Multi-AUV Collaborative Data Collection in Integrated Underwater Acoustic Communication and Detection Networks

Xiaoxiao Zhuo*, Tianhao Hu$^{\star\diamond}$, Wen Wu†, Liang Tang$^{\star\boxtimes}$, Fengzhong Qu*, and Xuemin (Sherman) Shen‡
Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China*
University of Chinese Academy of Sciences, Beijing 101408, China$^{\diamond}$
Frontier Research Center, Peng Cheng Laboratory, Shenzhen 518055, China†
Key Laboratory of Ocean Observation-Imaging Testbed of Zhejiang Province, Zhejiang University, Zhoushan 316000, China*
Engineering Research Center of Oceanic Sensing Technology and Equipment, Ministry of Education, Zhoushan 316000, China*
Department of Electrical and Computer Engineering, University of Waterloo, Waterloo N2L 3G1, Canada‡
Email: {zhuoxx, liang.tang}@mail.sim.ac.cn*, hutianhao22@mails.ucas.ac.cn$^{\star\diamond}$, wuw02@pcl.ac.cn†,
jimqufz@zju.edu.cn*, sshen@uwaterloo.ca‡

Abstract—**In this paper, we propose the multi-autonomous underwater vehicle (AUV) collaborative data collection in integrated underwater acoustic communication and detection networks (UCDNs). Specifically, multiple AUVs collaboratively traverse the sensor nodes to collect data while detecting the environment to avoid obstacles along the trajectory. We first propose a time division multiple access (TDMA)-based packet transmission and active bistatic sonar detection strategy for UCDNs to transmit the sensor data and detect the unknown environment. Furthermore, we formulate the collaborative data collection problem as a mixed combinatorial and sequential quadratic optimization problem to minimize the trajectory length of multiple AUVs. To solve this problem, we decouple it into two subproblems, i.e., the node traversal subproblem and the trajectory planning subproblem. The former subproblem is converted into the multi-traveling salesman problem (MTSP), which is solved by the Q-learning-based algorithm to improve the robustness. The latter subproblem is optimally planning each AUV's trajectory while avoiding obstacles, which is solved by the soft actor-critic (SAC) algorithm to online make continuous trajectory decisions. Simulation results demonstrate that the proposed scheme outperforms benchmarks in terms of energy consumption and overall trajectory length.**

I. Introduction

With the development of oceanic exploration, underwater data collection is indispensable in many applications, such as underwater pollutant monitoring, earthquake detection, and tsunami warning [1]. As an important way of long-distance underwater wireless communications, underwater acoustic communication networks (UACNs) have become the main way of underwater data collection [2], [3]. Traditional UACNs are designed to collect data from sensor nodes to sink nodes with multi-hop networks, which suffer from several disadvantages, including unbalanced energy consumption and unreliable communication links [4]. To cope with these issues, multi-autonomous underwater vehicles (AUV) collaborative data collection has been envisioned as viable alternative because AUVs can cruise to sensor nodes closely to save transmission power of sensor nodes and improve communication reliability [5].

Current research on multi-AUV collaborative data collection focuses on two aspects, i.e., the trajectory planning for AUVs to access sensor nodes and the data transmission among nodes. Firstly, most research on AUV trajectory planning focuses on planning fixed or pre-determined trajectory that is calculated based on sensor nodes' positions [6]–[8]. Nonetheless, fixed trajectory planning for AUV is not reliable without considering the volatile environment. Furthermore, the learning-based trajectory planning scheme has been proposed to adapt to the oceanic environment dynamically [9]–[13]. However, to avoid collision with obstacles, AUVs have to detect the unknown environment, which consumes large energy and makes online decisions. Secondly, with respect to packet transmission, most research focuses on how to reduce the energy consumption of sensor nodes while disregarding the energy consumption of AUVs [7]–[9], [14]. Underwater acoustic communication and detection are indispensable in enabling AUVs to detect the environment and collect information whereas they are loaded onto AUVs separately, resulting in large space occupation and low energy efficiency. Since the hardware and signal processing techniques of communication and detection are similar, researchers are dedicated to integrating detection and communication, which can reduce energy consumption [15]–[17]. For data collection in the integrated underwater acoustic communication and detection networks (UCDNs), multiple AUVs transmit the integrated underwater acoustic communication and detection (UCD) signal to transmit packets and detect oceanic environments to avoid collision with obstacles.

Multi-AUV collaborative data collection strategy in UCDNs needs to meet the following requirements. Firstly, multiple AUVs need to cooperate with each other to evade unknown obstacles and avoid duplicate data collection, so the environ-

$^{\boxtimes}$Liang Tang (liang.tang@mail.sim.ac.cn) is the corresponding author of this paper.

mental detection information and collection status need to be instantaneously shared. Secondly, the trajectory planning decisions of AUVs must be made in real-time to avoid obstacles based on the detection information.

In this paper, we propose a multi-AUV collaborative data collection scheme in UCDNs. We first propose a centralized network architecture for UCDNs to control information sharing among multiple AUVs. In this architecture, the controller can gather detection information and collection status from AUVs, based on which the controller can make decisions on data collection and trajectory planning for AUVs. Secondly, we propose the time division multiple access (TDMA) packet transmission strategy and the active bistatic/multi-static sonar detection strategy to transmit packets and detect the environment via UCD signals. Then, we formulate the collaborative data collection problem as a mixed combinatorial and sequential quadratic optimization problem to minimize the trajectory length of multiple AUVs. To solve this problem, we decouple it into two subproblems. The first subproblem is the node traversal problem to decide the traversal sequence of sensor nodes, which is converted into the multi-traveling salesman problem (MTSP) and solved by the Q-learning algorithm. The second subproblem is the online trajectory planning problem to reach the traversal node and avoid obstacles, which is solved by the soft actor-critic (SAC) algorithm to online make continuous trajectory decisions. Simulation results demonstrate that the proposed scheme outperforms the benchmarks in terms of energy consumption and overall trajectory length.

The contributions of this paper are summarized as follows:

1) We propose the communication and detection strategy in UCDNs via UCD signal to reduce energy consumption.
2) We decouple the collaborative data collection problem into two subproblems, i.e., the node traversal problem and the trajectory planning problem. We propose learning-based algorithms to plan the shortest trajectory length for AUVs while avoiding obstacles along the trajectory.

The remainder of this paper is organized as follows. The system model is given in Section II. Section III presents the problem formulation, followed by the proposed algorithm in Section IV. Simulation results are given in Section V, and the conclusion is drawn in Section VI.

II. System Model

A. Network Model

As shown in Fig. 1, the UCDNs for multi-AUV collaborative data collection is a centralized network, which consists of multiple AUVs in different clusters, a data center, several sensor nodes, and several control nodes.

- Sensor Nodes: Multiple sensor nodes equipped with sensing devices are randomly distributed in the area.
- AUVs: AUVs are equipped with integrated communication and detection modems to communicate among nodes and detect obstacles in the surrounding environment via UCD signals. Multiple AUVs are divided into different clusters under the control of a control node and are responsible for the allocated sensor nodes.

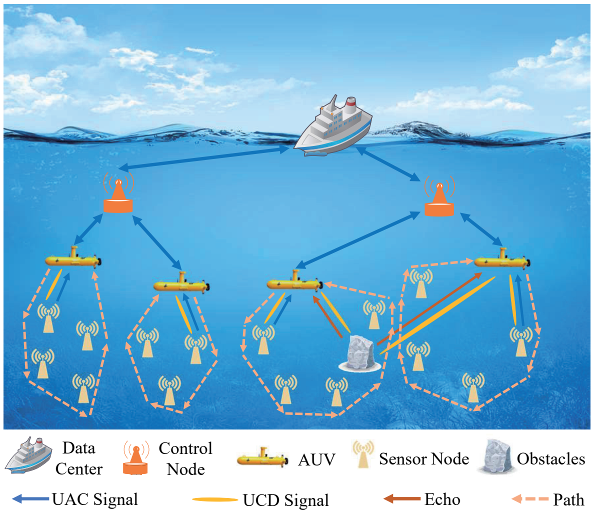


Fig. 1: Network model.

- Control Nodes: Underwater control nodes act as the cluster heads of AUVs. They are responsible for gathering data collection status and detection information from AUVs. The control nodes store the locations of the allocated sensor nodes to schedule the collection sequence of sensor nodes for AUVs. All decision-making processes of the AUV are handled by the control nodes, which are responsible for planning the trajectory for AUVs to collect data from sensor nodes and avoid obstacles.
- Data Center: The data center holds the locations of all sensor nodes to allocate the responsible area for each controller and its AUV cluster. It collects relevant information from all sensor nodes and control nodes to obtain the network's topology and overall state information.

B. *Integrated Communication and Detection in UCDNs*

In UCDNs, the UCD signals are used for both communication and detection to save energy. Therefore, we have designed an integrated communication and detection strategy in UCDNs. In this strategy, a TDMA-based packet transmission strategy is proposed to coordinate the packet transmission and echo signals receiving among sensor nodes, AUVs, control nodes, and obstacles. In addition, an active bistatic/multi-static detection scheme is proposed to detect the distance away from obstacles by analyzing the echo signals of UCD signals.

1) Communication Scheme: The communication scheme consists of three steps: Firstly, the controller broadcasts a packet periodically. Each AUV responds to the broadcast packets, based on which the controller can obtain a global network view and environment information. Based on the information, the controller can plan trajectories for each AUV to collect data from sensor nodes while avoiding collisions. Secondly, during the trajectory, each AUV sends the UCD signals periodically to search the sensor nodes and detect the environment. Finally, sensor nodes receive the UCD signals and upload data packets to AUVs. In a period of time, the AUVs and sensor nodes repeat the above steps until the next time round. Next, we

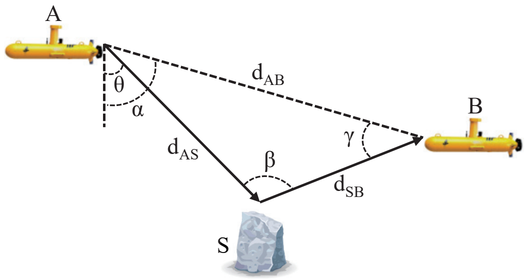


Fig. 2: An example of active bistatic detection in UCDNs.

will present the AUV-sensor communication and controller-AUV communication process in detail.

AUV-Sensor Communication: The AUV-sensor communication includes search packets that the AUV sends to sensor nodes and data packets that sensor nodes send to the AUV. Each AUV sends search packets periodically to request sensor nodes to upload data. These search packets also have detection functionality. Sensor nodes will reply response packets to upload the sensing data after receive search packets.

Controller-AUV Communication: The communication between the controller and AUV includes the downlink packets that the controller sends to the AUV and the uplink packets that the AUV sends to the controller. The controller issues instructions to the AUV and uploads information to the data center. When receiving the packets, each AUV demodulates the header, and if the target node is not itself, it ignores the subsequent content.

2) Detection Scheme: The detection process is based on active bistatic/multi-static sonar detection. In each time slot, AUVs periodically emit UCD packets to search for nearby sensor nodes. The packets can be reflected by obstacles. Other AUVs will receive these echo packets and demodulate the header of the packet to obtain the source node ID and transmission time of the packet. Based on this information, the AUV can determine the distance that the packet has traveled and send it to the controller. Then the controller can calculate the distance and angle of the obstacles related to each AUV, and thus determine the position of obstacles.

As shown in Fig. 2, an example of active bistatic detection is presented. If there exists an obstacle S, it will reflect the UCD signal sent by AUV A and produce an echo packet. Both AUV A and AUV B will receive the echo. The distance and angle of the obstacle are calculated by

$$\begin{cases} d_{AS}/\gamma = d_{SB}/(\alpha-\theta) = d_{AB}/\beta, \\ \gamma+\alpha+\beta-\theta=\pi. \end{cases} \tag{1}$$

Here, α, β, γ and θ are shown in Fig. 2. d_{AB} is the distance between AUV A and AUV B, which is monitored by the controller. d_{AS} and d_{SB} is the distance between AUV A, AUV B, and S, respectively, which can be calculated by AUV A and AUV B based on the echo signal, i.e.,

$$\begin{cases} d_{AS} = (t^r_A - t^s_A)v_s/2, \\ d_{SB} = (t^r_B - t^s_A)v_s - d_{AS}. \end{cases} \tag{2}$$

Here, t^r_A and t^r_B are the receiving time of the echo packet for AUV A and AUV B, respectively. t^s_A is the sending time of the UCD packet and v_s is the underwater acoustic speed.

Regarding the communication and detection strategy, the energy consumption for the AUV A_n can be calculated by

$$E_n = \sum_{t}^{\mathcal{T}_n}(P^A_n + P^{cd}_n + P^c_n)\Delta t. \tag{3}$$

Here, $\mathcal{T}_n$ is the data collection time for AUV A_n; P^A_n, P^{cd}_n, and P^c_n represent the power for AUV to move, detection and communication, and process information, respectively; Δt is the shortest unit of time in AUV operation. In traditional UACNs, P^{cd}_n is divided into the communication power and the detection power, which consumes more energy.

III. Problem Formulation

A. Original Optimization Problem

To collect data from K sensor nodes while avoiding obstacles efficiently, the following constraints should be considered.

1) Each AUV is released from the same starting point, and returns to the same point after collecting data, i.e.,

$$\sum_{n=1}^{N}\sum_{j=1}^{K} p^n_{0j} = \sum_{n=1}^{N}\sum_{i=1}^{K} p^n_{i0} = N, \tag{4}$$

where p^n_{ij} expresses that AUV A_n selects to traverse from sensor node k_i to sensor node k_j, $p^n_{0j}=1$ denotes that AUV is released from the starting point, and $p^n_{i0}=1$ denotes that AUV returns to the starting point.

2) Data from each sensor node is collected only once, i.e.,

$$\sum_{n=1}^{N}\sum_{i=1}^{K} p^n_{ij} = \sum_{n=1}^{N}\sum_{j=1}^{K} p^n_{ij} = 1, \ \forall\, i,j=1,...,K, \tag{5}$$

where $p^n_{ij}=1$ represents the trajectory selection of AUV A_n from node k_i to node k_j, and $p^n_{ij}=0$, otherwise.

3) The AUV should avoid all obstacles, i.e.,

$$d^0_{\min} - d(\mathbf{L}_n(t), \mathbf{S}_m) \leqslant 0, \ \forall\, t \in \mathcal{T}^n_{i,j}, n \in N, m \in M, \tag{6}$$

where $\mathbf{L}_n(t)$ represents the AUV A_n's locations at time t, $\mathbf{S}_m$ represents the location of obstacle S_m, $d(\mathbf{L}_n(t), \mathbf{S}_m)$ is the distance between the AUV A_n and the obstacle S_m, and $d^0_{\min}$ denotes the minimum safety distance between the AUV and the obstacles.

4) $D_n(t)$ indicates whether obstacles are avoided, i.e.

$$D_n(t) = \begin{cases} 0, & \text{if (6) holds}, \\ +\infty, & \text{otherwise}. \end{cases} \tag{7}$$

The trajectory length of AUV's each step is expressed by

$$l(n,t) = d(\mathbf{L}_n(t), \mathbf{L}_n(t-1)) + d(\mathbf{L}_n(t), \mathbf{S}_k)D_n(t). \tag{8}$$

The objective of the proposed scheme is to minimize the trajectory length of N AUV, which is expressed by

$$\mathbf{P}_0 : \min_{\mathbf{L}\in\mathcal{L}} \sum_{n=1}^{N}\sum_{t=1}^{\mathcal{T}^n} l(n,t) \tag{9a}$$

$$\text{s.t.} \quad (4),\ (5), \text{and } (6). \tag{9b}$$

In the above problem, $\mathcal{T}^n$ is time consumption for AUV A_n to traverse all allocated sensor nodes. The problem is a mixed combinatorial and sequential quadratic optimization problem.

B. Problem Decomposition

To solve this original optimization problem, we decouple it into the node traversal and the trajectory planning problem.

Subproblem 1: The first subproblem is to decide the traversal sequence of sensor nodes and can be modeled as the MTSP. The objectives of this subproblem are expressed as

$$\mathbf{P}_1: \min_{\mathcal{P}_n \in \mathcal{P}} \sum_{n=1}^{N} \sum_{(i,j)\in\mathcal{P}_n} d(k_i, k_j) p_{ij}^n \tag{10a}$$

$$\text{s.t.} \quad (4) \text{ and } (5). \tag{10b}$$

The objective function (10a) minimizes the linear distance that the AUVs collect data from all nodes, where $d(k_i, k_j)$ is the linear distance length between the sensor node k_i and k_j, $\mathcal{P}_n$ is the sequence of traversal sensor nodes for AUV A_n. To solve this combinatorial optimization problem, we propose the Q-learning-based node traversal algorithm.

Subproblem 2: The second subproblem is optimizing the optimal trajectory for AUV A_n to depart from the starting node to the destination node with the shortest trajectory and obstacle avoidance along the trajectory, which is defined as

$$\mathbf{P}_2: \min_{\mathbf{L}\in\mathcal{L},(i,j)\in\mathcal{P}_n} \sum_{t=1}^{\mathcal{T}_{i,j}^n} l(n,t) \tag{11a}$$

$$\text{s.t.} \quad (6) \text{ and } (7). \tag{11b}$$

Here, $\mathcal{T}_{i,j}^n$ is time consumption for A_n to move from node k_i to k_j. To solve this sequential quadratic optimization problem, we propose the SAC-based trajectory planning algorithm.

IV. Proposed Learning-Based Algorithm

A. Q-Learning-Based Node Traversal Algorithm

Due to the node traversal problem being NP-hard, the Q-learning-based algorithm is applied to solve this problem to avoid getting stuck in local optima and improve the robustness, aiming to collect data from all sensor nodes using the shortest traversal trajectory. Assuming there are K sensor nodes, we first set up a $K \times K$ Q-value table. The (s, a)-th element in the table represents the Q-value for selecting sensor node k_a as the next traversal node when starting from node k_s. The state refers to the current node that the AUV is traversal, and the action corresponds to selecting the next node. After the action, the AUV's state changes to the selected node. The reward of action is the negative value of the distance between the starting node and the target node. A larger total reward results in a shorter distance traveled by the AUV. The Q-value table is updated based on the Bellman equation [18], i.e.,

$$Q(s,a) = Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s,a)]. \tag{12}$$

The decision to select the next node is based on the generation of a random number. If the random number is less than Γ, a target node that has not been collected previously is chosen randomly. Otherwise, the next target node chosen will have the maximum Q-value from the current node to an uncollected node. The details of the Q-learning-based node traversal process are given in Algorithm 1.

B. SAC-Based Trajectory Planning Algorithm

The trajectory planning problem is challenging due to the characteristic of high-dimensional, continuous, and online. To solve this problem, we propose a SAC-based trajectory planning algorithm to online make continuous trajectory decisions. The state, action space, and reward are given as follows.

Algorithm 1 Q-learning based node traversal algorithm.

```
Initialize Q-values ($Q(s,a)$) arbitrarily;
for episode< $E$ do
  while All nodes are not fully traversed do
    for each AUV do
      if RandomNum< $\Gamma$ then
        Randomly select a node as the destination node of the nodes that have not been collected;
      else
        Select the node with the largest Q value in the Q value table as the destination node;
        Update status $s'$ and calculate rewards $r$;
        Update $Q(s,a)$ by following Eq. (12);
      end if
      if Different AUVs choose the same destination node then
        The AUV with a low Q value reselects the destination node;
      end if
    end for
  end while
end for
```

1) State Space: In addition to collecting information from each sensor node, AUVs need to achieve automatic obstacle avoidance in UCDNs. The state space of this system mainly consists of three parts: AUV's own state, the position of the traversal node, and the obstacle-related state. The AUV's own state includes the current position of the AUV A_n, i.e., $[x_{n,t}, y_{n,t}]^T$ and the communication connection index $O_{n,t}$. At the time t, when AUV A_n is within the controller's communication range, $O_{n,t} = 1$, and $O_{n,t} = 0$, otherwise. The position of the traversal node is $[x'_{n,t}, y'_{n,t}]^T$. The obstacle-related state contains the obstacle existence index $W_{n,t}$ for AUV A_n, the collision index $D_n(t)$, and the distance d^o and the motion angle ρ^o between AUV and obstacle. Finally, the state space $s_{n,t}$ of AUV A_n at time t can be given by $[O_{n,t}, x_{n,t}, y_{n,t}, x'_{n,t}, y'_{n,t}, W_{n,t}, D_n(t), d^o_{n,t}, \rho^o_{n,t}]$.

2) Action Space: The action of AUV can be determined by the velocities, which can be defined as $a_{n,t} = [v^x_{n,t}, v^y_{n,t}]$.

3) Reward: The reward consists of rewards for avoiding obstacles and approaching the traversal node. In addition, the penalty consists of the collision penalty, the disconnection penalty, and the timeout penalty. The reward for avoiding obstacles which is denoted by $r^o_{n,t} = W_{n,t}\omega_o(\overline{d^o_{n,t}} + \overline{\rho^o_{n,t}})$. where ω_o is a fixed value, $\overline{d^o_{n,t}}$ and $\overline{\rho^o_{n,t}}$ is the normalized value of $d^o_{n,t}$ and $\rho^o_{n,t}$. The reward for approaching the traversal node S is the normalized value of the reduction in the distance between the AUV and the traversal node. If there are no obstacles, the weight for it is 1. Otherwise, the weight is $1-\omega_o$, which is denoted by $r^a_{n,t} = (1 - W_{n,t}\omega_o)(\overline{\Delta d}(A_n, S))$. The collision penalty $p^1_{n,t}$ is 1 while $D_n(t) = +\infty$. The disconnection penalty $p^2_{n,t}$ is 1 while $O_{n,t} = 0$. The timeout penalty $p^3_{n,t}$ is 1 while $t > t_{\max}$. Finally, the reward for AUV

Algorithm 2 SAC-based trajectory planning algorithm.

```
Initialize $\pi^n(s_{n,t}|\mu^n)$ and $Q^n(s_{n,t},a_{n,t}|\theta^n)$;
And initialize $\pi^{n'}(s_{n,t}|\mu^{n'})$ and $Q^{n'}(s_{n,t},a_{n,t}|\theta^{n'})$;
for step $< K_n$ do
    Conduct Algorithm 1;
    for episode $< E$ do
        for $t < T$ do
            AUV $A_n$ executes $a_{n,t}$ by following $\pi^n(s_{n,t}|\mu^n)$ and gets reward $r_{n,t}$;
            Update new state $s_{n,t+1}$;
            Store $(s_{n,t},a_{n,t},r_{n,t},s_{n,t})$ in experience replay buffer and set $s_{n,t} \leftarrow s_{n,t+1}$;
            Get M random samples $(s_{n,t},a_{n,t},r_{n,t},s_{n,t})$ in experience and set target value $y_{n,t}$ by (14);
            Use gradient descent to update actor-network and critic-network based on (15) and (13);
            Update target critic and actor-network;
        end for
    end for
end for
```

A_n is calculated by $r_{n,t} = r^c + r^o_{n,t} - \omega_p(p^1_{n,t} + p^2_{n,t} + p^3_{n,t})$, where w_p is the weight value for the penalty.

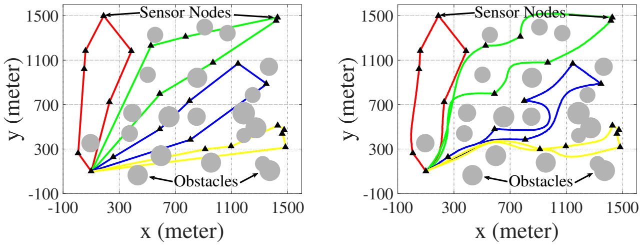
4) Training and Testing: This system employs the SAC algorithm where each AUV corresponds to four neural networks: actor network, target actor network, critic network, and target critic network. As this network is designed based on the proposed UCDNs, the controller can obtain relevant information on all AUVs in the cluster. Therefore, the neural networks of each AUV are loaded on the controller, each controller can iteratively update neural networks based on the information of all AUVs [19]. This architecture is known as centralized training decentralized execution (CTDE) architecture [20].

For AUV A_n, the algorithm is presented in detail as follows. Firstly, initialize the actor network $\pi^n(\cdot)$, target actor network $\pi^{n'}(\cdot)$, critic network $Q^n(\cdot)$, target critic network $Q^{n'}(\cdot)$, and their corresponding parameters μ^n, $\mu^{n'}$, θ^n, and $\theta^{n'}$. Set the initial values of $\mu^{n'}$ and $\theta^{n'}$ to the values of μ^n and θ^n, respectively. For each step, the Q-table trained by Algorithm 1 is used to determine the target node. Then, the actor network of AUV A_n makes a decision on the action $a_{n,t} = \pi^n(s_{n,t}|\mu^n)$ based on the state $s_{n,t}$. This results in a corresponding reward $r_{n,t}$ and an updated state $s_{n,t+1}$. The $(s_{n,t},a_{n,t},r_{n,t},s_{n,t+1})$ tuple is stored in the experience replay buffer. For every iteration of the neural network parameter update, M tuples are randomly sampled from the experience replay buffer to run the gradient descent algorithm.

The loss function for the critic network $Q^n(\cdot)$ used in the gradient descent is defined as

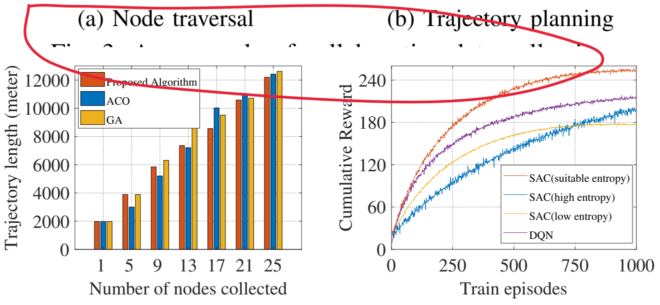
$$L(\theta^n) = \frac{1}{M}\sum_{j=1}^{M}\frac{(y_{n,j} - Q^n(s_{n,j},a_{n,j}|\mu^n))^2}{2}. \tag{13}$$

Here, y^n_j is the output of AUV A_n's target critic network combined with the reward, which is denoted by



(a) Node traversal (b) Trajectory planning

Fig. 3: An example of collaborative data collection



(a) Trajectory length. (b) Convergence curve.

Fig. 4: Comparison of algorithm performance.

$$\begin{aligned} y_{n,j} =& r_j + Q^{n'}(s_{n,j+1},a_{n,j+1}|\theta^{n'}) \\ & - \varepsilon\log(\pi^{n'}(a_{n,j+1}|s_{n,t};\mu^{n'})). \end{aligned} \tag{14}$$

The loss function for the actor network $\pi^n(\cdot)$ used in the gradient descent is defined as

$$\begin{aligned} L(\mu^n) =& E[\varepsilon\log(\pi^n(a_{n,j+1}|s_{n,t};\mu^n)] \\ & - Q^n(s_{n,j},a_{n,j}|\theta^n), \end{aligned} \tag{15}$$

where ε is the ratio parameter.

The parameters for the target actor network $\pi^{n'}(\cdot)$ and target critic network $Q^{n'}(\cdot)$ are updated by $\mu^{n'} = \sigma\mu^n + (1 - \sigma\mu^{n'})$ and $\theta^{n'} = \sigma\theta^n + (1 - \sigma\theta^{n'})$, where σ is the ratio hyper-parameter. The details of the SAC-based trajectory planning are given in Algorithm 2.

V. Performance Evaluation

A. Simulation Setup

In the simulation, 25 sensor nodes are randomly distributed, and 20 obstacles are randomly located in the area of 1,500 m by 1,500 m. Four AUVs jointly perform the data collection task. The power of moving, detection, communication, and UCD signal is set to 100 W, 30 W, 5 W, and 30 W, respectively. The obstacle warning distance d^w is 100 m.

B. Simulation Results

As shown in Fig. 3, the trajectory of the AUVs completing the data collection task is illustrated. As depicted in Fig. 3(a), the node traversal trajectory is generated based on the Q-learning algorithm. When the proposed SAC-based trajectory planning algorithm is employed, AUVs can complete their data collection tasks while avoiding obstacles as shown in Fig. 3(b).

As shown in Fig. 4, we present the performance comparison among the proposed Q-learning algorithm, genetic algorithm (GA), and ant colony optimization (ACO) in solving the node traversal problem. From Fig. 4(a), it can be seen that, as

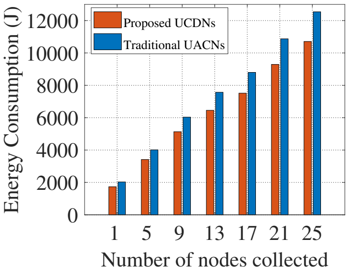


Fig. 5: Comparison of energy consumption.

compared to traditional heuristic algorithms, the proposed Q-learning is less likely to get trapped in locally optimal solutions, thus shortening the trajectory lengths. As shown in Fig. 4(b), we present a comparison of the results between the SAC-based algorithm with different parameter settings and the deep Q-network (DQN) algorithm in solving the trajectory planning problem. It can be seen that, if the amount of action entropy is too high, convergence is slower, and if the amount is too low, the exploration process is too short and the convergence is too fast, resulting in suboptimal trajectory planning. When SAC algorithm is properly configured, the convergence speed and the sum of the reward of the proposed algorithm perform 30% and 15% better than that of the DQN based algorithm.

As shown in Fig. 5, we compare the energy consumption of UCDNs and that of traditional UACNs. From Fig. 5, the energy consumption of the proposed UCDNs is 14.6% higher than that of the traditional UACNs. Different from the traditional UACNs, the transmission of detection and communication packets are integrated together to save energy in UCDNs. In addition, operations that require advanced processors, such as motion decisions and detection calculations, are offloaded to the controller. AUV only needs simple processors and integrated modules for detection and communication. Therefore, the energy consumption of AUVs in UCDNs can be reduced.

VI. Conclusion

In this paper, we have investigated the multi-AUV collaborative data collection problem in UCDNs. We have formulated the collaborative data collection problem as a mixed combinatorial and sequential quadratic optimization problem to minimize the trajectory length, which is solved by Q-learning and SAC algorithms to traverse all sensor nodes while avoiding obstacles. Simulation results show that the proposed scheme can achieve lower energy consumption, faster convergence speed, and shorter AUV trajectory length. For future work, we aim to jointly optimize the AUV trajectory planning and packet transmission scheme.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62225114 and 62201311; in part by XTCX-KJ-2022-2-02; in part by Key Research and Development Program of Zhejiang under Grant 2022C03027; in part by the Peng Cheng Laboratory Major Key Project under Grant PCL2021A09-B2.

References

[1] M. Jahanbakht, W. Xiang, L. Hanzo, and M. Rahimi Azghadi, "Internet of underwater things and big marine data analytics-A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 904–956, Jan. 2021.

[2] T. Qiu, Z. Zhao, T. Zhang, C. Chen, and C. P. Chen, "Underwater Internet of things in smart ocean: System architecture and open issues," *IEEE Trans. Ind. Informat*, vol. 16, no. 7, pp. 4297–4307, Oct. 2019.

[3] X. Shen, J. Gao, W. Wu, M. Li, C. Zhou, and W. Zhuang, "Holistic network virtualization and pervasive network intelligence for 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 1–30, 1st. Quart. 2022.

[4] 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, Aug. 2018.

[5] R. Su, D. Zhang, C. Li, Z. Gong, R. Venkatesan, and F. Jiang, "Localization and data collection in AUV-aided underwater sensor networks: Challenges and opportunities," *IEEE/ACM Trans. Netw.*, vol. 33, no. 6, pp. 86–93, Dec. 2019.

[6] I. Jawhar, N. Mohamed, J. Al-Jaroodi, and S. Zhang, "An architecture for using autonomous underwater vehicles in wireless sensor networks for underwater pipeline monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1329–1340, Mar. 2019.

[7] Z. Liu, X. Meng, Y. Liu, Y. Yang, and Y. Wang, "AUV-aided hybrid data collection scheme based on value of information for Internet of underwater things," *IEEE Internet Things J.*, vol. 9, no. 9, pp. 6944–6955, May 2022.

[8] 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. 10 010–10 022, Oct. 2020.

[9] X. Hou, J. Wang, T. Bai, Y. Deng, Y. Ren, and L. Hanzo, "Environment-aware AUV trajectory design and resource management for multi-tier underwater computing," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 474–490, Feb. 2023.

[10] W. Wu, C. Zhou, M. Li, H. Wu, H. Zhou, N. Zhang, S. Xuemin, and W. Zhuang, "AI-native network slicing for 6G networks," *IEEE Wireless Commun.*, vol. 29, no. 1, p. 96–103, Feb. 2022.

[11] Z. Chu, F. Wang, T. Lei, and C. Luo, "Path planning based on deep reinforcement learning for autonomous underwater vehicles under ocean current disturbance," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 108–120, Jan. 2023.

[12] M. Cheng, Q. Guan, F. Ji, J. Cheng, and Y. Chen, "Dynamic-detection-based trajectory planning for autonomous underwater vehicle to collect data from underwater sensors," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13 168–13 178, Aug. 2022.

[13] G. Han, A. Gong, H. Wang, M. Martínez-García, and Y. Peng, "Multi-AUV collaborative data collection algorithm based on Q-learning in underwater acoustic sensor networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9294–9305, Jul. 2021.

[14] W. Wu, N. Chen, C. Zhou, M. Li, X. Shen, W. Zhuang, and X. Li, "Dynamic RAN slicing for service-oriented vehicular networks via constrained learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2076–2089, Jul. 2021.

[15] M. Dai, Y. Li, P. Li, Y. Wu, L. Qian, B. Lin, and Z. Su, "A survey on integrated sensing, communication, and computing networks for smart oceans," *Sensor. Actuator Netw J*, vol. 11, no. 4, pp. 70–70, Oct. 2022.

[16] Y. Wang, Z. Shi, X. Ma, and L. Liu, "A joint sonar-communication system based on multicarrier waveforms," *IEEE Signal Process. Lett.*, vol. 29, pp. 777–781, Mar. 2022.

[17] J. Yin, W. Men, X. Han, and L. Guo, "Integrated waveform for continuous active sonar detection and communication," *IET Radar, Sonar & Nav*, vol. 14, no. 9, pp. 1382–1390, Jul. 2020.

[18] J. Clifton and E. Laber, "Q-learning: theory and applications," *Annu Rev Stat Appl*, vol. 7, pp. 279–301, Mar. 2020.

[19] W. Wu, M. Li, K. Qu, C. Zhou, X. Shen, W. Zhuang, X. Li, and W. Shi, "Split learning over wireless networks: Parallel design and resource management," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 4, pp. 1051–1066, 2023.

[20] C. Lin, G. Han, T. Zhang, S. B. H. Shah, and Y. Peng, "Smart underwater pollution detection based on graph-based multi-agent reinforcement learning towards AUV-based network ITS," *IEEE Trans. Intell. Transp. Syst.*, to appear, Apr. 2022, 10.1109/TITS.2022.3162850.