [[CrossRef](#)]
26. Cao, J.; Li, Y.; Zhao, S.; Bi, X. Genetic-Algorithm-Based Global Path Planning for AUV. In Proceedings of the 2016 9th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 10–11 December 2016; pp. 79–82. [[CrossRef](#)]
27. Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 7 January 2000; pp. 3005–3014. [[CrossRef](#)]
28. Lee, J.; Cheng, W. Fuzzy-logic-based clustering approach for wireless sensor networks using energy predication. *IEEE Sens. J.* **2012**, *12*, 2891–2897. [[CrossRef](#)]
29. Mohapatra, H.; Rath, A.K. Fault tolerance in WSN through PE-LEACH protocol. *IET Wire. Sens. Sys.* **2019**, *9*, 358–365. [[CrossRef](#)]
30. Wang, K.; Gao, H.; Xu, X.; Jiang, J.; Yue, D. An energy-efficient reliable data transmission scheme for complex environmental monitoring in underwater acoustic sensor networks. *IEEE Sens. J.* **2016**, *16*, 4051–4062. [[CrossRef](#)]
31. Fei, W.; Hexiang, B.; Deyu, L.; Jianjun, W. Energy-efficient clustering algorithm in underwater sensor networks based on fuzzy c means and moth-flame optimization method. *IEEE Access* **2020**, *8*, 97474–97484. [[CrossRef](#)]
32. Krishnaswamy, V.; Manvi, S.S. Fuzzy and PSO based clustering scheme in underwater acoustic sensor networks using energy and distance parameters. *Wireless Personal Commun.* **2019**, *108*, 1529–1546. [[CrossRef](#)]
33. Wang, S.; Nguyen, T.L.N.; Shin, Y. Energy-efficient clustering algorithm for magnetic induction-based underwater wireless sensor networks. *IEEE Access* **2019**, *7*, 5975–5983. [[CrossRef](#)]
34. Islam, T.; Lee, Y.K. A cluster based localization scheme with partition handling for mobile underwater acoustic sensor networks. *Sensors* **2019**, *19*, 1039. [[CrossRef](#)]
35. Wan, Z.; Liu, S.; Ni, W.; Xu, Z. An energy-efficient multi-level adaptive clustering routing algorithm for underwater wireless sensor networks. *Clust. Comput.* **2019**, *22*, 14651–14660. [[CrossRef](#)]
36. Bansal, R.; Maheshwari, S.; Awwal, P. Energy-Efficient Multilevel Clustering Protocol for Underwater Wireless Sensor Networks. In Proceedings of the 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 10–11 January 2019; pp. 107–113. [[CrossRef](#)]
37. Zou, Z.; Lin, X.; Sun, J. A Cluster-Based Adaptive Routing Algorithm for Underwater Acoustic Sensor Networks. In Proceedings of the 2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS), Chongqing, China, 6–8 December 2019; pp. 302–310. [[CrossRef](#)]

38. Cao, Y.; Zhang, L. Data Fusion of Heterogeneous Network Based on BP Neural Network and Improved SEP. In Proceedings of the 2017 9th International Conference on Advanced Infocomm Technology (ICAIT), Chengdu, China, 22–24 November 2017; pp. 138–142. [[CrossRef](#)]
39. Yue, Y.; Fan, H.; Li, J.; Qin, Q. Large-Scale Mobile Wireless Sensor Network Data Fusion Algorithm. In Proceedings of the 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, China, 12–14 March 2016; pp. 1–5. [[CrossRef](#)]
40. Goyal, N.; Dave, M.; Verma, A.K. Fuzzy Based Clustering and Aggregation Technique for Under Water Wireless Sensor Networks. In Proceedings of the 2014 International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 13–14 February 2014; pp. 1–5. [[CrossRef](#)]
41. Harb, H.; Makhoul, A.; Couturier, R. An enhanced k-means and ANOVA-based clustering approach for similarity aggregation in underwater wireless sensor networks. *IEEE Sens. J.* **2015**, *15*, 5483–5493. [[CrossRef](#)]
42. Wang, S.; Zhao, B.; Li, D.; Du, T. Data Fusion Algorithm of Wireless Sensor Based on Combination between Cluster Head Election Improvement and Neural Network. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 6386–6391. [[CrossRef](#)]
43. Gang, W.; Xiangyang, L.; Guangen, W.; Yong, G.; Simin, M. Research on Data Fusion Method Based on Rough Set Theory and BP Neural Network. In Proceedings of the 2020 International Conference on Computer Engineering and Application (ICCEA), Guangzhou, China, 18–20 March 2020; pp. 269–272. [[CrossRef](#)]
44. Lin, Z.; Keh, H.C.; Wu, R.; Roy, D.S. Joint Data Collection and Fusion Using Mobile Sink in Heterogeneous Wireless Sensor Networks. *IEEE Sens. J.* **2020**, *21*, 2364–2376. [[CrossRef](#)]
45. Lorenzo, B.; Glisic, S. Optimal routing and traffic scheduling for multihop cellular networks using genetic algorithm. *IEEE Trans. Mob. Comput.* **2013**, *12*, 2274–2288. [[CrossRef](#)]
46. Lu, T.; Zhu, J. Genetic algorithm for energy-efficient QoS multicast routing. *IEEE Commun. Lett.* **2013**, *17*, 31–34. [[CrossRef](#)]
47. Silva, E.; Guardieiro, P. An efficient genetic algorithm for anycast routing in delay/disruption tolerant networks. *IEEE Commun. Lett.* **2010**, *14*, 315–317. [[CrossRef](#)]
48. Al-Shalabi, M.; Anbar, M.; Wan, T.C.; Alqattan, Z. Energy efficient multi-hop path in wireless sensor networks using an enhanced genetic algorithm. *Inf. Sci.* **2019**, *500*, 259–273. [[CrossRef](#)]
49. ThamaraiKannan, N.; Kamalraj, S. Utilization of compact genetic algorithm for optimal shortest path selection to improve the throughput in mobile Ad-Hoc networks. *Clust. Comput.* **2019**, *22*, 3715–3726. [[CrossRef](#)]
50. Xin, J.; Zhong, J.; Yang, F.; Cui, Y.; Sheng, J. An improved genetic algorithm for path-planning of unmanned surface vehicle. *Sensors* **2019**, *19*, 2640. [[CrossRef](#)]
51. Han, G.; Jiang, J.; Shu, L.; Xu, Y.; Wang, F. Localization algorithms of underwater wireless sensor networks: A survey. *Sensors* **2012**, *12*, 2026–2061. [[CrossRef](#)]
52. Sozer, E.M.; Stojanovic, M.; Proakis, J.G. Underwater acoustic networks. *IEEE J. Ocean. Eng.* **2000**, *25*, 72–83. [[CrossRef](#)]