IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 11, 2022, accepted March 30, 2022, date of publication April 5, 2022, date of current version April 15, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3165046

Reinforcement Learning-Based Trajectory Optimization for Data Muling With Underwater Mobile Nodes

QIANG FU[1], AIJUN SONG[1], (Member, IEEE), FUMING ZHANG[2], (Senior Member, IEEE), AND MIAO PAN[3], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487, USA
[2]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
[3]Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204, USA

Corresponding author: Aijun Song (song@eng.ua.edu)

This work was supported in part by the National Science Foundation (NSF) under Grant CNS 1801861, Grant 1801925, Grant 1828678, Grant 2016726, and Grant 2048188.

ABSTRACT This paper addresses trajectory optimization problems for underwater data muling with mobile nodes. In the underwater data muling scenario, multiple autonomous underwater vehicles (AUVs) sample a mission area, and autonomous surface vehicles (ASVs) visit the navigating AUVs to retrieve the collected data. The optimization objectives are to simultaneously maximize fairness in data transmissions and minimize the travel distance of the surface nodes. We propose an nearest-K reinforcement learning algorithm, which chooses only from the nearest-K AUVs as candidates for the next node for data transmissions. We use the distance between AUVs and the ASV as the state, selected AUVs as the action. A reward is designed as the function of both the data volume transmitted and the ASV travel distance. In the scenario with multiple ASVs, an AUV association strategy is presented to support the use of multiple surface nodes. We conduct computer simulations for performance evaluation. The effects from the number of AUVs, the size of the mission area, and the state number are investigated. The simulation results show that the proposed algorithm outperforms traditional methods in terms of the fairness and ASV travel distance.

INDEX TERMS Underwater acoustic communications, data muling, autonomous underwater vehicles, autonomous surface vehicles, trajectory optimization, reinforcement learning.

I. INTRODUCTION

Mobile platforms, such as autonomous underwater vehicles (AUVs) and autonomous surface vehicles (ASVs), have been used as effective tools in ocean monitoring and exploration [1]. Compared with fixed ocean monitoring networks, AUVs and ASVs have clear advantages in terms of operational costs and mission flexibility [2]. In this paper, we mainly focus on the underwater data muling using these mobile nodes, AUVs and ASVs.

Underwater data muling with mobile nodes promises a new way to achieve data collection in oceans [3]. In this scenario, a mission area is divided into several sub-missions, which are assigned to individual AUVs. Surface vehicles, ASVs, ferry data from AUVs using underwater acoustic communications and then transfer the data to the control center using terrestrial wireless communications. Due to the limited communication range of underwater acoustic communications, it is impossible for an ASV to cover a large geographic area. The surface node needs to visit each AUV so that these AUVs can obtain a reasonable amount of time for data transmission.

The use of ASVs to retrieve the AUV data provides several benefits. First, a low latency can be achieved. AUV measurements can be accessed before the vehicle recovery. Second, the AUV energy is conserved since AUVs do not need to be on the surface for communication with the control center. AUVs can operate underwater for longer periods of time. Third, multiple surface vehicles provide the possibility of expediting data transmission.

The trajectory of ASVs needs to be optimized to minimize the travel distance. Due to limited energy, ASVs are required to select the shortest route to approach each AUV. In this way, energy efficiency is achieved and the ASV mission time is extended. Access fairness among AUVs needs to be ensured too. Often AUVs are expected to transmit equal data volumes. Unfairness among users can cause large package delay [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Miaohui Wang.

This requires the surface vehicle to balance the transmitted data volume among all AUVs.

The requirements of energy-saving and fairness impose a trade-off in ASVs trajectory planning. On one hand, for limited energy resources, ASVs need to design a visiting sequence in an energy-efficient manner. Therefore, ASVs prefer to move as little as possible to save energy. On the other hand, for the low-latency requirement, ASVs should visit each AUV and fetch an equal amount of data among all AUVs. It requires each AUV not only to be given the same communication time but also the same visiting frequency. It is undesirable to approach one AUV for a long time and leave the rest barely connected. In other words, ASVs are required to hop among AUVs.

Multiple applications of mobile nodes were reported in wireless networks [5]–[8]. In [5], mobile sinks were used to collect the data to bypass the hot-spot problem. In [6], an ant colony-based path determination algorithm was proposed for wireless sensor networks. In [7], instead of visiting each sink node, the mobile sink visited each rendezvous point. One common characteristic of these wireless applications was that sensor node locations remained stationary. Therefore, the developed solutions were not applicable to the problem of interest in this paper.

One related application of underwater mobile platforms was the AUV-aided data muling in underwater acoustic sensor networks [9]–[14]. The network consisted of a variable number of fixed sensors that are deployed to perform collaborative data collection over a wide area. The AUV-aided data muling was proposed to extend the lifetime of fixed sensors [12]. Survey data were collected by using one or multiple AUVs rather than being transmitted among fixed nodes.

Multiple schemes were proposed in the AUV-aided underwater data muling [13], [15]–[21]. Early research focused on hardware design and simple path planning algorithms [13]. Several energy-efficient protocols were proposed to address this problem. These protocols reduced the AUV energy used by either designing the trajectory of the AUV [15]–[19] or grouping the underwater acoustic sensors into several clusters [20], [21]. None of the efforts in the literature used the multiple AUVs cooperation or the mobile sensor problem in the data muling protocol.

The data muling was formulated as the traveling salesman problem (TSP) [22]. As one of the most widely studied optimization problems [23], the objective of the TSP was to find the shortest route from a list of cities. There were two types of algorithms: exact algorithms and heuristic algorithms [24]. However, exact algorithms were impractical for a large number of cities. Heuristic algorithms, such as the nearest neighbor algorithm and the ant colony algorithm, were able to find a sub-optimal solution for large-scale problems within a fair time expense and acceptable accuracy ($97\% - 98\%$) [25]. Multiple objectives TSP optimization was explored in [12], [26]. These research only considered the scenario with static nodes. Research on mobile TSP problems mainly assumed that the target moves along a straight line in a two-dimensional space [27]. Mobile TSP problems focus only on a single optimization objective, that is the shortest route [28], [29].

Recent successes in the application of the reinforcement learning to optimization problems created interest in many areas [30]. In the area of underwater wireless sensor networks (UWSNs), reinforcement learning-based methods were widely used in oceanographic data collection. Energy efficiency [31]–[37] and end-to-end delay [38], [39] were two main concerns in UWSNs protocol design. In [40], a reinforcement learning-based congestion-avoided routing (RCAR) protocol was designed to reduce the end-to-end delay and energy consumption. In [38], the Q-learning based energy-efficient and balanced data gathering routing protocol (QL-EEBDG) was proposed to enhance the network lifespan.

Unmanned aerial vehicles (UAVs) play a similar role in terrestrial communications to ASVs in underwater environments. Reinforcement learning-based algorithms were proposed for UAV-aided terrestrial communications [41]–[44]. Current applications of the reinforcement learning often focus on a single optimization objective, such as minimizing energy consumption, minimizing delay and maximizing coverage area. The roles of UAVs in the protocol, the characteristics of the communication channel, the speed of AUVs and ASVs are different. These algorithms cannot be applied to the underwater environment. The reinforcement learning-based trajectory optimization for ASVs has not been investigated yet.

In this paper, we propose a nearest-K reinforcement learning-based trajectory optimization for the data muling with underwater mobile nodes. The ASV tracks are optimized for the surface vehicles to visit AUVs in a certain sequence. We use the distance between ASVs and AUVs as states; and the selected AUVs as actions. We design the reward as a function of the ASV travel distance and the data volume of AUVs. In this way, we achieve a balanced optimization objective: maximizing fairness among AUVs and minimizing the ASV travel distance. The reinforcement learning algorithm is simplified by limiting only the nearest K AUVs as candidates, which are selected as the next target AUV to be approached. We also design a user association algorithm such that multiple AUVs can be assigned to different ASVs. Multiple ASVs can work together to serve a group of AUVs.

The major contributions of this paper are summarized as follows. First, we propose a nearest-K reinforcement learning-based trajectory optimization for data muling with a single ASV. Second, we design an AUVs association strategy for a multiple ASVs scenarios. Third, we demonstrate the performance advantage of the proposed algorithm in a multi-AUVs and ASVs cooperation scenario with four and eight AUVs. The proposed algorithm is able to design an optimized track for ASVs, achieve a balanced optimization objective: minimizing the ASV travel distance and maximizing the fairness.

The paper is organized as follows. In Section 2, we describe the underwater data muling scenario and related problem

statements. In Section 3, we introduce the nearest-K reinforcement learning-based trajectory optimization algorithm. In Section 4, we demonstrate the performance of the proposed algorithm. In Section 5, we provide concluding remarks.

II. SYSTEM MODEL AND PROBLEM STATEMENT

A. SCENARIO DESCRIPTION

We consider an oceanographic mission with multiple ASVs. The environmental characteristics of a mission area such as temperature, water depth, and salinity need to be collected. These measurements are required to send to a control center onshore. A group of AUVs and ASVs are operated to execute this mission, as illustrated in Fig. 1.

The mission is often divided into multiple sub-missions, which are carried out by individual AUVs. Each AUV is assigned a small area to explore. ASVs are used to collect the survey data from AUVs via acoustic communications. Those data are sent to the control center by terrestrial wireless communications. In this scenario, AUVs do not need to surface or to directly transfer the collected data to the control center. Instead, ASVs approach each AUV and collect data by using underwater acoustic communications and underwater optical communications. The AUV mission time underwater can be extended. If ASVs visit each AUV more frequently, the control center gets data with less delay.

We first consider a scenario with N_v AUVs supported by a single ASV. We define an episode as a period of time in which the ASV visits one or multiple AUVs for data transmission. Each episode has two phases: capturing and trailing phases. During the capturing phase, the ASV chases a selected AUV with a full speed, v_t, which is higher than the AUV speed. To maximize the data volume, the selected AUV sends data to the ASV in the capturing phase via acoustic communications. AUVs and ASVs use acoustic communication with a low data rates. The data rate changes based on communication distance. The AUV is considered captured when the ASV-AUV distance reduces to a threshold of, for example, 50 m. In the trailing phase, the ASV trails the AUV at a lower speed of v_c for data transmissions. The ASV and ASVs use optical communications with a high data rate. The ASV can communicate with multiple AUVs to fully utilize the communication resources within the operating range of optical communications.

Different from the application in [5], the scenario in this paper only includes moving nodes such as AUVs and ASVs. The AUVs in our scenario move along a designed track to collect data for a target area. Because deploying these AUVs at different depths does not bring additional benefits, we assume that AUVs are deployed at the same depth in the ocean.

B. PROBLEM FORMULATION

We define the equivalent data volume as:

$$B_i'(k) = \omega_i(k) \times b_i(k), \tag{1}$$

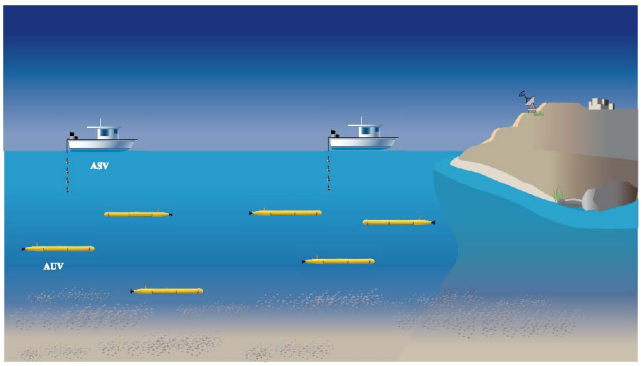


FIGURE 1. Data muling with the mobile nodes scenario. Multiple AUVs collect data from their assigned area. ASVs visit each AUV to ferry the data.

where b_i is successfully received bits from the i-th AUV in the k-th episode. We define the average data volume as the average value of the data volume in a time window length of T_c.

The average equivalent data volume of the i-th user can be calculated as:

$$B_i(k) = \begin{cases} \left(1 - \frac{1}{T_c}\right) B_i(k-1) + \frac{1}{T_c} B_i'(k) & i\text{-th user} \\ \left(1 - \frac{1}{T_c}\right) B_i(k-1) & \text{otherwise.} \end{cases} \tag{2}$$

In Eq. (2), the average equivalent data volume is initialized as one for all AUVs, $B_i(0) = 1$. A weighting factor $\omega_i(k)$ is defined to account for the user priority and fairness. It is the inverse of the equivalent transmitted data volume of the previous episode,

$$\omega_i(k) = \frac{1}{B_i(k-1)}. \tag{3}$$

The total equivalent data volume is the summation of the equivalent data volume from all AUVs:

$$B = \sum_{k=1}^{K} \sum_{i=1}^{N_v} \tau_i(k) B_i(k), \tag{4}$$

where N_v is the number of AUVs and $\tau_i(k)$ is an indicator function. When an AUV is selected at k-th episode, $\tau_i(k)$ is equal to 1, otherwise, $\tau_i(k)$ is equal to 0. Therefore,

$$\tau_i(k) = \begin{cases} 1 & i-\text{th AUV is selected} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The total ASV travel distance can be expressed as:

$$\begin{aligned} D &= \sum_{k=1}^{K} \sum_{i=1}^{N_v} \tau_i(k) (d_{c,i}(k) + d_{t,i}(k)) \\ &= \sum_{k=1}^{K} \sum_{i=1}^{N_v} \tau_i(k) D_i(k), \end{aligned} \tag{6}$$

where $d_{c,i}(k)$ is the distance that the ASV uses to capture a target AUV, $d_{t,i}(k)$ is the distance that the ASV trails with the AUV in k-th episode. In Eq. (6), $D_i(k) = d_{c,i}(k) + d_{t,i}(k)$.

We define a metric R below to represents that how much data volume can be transmitted when an ASV travels one unit of distance,

$$R = \frac{B}{D} = \sum_{k=1}^{K} \frac{\sum_{i=1}^{N_v} \tau_i(k) \times \omega_i(k) \times b_i(k)}{\sum_{i=1}^{N_v} \tau_i(k) \times D_i(k)} \tag{7}$$

The optimization objective is to minimize R, which is equivalent to maximizing the equivalent data volume while minimizing the travel distance. Therefore, the optimization objective becomes to select suitable AUVs at all episodes such that the value of R is minimized.

The data volume at the i-th AUV for all K episodes can be calculated as

$$b_i' = \sum_{k=1}^{K} \tau_i(k) b_i(k). \tag{8}$$

The total transmitted data volume from all AUVs during all K episodes is

$$b' = \sum_{k=1}^{K} \sum_{i=1}^{N_v} \tau_i(k) b_i(k). \tag{9}$$

C. ENERGY CONSUMPTION MODEL

In this paper, we focus on designing an ASV track with considerations of energy savings and communication connectivity. The energy consumption of the ASVs originates two aspects. The first part is related to vehicle movement. Let f_c and f_t be the resistant forces for the capture and trailing phase. The energy consumption for the ASV to capture and trail the i-th underwater node during the k-th episode is expressed as:

$$E_{m,i}(k) = f_c d_{c,i}(k) + f_t d_{t,i}(k), \tag{10}$$

which increases linearly with the travel distance. In the capturing phase, the ASV moves at a high speed to catch up with an AUV. We consider the $f_c = 300\ N$ for the speed of 10 kn based on [45]. In the trailing phase, the ASV moves at a lower speed. We consider $f_t = 50\ N$ at a speed of 2 kn and $f = 100$ N at a speed of 4 kn.

The other part of the energy consumption comes from the data transmission. We assume a constant power level for acoustic transmission, P_c. The associated energy consumption can be expressed as:

$$E_c = P_c \times T_c, \tag{11}$$

where the value of P_c is set up as 10 W.

Combining (10) and (11), the total energy consumption of the ASV is:

$$E = \sum_{k=1}^{K} \sum_{i=1}^{N_v} (\tau_i(k) E_{m,i}(k) + E_c). \tag{12}$$

III. Q-LEARNING BASED ALGORITHM FOR THE ASV TRAJECTORY OPTIMIZATION

A. PROPOSED ALGORITHM

We adopt the Q-learning algorithm to optimize the trajectory of the ASV. The Q-learning algorithm includes an agent, a set of state, actions, and a reward. The agent executes an action and gets a reward, while the state of the agent transits from one to another. The Q-learning algorithm selects actions to maximize the total reward. The following parameters need to be defined based on our scenario:

1) State: The state is defined as the distance that the ASV needs to move in the capturing and trailing phases. Therefore, the state in the k-th episode is defined as a vector $\mathbf{D}(k)$:

$$\mathbf{D}(k) = \begin{bmatrix} D_1(k) & D_2(k) & \cdots & D_{N_v}(k) \end{bmatrix}, \tag{13}$$

where $D_i(k)$ is the distance that the ASV needs to travel when the k-th AUV is selected. $D_i(k)$ is the distance that the ASV needs to travel when the i-th AUV is selected. The value $D_i(k)$ includes two parts, the ASV distance in the capturing phase ($d_{c,i}(k)$) and the distance in the trailing phase ($d_{t,i}(k)$).
The process of the ASV capturing AUVs is formulated as a differential equation. The location of the ASV is set as the coordinate origin. We assume the i-th AUV is located at $\mathbf{a}_i(k) = (a_{i,1}(k), a_{i,2}(k))$ in the k-th episode. The AUV moves along the y axis at the speed of v. At time Δ_t, this AUV arrives at $(a_{i,1}(k), a_{i,2}(k) + v\Delta_t)$. At the same time, the ASV moves to the point $(x(\Delta_t), y(\Delta_t))$, traveling at the direction towards to the AUV. Therefore, the differential equation model can be express as,

$$\frac{dy}{dx} = \frac{y(\Delta_t) - a_{i,2}(k) - v\Delta_t}{x(\Delta_t) - a_{i,1}(k)}. \tag{14}$$

The ASV travel distance in the capturing phase $d_{c,i}(k)$ is calculated by the ASV track $((x(t), y(t)))$ determined in this differential equation. The ASV travel distance in the trailing phase $d_{t,i}(k)$ is calculated by using data transmission and the ASV-AUV trailing speed.

2) Reward: The reward is defined as:

$$R' = \sum_{k=1}^{K} \gamma^k \frac{\sum_{i=1}^{N_v} \tau_i(k) \times \omega_i(k) \times b_i(k)}{\sum_{i=1}^{N_v} \tau_i(k) \times D_i(k)}, \tag{15}$$

where $0 < \gamma < 1$ is the discount rate.
To compare the proposed algorithm with single-objective optimizations, we define another reward function to minimize the ASV travel distance,

$$D' = \sum_{k=1}^{K} \gamma^k \sum_{i=1}^{N_v} \tau_i(k) \times (-D_i(k)). \tag{16}$$

3) Action: The ASV captures different AUVs to fetch the data in k-th episode. Therefore, the action here is defined as the selection of AUVs. Here we set the maximum number that the ASV can transmit data

simultaneously as two. Thus, the action candidates are $AUV_1, AUV_2, \ldots, AUV_{N_v}, AUV_1+AUV_2, \ldots, AUV_{N_v-1}+AUV_{N_v}$.

B. NEAREST-K MODIFICATION

To implement the Q-learning algorithm, we need to discretize the vector $\mathbf{D}(k)$ as the state. In the later implementation, the $D_i(k)$ values are discretized in the step size of $\Delta_d = 80$ m. Even with this quantization step, there are a large number of states when the number of AUVs is moderate number. With eight AUVs in a medium mission area, for example, 3 $km \times$ 3 km, the number of possible states is 38^8. This is impractical for implementation.

We solve this issue by integrating a nearest-K algorithm. Instead of using the distance between the ASV and all AUVs, we use the distance between the ASV and the nearest-K AUVs as the state:

$$\mathbf{D}(k) = \begin{bmatrix} D_1(k) & D_2(k) & \cdots & D_K(k) \end{bmatrix}. \tag{17}$$

The number of states can be drastically decreased for a small number for K.

C. USER ASSOCIATION FOR MULTIPLE ASVS

When there are a large number of AUVs, multiple ASVs can be used to reduce the access delay. We propose a geometry-based association algorithm to divide the AUVs into subgroups, each of which is associated with a single ASV. When there are N_v AUVs and M ASVs, the algorithm separates N_v AUVs into M sub-groups. AUVs with similar locations are assigned into the same group and associated with an ASV.

The N_v AUVs have their location vectors: $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N$ ($\mathbf{a}_i \in \mathbf{A}$). The M ASVs have the initial locations: $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M$ ($\mathbf{c}_i \in \mathbf{C}$). Each AUV is associated to an ASV based on the Euclidean distance $d(\mathbf{a}_i, \mathbf{c}_i)$,

$$\arg\min_{\mathbf{c}_i \in \mathbf{C}} d(\mathbf{c}_i, \mathbf{a}_i) = \arg\min_{\mathbf{c}_i \in \mathbf{C}} \sqrt{\sum_{j=1}^{2} \left(a_{i,j} - c_{i,j}\right)^2}. \tag{18}$$

The strategy of the association algorithm is to divide the AUVs into M sub-groups $\mathbf{S}_i$ with pre-determined group sizes, $N_{v_i} = |\mathbf{S}_i|$. Each sub-group has a centroid, which can be considered as the starting location of an surface vehicle. The AUVs in a particular sub-group have the shortest distance to its centroid. The association algorithm uses the following procedure:

1) Preset the sizes N_{v_i} for subgroups of AUVs so that $\sum_{i=1}^{M} N_{v_i} = N_v$. Randomly choose the locations of M AUVs as the centroids.
2) Calculate the Euclidean distances between each AUV and the M centroids.
3) Assign each AUV to the closest centroid using the Euclidean distance calculated in Step 2. If the closest centroid has associated an adequate number of AUVs N_{v_i}, assign this AUV to the second closest centroid.
4) Update the centroid location of each sub-group by

$$\mathbf{c}_i = \frac{1}{|\mathbf{S}_i|} \sum_{\mathbf{a}_i \in \mathbf{S}_i} \mathbf{a}_i. \tag{19}$$

5) Repeat Steps 2 to 4 until the centroids do not change. Those centroids are the initial locations of ASVs.

Algorithm 1: User Association

Data: number of ASVs M, size number of AUVs in each group N_{v_i}, the AUVs location vectors: $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{N_v}$ ($\mathbf{a}_i \in \mathbf{A}$)
Result: M ASVs locations: $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M$ ($\mathbf{c}_i \in \mathbf{C}$)

```
Initial c_1, c_2, ..., c_M (c_i ∈ C) at random;
while C has not converged do
    for i ← 1 to N_v do
        if |S| < N_{v_i} then
            l ← arg min_{c_i ∈ C} d(c_i, a_i);
            S_i = S_i ∪ {a_l};
        end
        else
            w ← arg min_{c_i ∈ C\{c_l}} d(c_i, a_i);
            S_i = S_i ∪ {a_w};
        end
    end
    c_i = (1/|S_i|) Σ_{a_i ∈ S_i} a_i;
end
```

D. NON-LEARNING BASED ASV TRACK PLANNING ALGORITHM

For comparison, we implement the ASV trajectory planning based on the non-learning nearest-K algorithm in the single ASV scenario. The surface vehicle selects the nearest K AUVs as the candidates for their next visits. The nearest-K algorithm is described as follows:

1) Initialization: Select an AUV randomly as the first target to be visited. This AUV is referred to the current stop.
2) At the k-th episode: calculate the distance $D_i(k)$, from the currently selected AUV, or the current stop, to unvisited AUVs.
3) Choose the nearest K AUVs to form an unvisited AUV pool based on the distance $D_i(k)$.
4) Select one, from the unvisited AUV pool, which has a minimum distance. Mark the selected AUVs as visited.
5) Repeat Steps 2 to 5 until all AUVs are marked as visited.
6) Mark all AUVs as unvisited. Repeat Steps 2 to 6.

In the single ASV scenario, the algorithm calculates the distance $D_i(k)$. The AUV with the shortest distance is chosen as the new target AUV, which is expressed in Eq. (20). In the multiple ASVs scenario, AUVs are associated to a ASV first.

The ASV visits the associated AUVs by using the nearest-K algorithm as below:

$$D = \sum_{k=1}^{K} \min\left(D_1(k), \quad D_2(k), \quad \cdots, \quad D_K(k)\right). \tag{20}$$

Algorithm 2: Nearest-K Algorithm

Data: number of ASV M, number of episodes K, N_v AUVs location vectors: $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N$ ($\mathbf{a}_i \in \mathbf{A}$)
Result: M ASVs locations: $w_1, w_2, \ldots, w_K$ ($w_i \in \mathbf{W}$)

Select the j-th AUV arbitrarily. Initial $\mathbf{S} = \{1, 2, \ldots, N_v\} \setminus \{j\}, k \leftarrow 1$;
while $k <= K$ *episodes* **do**
  while $\mathbf{S} \neq \varnothing$ **do**
    $l \leftarrow \arg\min_{l \in \mathbf{S}} d\left(\mathbf{a}_j, \mathbf{a}_l\right)$;
    $w_k \leftarrow l$;
    $\mathbf{S} \leftarrow \mathbf{S} \setminus \{l\}$;
    $j \leftarrow l$;
    $k++$;
  end
  Initial $\mathbf{S} = \{1, 2, \ldots, N_v\}$;
  Select the j-th AUV arbitrarily. Initial $\mathbf{S} = \{1, 2, \ldots, N_v\} \setminus \{j\}$;
  $k++$;
end

Note that we deal with a new application where the ASVs work with mobile platforms to perform data muling, which is different from wireless sensor networks [46].No learning-based algorithms are applicable to this new application. We use a non-learning-based algorithm for comparison. The complexity of the non-learning algorithm is lower than that of the proposed learning-based algorithm.

IV. SIMULATION AND RESULTS

We conducted computer simulations to evaluate the performance of the proposed algorithm. Standard datasets are critical for performance evaluations [47]. However, no standard datasets are available for our research problems. Therefore, to evaluate the algorithm performance, we created multiple scenarios with different mission areas and different numbers of AUVs.

The ASV-AUV communication data rates were simulated based on the transmitter-receiver distance. It was assumed that the product of the bit rate and distance remained constant. This constant number C was 5 *kbps* × *km*. In the capturing phase, the ASV speed v_t was equal to 10 *kn*. The data rate decreased with the increase of the ASV-AUV range. We assumed that the maximum communication distance was 1 *km* with the lowest bit rate of 5 *kbps*. In the trailing phase, the data rate was 200 *kbps* within the communication range of 50 *m*. The ASV used the same speed as the target AUVs, v_c. The discretization step size Δ_d was 80 *m*. We calculated the total ASV travel distance D for 100 episodes.

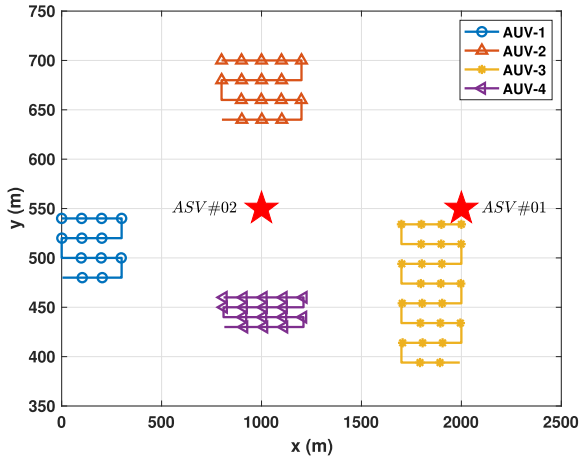


FIGURE 2. Scenario 1: small mission area, 2 *km* × 0.4 *km*, with four AUVs. Four AUVs used the lawn-mowing tracks with a track spacing of 20 *m*. Three vehicles, AUV-1, AUV-3 and AUV-4, had a higher speed, 2 *kn*, while the remaining one had a lower speed, AUV-2 at 1 *kn*.

TABLE 1. ASV travel distance D and Energy consumption E under the single-objective optimization (scenario 1).

	ASV travel distance (km)		Energy consumption (KJ)	
	Initial loc: $ASV\#01$	Initial loc: $ASV\#02$	Initial loc: $ASV\#01$	Initial loc: $ASV\#02$
Nearest-K	87.1	86.4	2.62×10^4	2.60×10^4
Q-learning	81.9	80.6	2.46×10^4	2.42×10^4

A. SCENARIO 1: SMALL MISSION AREA WITH FOUR AUVS

First, we considered a small mission area, 2 *km* by 0.4 *km*, with a small number of AUVs, as shown in Fig. 2. Four AUVs used the same lawnmower track pattern with a track space of 20 *m*. Three vehicles, AUV-1, AUV-3 and AUV-4, had a higher speed, 2 *kn*, while the remaining one had a lower speed, AUV-2 at 1 *kn*.

To validate the proposed algorithm, we used it to optimize for a single objective, minimizing the ASV travel distance D. The reward was set based on Eq. (16). The nearest-K algorithm was used for comparison.

Two initial ASV locations were tested. The results are presented in Table 1. With the initial location as $ASV\#01$, the ASV traveled 81.9 *km* when using the Q-learning algorithm. The ASV traveled 87.1 *km* when using the nearest-K algorithms. The ASV travel distance was shortened by 5.2 *km* when the proposed algorithm was used. Similar results were obtained for the initial location of $ASV\#02$. Two ASV initial locations generated similar ASV travel distance for each of the two algorithms. The proposed algorithm generated a shortened distance of 5.8 *km*. Correspondingly, the energy consumption decreased 5.1%. The results show that the proposed algorithm is effective in optimizing the track distance of the ASV. It performed better than the nearest-K algorithms.

Next, the Q-learning algorithm was tested with the combined objective of the ASV travel distance and data trans-

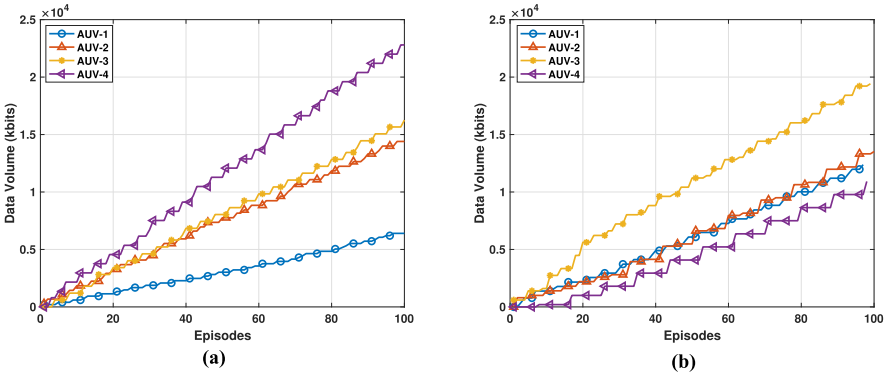


FIGURE 3. AUV data volume b'_i in Scenario 1 for two algorithms: (a) Nearest-K algorithm and (b) Q-learning algorithm. The Nearest-K algorithm used the optimization objective in Eq.(20). The Q-learning algorithm used the reward in Eq.(15).

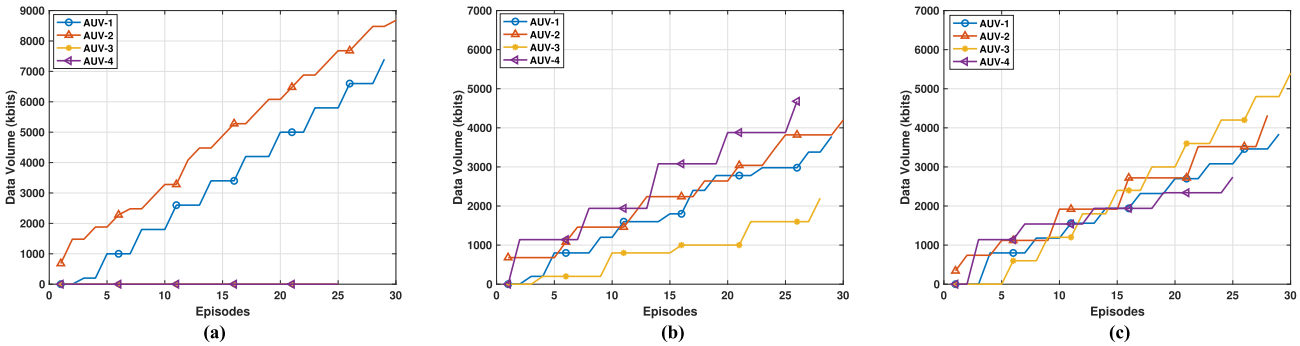


FIGURE 4. Effects of the range discretization. Three range steps were tested, including a) $\Delta_d = 400m$, (b) $\Delta_d = 90m$, and (c) $\Delta_d = 80m$.

TABLE 2. AUV data volume under the single-objective optimization (scenario 1).

Data volume ($kbits$)	AUV-1	AUV-2	AUV-3	AUV-4	Total
Nearest-K	6,400	16,260	14,400	22,800	5,9860
Q-learning	12,380	13,320	19,420	10,920	56,040

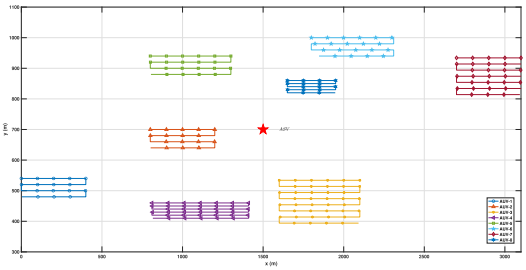


FIGURE 5. Scenario 2: Mission with eight AUVs. Eight AUVs had same lawnmower tracks with a track space of 20 m. AUV-1, AUV-3, AUV-4, AUV-6, and AUV-8 had the speed of 2 kn. AUV-2, AUV-5, and AUV-7 had the speed of 1 kn.

mission fairness, which was achieved based on the reward function Eq.(15). Fig. 3 shows the transmitted data volume for each AUV. In the nearest-K algorithms, the data volume of AUV-4 increased the fastest among all AUVs. The data volume of AUV-1 had the lowest increase. In comparison, the Q-learning algorithm generated balanced data volumes across four AUVs.

The two algorithms generated different ASV travel distances. The nearest-K algorithm produced an ASV travel distance of 26.4 km. The Q-learning algorithm led to an ASV travel distance of 30.1 km, 3.7 km increase. The proposed algorithm optimized for balanced data transmission. Each AUV had fair transmission opportunities. The ASV visited the AUV with a lower data volume more frequently to improve the fairness. Therefore the ASV travel distance was longer compared with the nearest-K algorithm. This confirms that the proposed algorithm was effective in implementing the combined objective of the ASV travel distance and data transmission fairness.

Table 2 shows individual AUV data volume b'_i. In the nearest-K algorithm, AUV-1 achieved the lowest data volume, 6,400 $kbits$. AUV-4 achieved the highest data volume, 22,800 $kbits$ which was 3.56 times the data volume of AUV-1. In the Q-learning algorithm, AUV-3 transmitted 19,420 $kbits$ data, which was the highest among four AUVs. AUV-4 achieved the lowest data volume, only 10,920 $kbits$. And AUV-3 transmitted 1.7 times more data than AUV-4.

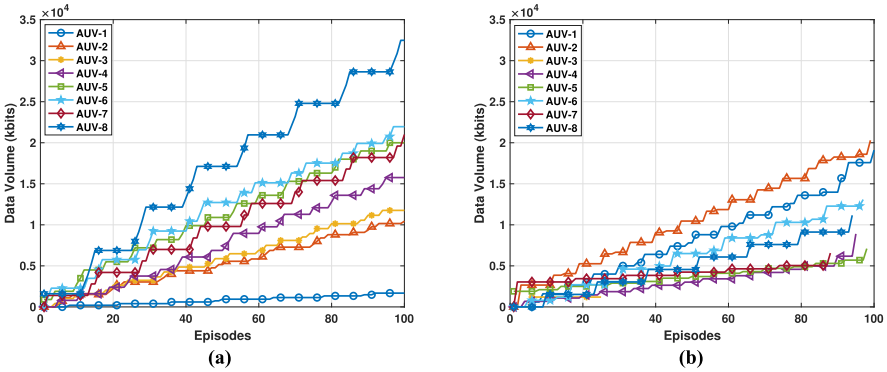


FIGURE 6. AUV data volume b'_i in Scenario 2 for two algorithms: (a) Nearest-K algorithm and (b) Q-learning algorithm. The Nearest-K algorithm use the optimization objective in Eq. (20). The Q-learning algorithm used the reward in Eq. (15).

Table 2 also shows the total data volume b' for all AUVs. In the nearest-K algorithm, the total data volume was 17,699 *kbits*. And the total data volume was 16,299 *kbits* in the Q-learning algorithm, 1,400 *kbits* lower than the nearest-K algorithms.

The results show that the proposed algorithm can maintain a more evenly data volume among AUVs. The AUV which has a longer distance has a lower data volume in the traditional algorithm. The proposed algorithm assigned more visiting opportunities to those AUVs so that they have more opportunities to transmit the data. Because of this, the total data volume decreased slightly compared to the traditional algorithm.

Next we investigated the effects of the distance discretization. Three step sizes, $\Delta_d = 400$, 90, and 80 m were tested. The results are shown in Fig. 4. When $\Delta_d = 400$ m, the Q-learning algorithm did not work properly. Only two AUVs were given opportunities to transmit data. With a smaller discretization step size, the performance of the algorithm improved. Fairness among AUVs was best ensured when the discretization step size was 80 m. In the following simulations, the discretization step size was setup to 80 m.

B. SCENARIO 2: LARGE MISSION AREA WITH EIGHT AUVS

In this subsection, we tested the proposed algorithm with a large group of eight AUVs in a relatively large mission area. Scenario 2 had a survey area of 3 *km* by 1 *km*, as illustrated in Fig. 5. The eight AUVs had same lawnmower tracks with a track space of 20 m. Five AUVs, AUV-1, AUV-3, AUV-4, AUV-6, and AUV-8, had the speed of 2 *kn*. The other three, AUV-2, AUV-5 and AUV-7, had the speed of 1 *kn*.

We first compared the AUV data volume b'_i as the episode count increased, as shown in Fig. 6. The Q-learning algorithm used the reward in Eq. (15). In the nearest-K algorithms, the data volume of AUV-8 increased fastest. The data volume of AUV-1 had the lowest increasing speed. In the Q-learning algorithm, the data volume of each AUV increased at a similar speed.

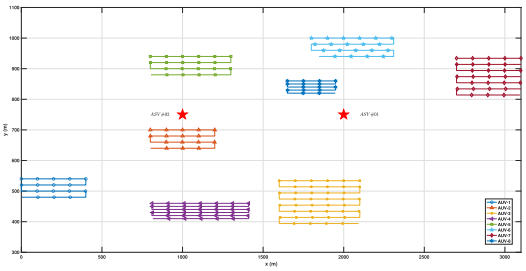


FIGURE 7. Scenario 3: Mission with eight AUVs. Eight AUVs had same lawnmower tracks with a track space of 20 *m*. AUV-1, AUV-3, AUV-4, AUV-6, and AUV-8 had the speed of 2 *kn*. AUV-2, AUV-5, and AUV-7 had the speed of 1 *kn*.

Table 3 shows the AUV data volume b'_i. In the nearest-K algorithms, AUV-1 achieved the lowest data volume, 1,680 *kbits*. AUV-8 achieved the highest data volume, 32,480 *kbits* which was 19.3 times more the data volume of AUV-1. In the Q-learning algorithm, AUV-6 transmitted 11,960 *kbits* data, which was the highest number among eight AUVs. AUV-4 achieved the lowest data volume, only 3,200 *kbits*. And AUV-3 transmitted 3.73 times the data volume of AUV-6.

Table 3 also shows the total data volume b'. The data volume of each AUV was added up together. In the nearest-K algorithm, the total data volume b' was 134,680 *kbits*. The total data volume was 84,038 *kbits* in the Q-learning algorithm, 37% lower than the nearest-K algorithm.

The ASV travel distance D differed between the nearest-K and Q-learning algorithms. The former led to a travel distance of 84.9 *km*. The latter 100.1 *km*. The travel distance increased 17% when using the Q-learning algorithm. The energy expense of nearest-K algorithm was 2.5×10^4 KJ. The Q-learning algorithm led to a higher energy expense, 3.0×10^4 KJ.

The results show that the proposed algorithm was effective for a larger group of eight AUVs. The ASV visited farther underwater nodes more often to achieve the fairness in the data volume. Therefore, the ASV traveled a longer distance than in the nearest-K algorithm. The nearest-K modification

TABLE 3. AUV data volume b'_i under the combined objective optimization (scenario 2).

Data volume ($kbits$)	AUV-1	AUV-2	AUV-3	AUV-4	AUV-5	AUV-6	AUV-7	AUV-8	Total
Nearest-K	1,680	10,440	11,760	15,760	19,600	21,960	21,000	32,480	134,680
Q-learning	21,260	11,100	3,200	14,740	7,099	11,960	5,760	8,919	84,038

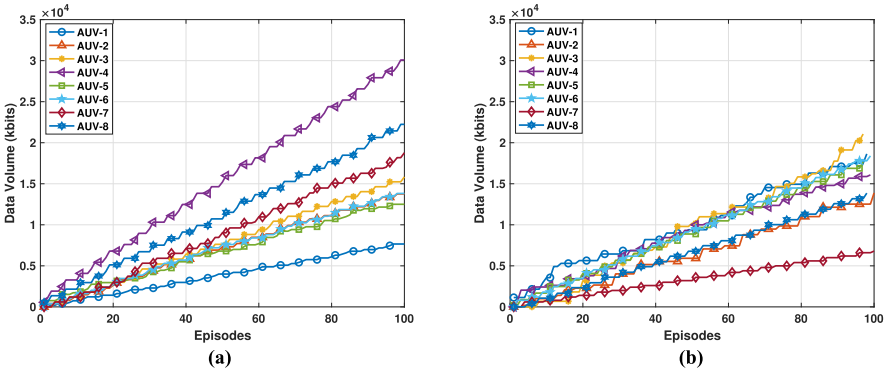


FIGURE 8. AUV data volume b'_i in Scenario 3 for two algorithms: (a) Nearest-K algorithm and (b) Q-learning algorithm. The Nearest-K algorithm use the optimization objective in Eq. (20). The Q-learning algorithm used the reward in Eq. (15).

TABLE 4. The AUV data volume b'_i under the combined objective optimization (scenario 3).

Data volume ($kbits$)	AUV-1	AUV-2	AUV-3	AUV-4	AUV-5	AUV-6	AUV-7	AUV-8	Total
Nearest-K-2-steps	7,660	13,820	15,840	30,080	12,500	13,820	18,780	22,240	134,740
Q-learning	18,640	13,880	21,060	16,080	17,680	18,400	6,799	13,860	126,399

decreased the number of states while not bringing visible negative impacts on the algorithm performance.

Compared to the performance in Scenario 1, the data volume became more uneven among the AUVs. The was because that the Q-learning algorithm only considered the nearest four AUVs in each episode to plan the track of the ASV. This increased the discrepancies in data volumes among the AUVs.

C. SCENARIO 3: LARGE MISSION AREA WITH TWO ASVS

In this subsection, we tested the proposed AUVs association strategy in the scenario of two ASVs and eight AUVs, as shown in Fig. 7. AUVs tracks and the optimization objective remained the same as Scenario 2. The node number, track spacing, and vehicle speed were kept the same with Scenario 2.

The AUV data volumes are shown in Fig. 8. In the nearest-K algorithms, the data volume of eight AUVs had a similar trend with that for a single ASV. In the Q-learning algorithm, the data volume of four AUVs increased more evenly than that for a single ASV.

Table 4 shows the AUV data volume b'_i and the total data volume b'. In the nearest-K algorithm, AUV-1 achieved the lowest data volume, 7,660 *kbits*. AUV-4 achieved the highest data volume, 30,080 *kbits*, which was 3.9 times the data volume of AUV-1. In the Q-learning algorithm, AUV-3 transmitted 21,060 *kbits* data, which was the highest number among eight AUVs. AUV-7 achieved the lowest data volume, only 6,799 *kbits*. And AUV-3 transmitted 3.09 times the data volume of AUV-6.

The total data volume b' of two algorithms showed a minor differences. The nearest-K algorithm led of total data volume of 134,740 *kbits*. In comparison, the Q-learning algorithm had 126,399 *kbits*, only 6.1% percent lower than the nearest-K algorithms.

The ASV travel distance D had a significant difference. The ASV travel distance was 73.4 *km* in the nearest-K algorithm. The two ASVs traveled 124.1 *km* in the Q-learning algorithm. The distance increased by 69% when using the Q-learning algorithm. The energy expense of the nearest-K algorithm is 2.1×10^4 KJ. The energy expense of the Q-learning algorithm was 3.7×10^4 KJ.

The ASVs visited AUVs with lower data volume more so that the data volume of each AUV increased more evenly. The ASV moved a longer distance than the nearest-K algorithm. This indicates that the proposed algorithm assigned more opportunities to the AUV which has a longer distance from the ASV. Same as before, the results show that each AUV had more fairly transmission opportunities in this scenario. Compared to the performance in one ASV scenarios, the AUV data volumes were more evenly distributed among the AUVs.

The usage of two ASVs brought two benefits with the Q-learning algorithm. First, the total data volume increased from 84,038 *kbits* to 126,399 *kbits*, which is a 50.4 % increase with one more ASV introduced. Second, the ratio between the

highest and lowest data volumes decreased from 3.73 to 3.09. This meant the data volumes among eight AUVs became more evenly. The introduction of multiple ASVs made each ASV serving fewer AUVs than in Scenario2. The AUVs had more communication opportunities.

V. CONCLUSION

In this paper, we proposed a nearest-K reinforcement learning-based trajectory optimization for underwater data muling with mobile nodes. The main idea was to use a reinforcement learning algorithm to optimize the track of ASVs. ASVs approached each AUVs along its track and collected data from the underwater nodes. This optimized track maximized the fairness among AUVs and simultaneously minimized the ASV travel distance. We simplified the reinforcement learning algorithm by limiting only the nearest-K AUVs as candidates when the learning algorithm calculated optimized ASV trajectories. We also designed a user association strategy, which supported multiple ASVs works together.

REFERENCES

[1] H. Zheng, N. Wang, and J. Wu, "Minimizing deep sea data collection delay with autonomous underwater vehicles," *J. Parallel Distrib. Comput.*, vol. 104, pp. 99–113, Jun. 2017.

[2] H. R. Kolar, J. Cronin, P. Hartswick, A. C. Sanderson, J. S. Bonner, L. Hotaling, R. F. Ambrosio, Z. Liu, M. L. Passow, and M. L. Reath, "Complex real-time environmental monitoring of the Hudson river and estuary system," *IBM J. Res. Develop.*, vol. 53, no. 3, p. 4, May 2009.

[3] J. Yuh, G. Marani, and D. R. Blidberg, "Applications of marine robotic vehicles," *Intell. Service Robot.*, vol. 4, no. 4, pp. 221–231, 2011.

[4] C. Wang, H.-K. Lo, and S.-H. Fang, "Fairness analysis of throughput and delay in WLAN environments with channel diversities," *EURASIP J. Wireless Commun. Netw.*, vol. 2011, no. 1, pp. 1–14, Jul. 2011.

[5] P. K. Donta, B. S. P. Rao, T. Amgoth, C. S. R. Annavarapu, and S. Swain, "Data collection and path determination strategies for mobile sink in 3D WSNs," *IEEE Sensors J.*, vol. 20, no. 4, pp. 2224–2233, Feb. 2020.

[6] D. P. Kumar, A. Tarachand, and C. S. R. Annavarapu, "Aco-based mobile sink path determination for wireless sensor networks under non-uniform data constraints," *Appl. Soft Comput.*, vol. 69, pp. 528–540, Aug. 2018.

[7] B. G. Gutam, P. K. Donta, C. S. R. Annavarapu, and Y.-C. Hu, "Optimal rendezvous points selection and mobile sink trajectory construction for data collection in WSNs," *J. Ambient Intell. Humanized Comput.*, Oct. 2021. [Online]. Available: https://link.springer.com/article/10.1007/s12652-021-03566-2#citeas, doi: 10.1007/s12652-021-03566-2.

[8] D. P. Kumar, A. Tarachand, and C. S. R. Annavarapu, "Machine learning algorithms for wireless sensor networks: A survey," *Inf. Fusion*, vol. 49, pp. 1–25, Sep. 2019.

[9] R. W. L. Coutinho and A. Boukerche, "Exploiting mobility to improve underwater sensor networks," in *Proc. 16th ACM Int. Symp. Mobility Manage. Wireless Access (MobiWac)*, 2018, pp. 89–94.

[10] E. Simetti and G. Casalino, "Manipulation and transportation with cooperative underwater vehicle manipulator systems," *IEEE J. Ocean. Eng.*, vol. 42, no. 4, pp. 782–799, Oct. 2017.

[11] N. Tsiogkas, V. De Carolis, and D. M. Lane, "Energy-constrained informative routing for AUVs," in *Proc. OCEANS-Shanghai*, Apr. 2016, pp. 1–5.

[12] K. Li, C.-C. Shen, and G. Chen, "Energy-constrained bi-objective data muling in underwater wireless sensor networks," in *Proc. 7th IEEE Int. Conf. Mobile Ad-Hoc Sensor Syst. (IEEE MASS)*, Nov. 2010, pp. 332–341.

[13] M. Dunbabin, P. Corke, I. Vasilescu, and D. Rus, "Data muling over underwater wireless sensor networks using an autonomous underwater vehicle," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2006, pp. 2091–2098.

[14] P. Gjanci, C. Petrioli, S. Basagni, C. A. Phillips, L. Bölöni, and D. Turgut, "Path finding for maximum value of information in multi-modal underwater wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 2, pp. 404–418, Feb. 2018.

[15] M. Doniec, "Autonomous underwater data muling using wireless optical communication and agile AUV control," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2013.

[16] H. Nam, "Data-gathering protocol-based AUV path-planning for long-duration cooperation in underwater acoustic sensor networks," *IEEE Sensors J.*, vol. 18, no. 21, pp. 8902–8912. Nov. 2018.

[17] G. Han, S. Shen, H. Wang, J. Jiang, and M. Guizani, "Prediction-based delay optimization data collection algorithm for underwater acoustic sensor networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6926–6936, Jul. 2019.

[18] S. Chen, Y. Chen, J. Zhu, and X. Xu, "Path-planning analysis of AUV-aided mobile data collection in UWA cooperative sensor networks," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Aug. 2020, pp. 1–5.

[19] F. B. Teixeira, N. Moreira, R. Campos, and M. Ricardo, "Data muling approach for long-range broadband underwater communications," in *Proc. Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2019, pp. 1–4.

[20] G. Han, Z. Zhou, T. Zhang, H. Wang, L. Liu, Y. Peng, and M. Guizani, "Ant-colony-based complete-coverage path-planning algorithm for underwater gliders in ocean areas with thermoclines," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8959–8971, Aug. 2020.

[21] X. Zhuo, M. Liu, Y. Wei, G. Yu, F. Qu, and R. Sun, "AUV-aided energy-efficient data collection in underwater acoustic sensor networks," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10010–10022, Oct. 2020.

[22] J. Mcmahon and E. Plaku, "Autonomous underwater vehicle mine countermeasures mission planning via the physical traveling salesman problem," in *Proc. OCEANS-MTS/IEEE Washington*, Oct. 2015, pp. 1–5.

[23] J. B. Robinson, *On the Hamiltonian Game (a Traveling-Salesman Problem)*. Santa Monica, CA, USA: RAND Corporation, 1949.

[24] G. Laporte, "The traveling salesman problem: An overview of exact and approximate algorithms," *Eur. J. Oper. Res.*, vol. 59, no. 2, pp. 231–247, Jun. 1992.

[25] J. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.

[26] Z. Wang, J. Guo, M. Zheng, and Y. Wang, "Uncertain multiobjective traveling salesman problem," *Eur. J. Oper. Res.*, vol. 241, no. 2, pp. 478–489, 2015.

[27] Q. Jiang, R. Sarker, and H. Abbass, "Tracking moving targets and the non-stationary traveling salesman problem," *Complex. Int.*, vol. 11, pp. 171–179, Nov. 2005.

[28] C. S. Helvig, G. Robins, and A. Zelikovsky, "The moving-target traveling salesman problem," *J. Algorithms*, vol. 49, no. 1, pp. 153–174, Oct. 2003.

[29] C. Gambella, J. Naoum-Sawaya, and B. Ghaddar, "The vehicle routing problem with floating targets: Formulation and solution approaches," *INFORMS J. Comput.*, vol. 30, no. 3, pp. 554–569, Aug. 2018.

[30] R. S. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[31] S. Wang and Y. Shin, "Efficient routing protocol based on reinforcement learning for magnetic induction underwater sensor networks," *IEEE Access*, vol. 7, pp. 82027–82037, 2019.

[32] X. Li, X. Hu, W. Li, and H. Hu, "A multi-agent reinforcement learning routing protocol for underwater optical sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–7.

[33] V. Di Valerio, F. L. Presti, C. Petrioli, L. Picari, D. Spaccini, and S. Basagni, "CARMA: Channel-aware reinforcement learning-based multi-path adaptive routing for underwater wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2634–2647, Nov. 2019.

[34] W. Xia, C. Di, H. Guo, and S. Li, "Reinforcement learning based stochastic shortest path finding in wireless sensor networks," *IEEE Access*, vol. 7, pp. 157807–157817, 2019.

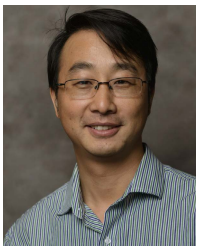
[35] K. G. Omeke, M. S. Mollel, L. Zhang, Q. H. Abbasi, and M. A. Imran, "Energy optimisation through path selection for underwater wireless sensor networks," in *Proc. Int. Conf. U.K.-China Emerg. Technol. (UCET)*, Aug. 2020, pp. 1–4.

[36] Y. Chen, J. Zhu, L. Wan, S. Huang, X. Zhang, and X. Xu, "ACOA-AFSA fusion dynamic coded cooperation routing for different scale multi-hop underwater acoustic sensor networks," *IEEE Access*, vol. 8, pp. 186773–186788, 2020.

[37] S. Han, L. Li, and X. Li, "Deep Q-network-based cooperative transmission joint strategy optimization algorithm for energy harvesting-powered underwater acoustic sensor networks," *Sensors*, vol. 20, no. 22, p. 6519, Nov. 2020.

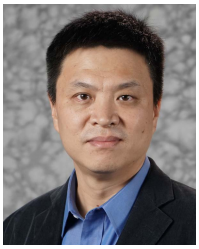
[38] N. Javaid, O. A. Karim, A. Sher, M. Imran, A. U. H. Yasar, and M. Guizani, "Q-learning for energy balancing and avoiding the void hole routing protocol in underwater sensor networks," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 702–706.
[39] X. Ye and L. Fu, "Deep reinforcement learning based MAC protocol for underwater acoustic networks," in *Proc. Int. Conf. Underwater Netw. Syst.*, Oct. 2019, pp. 1–5.
[40] Z. Jin, Q. Zhao, and Y. Su, "RCAR: A reinforcement-learning-based routing protocol for congestion-avoided underwater acoustic sensor networks," *IEEE Sensor J.*, vol. 19, no. 22, pp. 10881–10891, Nov. 2019.
[41] C. H. Liu, Z. Chen, J. Tang, J. Xu, and C. Piao, "Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2059–2070, Sep. 2018.
[42] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani, "Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges," *IEEE Access*, vol. 7, pp. 48572–48634, 2019.
[43] C. T. Cicek, H. Gultekin, B. Tavli, and H. Yanikomeroglu, "UAV base station location optimization for next generation wireless networks: Overview and future research directions," in *Proc. 1st Int. Conf. Unmanned Vehicle Syst.-Oman (UVS)*, Feb. 2019, pp. 1–6.
[44] A. Sharma, P. Vanjani, N. Paliwal, C. M. W. Basnayaka, D. N. K. Jayakody, H.-C. Wang, and P. Muthuchidambaranathan, "Communication and networking technologies for UAVs: A survey," *J. Netw. Comput. Appl.*, vol. 168, Oct. 2020, Art. no. 102739.
[45] J. Busquets, F. Zilic, C. Aron, and R. Manzoliz, "AUV and ASV in twinned navigation for long term multipurpose survey applications," in *Proc. MTS/IEEE OCEANS-Bergen*, Jun. 2013, pp. 1–10.
[46] P. K. Donta, T. Amgoth, and C. S. R. Annavarapu, "Delay-aware data fusion in duty-cycled wireless sensor networks: A Q-learning approach," *Sustain. Comput., Informat. Syst.*, vol. 33, Jan. 2022, Art. no. 100642.
[47] K. Dinesh, P. Donta, and T. Amgoth, "EDGF: Empirical dataset generation framework for wireless sensor networks," *Comput. Commun.*, vol. 180, pp. 48–56, Dec. 2021.

QIANG FU received the master's degree from the College of Ocean and Earth Science, Xiamen University, China, in 2016. He is currently pursuing the Ph.D. degree with The University of Alabama, Tuscaloosa, AL, USA. His research interests include underwater acoustic communications and signal processing.



AIJUN SONG (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Delaware, Newark, DE, USA, in 2005.

From 2005 to 2008, he was a Postdoctoral Research Associate with the College of Earth, Ocean, and Environment, University of Delaware, and also an Office of Naval Research Postdoctoral Fellow. He was a Research Professor at the University of Delaware, from 2008 to 2015. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL, USA. His current research interests include digital communications and signal processing techniques for radio-frequency and underwater acoustic channels, ocean acoustics, sensor networks, and ocean monitoring and exploration. He was a recent recipient of the NSF CAREER Award in 2021. He served as the General Co-Chair for the 2018 March NSF Workshop on Underwater Wireless Communications and Networking and the 2018 November NSF Workshop on Underwater Wireless Infrastructure. He was the General Co-Chair of the 14th International Conference on Underwater Networks & Systems.



FUMING ZHANG (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1995 and 1998, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 2004. From 2004 to 2007, he was a Lecturer and a Postdoctoral Research Associate with the Mechanical and Aerospace Engineering Department, Princeton University, Princeton, NJ, USA. In 2007, he joined the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, where he is currently a Professor. His research interests include marine autonomy, mobile sensor networks, and theoretical foundations for battery-supported cyber-physical systems. He received the National Science Foundation CAREER Award in 2009 and the Office of Naval Research Young Investigator Program Award in 2010.

MIAO PAN (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Dalian University of Technology, Dalian, China, in 2004, the M.A.Sc. degree in electrical and computer engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2007, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2012. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA. His research interests include wireless/AI for AI/wireless, deep learning privacy, cybersecurity, underwater communications and networking, and cyber-physical systems. He is a member of AAAI and ACM. He was a recipient of the NSF CAREER Award in 2014. His work won IEEE TCGCC Best Conference Paper Awards 2019, Best Paper Awards in ICC 2019, VTC 2018, GLOBECOM 2017, and GLOBECOM 2015. He has also been a Technical Organizing Committee for several conferences, such as a TPC Co-Chair for Mobiquitous 2019 and ACM WUWNet 2019. He is an Editor of the IEEE Open Journal of Vehicular Technology and an Associate Editor of the IEEE Internet of Things (IoT) Journal (Area 5: Artificial Intelligence for IoT). He was an Associate Editor of the IEEE Internet of Things (IoT) Journal (Area 4: Services, Applications, and Other Topics for IoT), from 2015 to 2018.