Environment and Energy-Aware AUV-Assisted Data Collection for the Internet of Underwater Things

Zekai Zhang*, Jingzehua Xu*, Guanwen Xie, Jingjing Wang, *Senior Member, IEEE*, Zhu Han, *Fellow, IEEE*, and Yong Ren, *Senior Member, IEEE*

***Abstract*—Considering the wide-area distribution and limited transmission power of sensing devices in the Internet of Underwater Things (IoUT), employing autonomous underwater vehicles (AUVs) to collect data is considered a promising solution. While most existing AUV-assisted data collection schemes primarily focus on enhancing data collection throughput and identifying the shortest path, they often overlook the influence of the underwater environment on AUV and the timeliness of data collection. In this paper, we design a multi-AUV-assisted data collection system, in which AUVs select their own target devices to collect data according to the data upload urgencies of IoUT devices. Considering the disturbance of turbulent ocean environment and the limited energy of AUV, we propose an environment and energy-aware AUV-assisted data collection scheme. This scheme aims to conduct path planning for multiple AUVs based on perceived environmental information, including turbulent fields and device statuses. The primary goals are to maximize the sum data collection rate and total data throughput, minimize AUV energy consumption, reduce the average data overflow times. To solve this high-dimensional NP-hard problem, we first model the problem as a Markov decision process, and propose a multi-agent independent soft actor-critic to solve it. Extensive simulations validate the effectiveness and adaptability of our approach.**

***Index Terms*—Internet of Underwater Things, autonomous underwater vehicle, data upload urgency, data collection, path planning, multi-agent independent soft actor-critic.**

This work of Jingjing Wang was partly supported by the National Natural Science Foundation of China under Grant No. 62071268 and No. 62222101, partly supported by the Young Elite Scientist Sponsorship Program by the China Association for Science and Technology under Grant No. 2020QNRC001, and partly supported by the Fundamental Research Funds for the Central Universities. This work of Yong Ren was partly supported by the National Natural Science Foundation of China under Grant 62127801, partly supported by the National Key Research and Development Program of China under Grant 2020YFD0901000. This work of Zhu Han was partially supported by NSF CNS-2107216, CNS-2128368, CMMI-2222810, ECCS-2302469, US Department of Transportation, Toyota and Amazon. (*Corresponding author: Jingjing Wang*)

Z. Zhang and J. Xu are with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China. E-mail: {zhangzej21, xjzh23}@mails.tsinghua.edu.cn.

G. Xie is with the Ocean College, Zhejiang University, Zhoushan, 316000, China. E-mail: 3200101418@zju.edu.cn.

J. Wang is with the School of Cyber Science and Technology, Beihang University, Beijing 100191, China. E-mail: drwangjj@buaa.edu.cn.

Z. Han is with the Department of Electrical and Computer Engineering at the University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul, South Korea, 446-701. E-mail: hanzhu22@gmail.com.

Y. Ren is with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China. E-mail: reny@tsinghua.edu.cn.

*These authors contributed equally to this work.

I. Introduction

ORIGINATING from the Internet of Things, the Internet of Underwater Things (IoUT) is an intelligent network connected by various types of underwater sensing devices [1], [2]. IoUT uses these devices and related technologies to perceive, interpret, and transmit data about the underwater environment, facilitating activities such as underwater searching and resource exploration [3]. Unlike terrestrial wireless communication environments, the underwater environment is complex and changeable, and the electromagnetic signal is seriously attenuated in underwater, resulting in the inapplicability of radio frequency communication technology to IoUT, and underwater acoustic communication is widely used as a remedy [4]. However, IoUT networks based on acoustic communication face the shortcomings of low bandwidth and long delay, thus making them unsuitable for long-distance transmission of large amounts of data. Consequently, devising efficient and reliable data collection schemes within IoUT networks remains a pivotal challenge for ongoing research.

Traditional data collection methods typically utilize multi-hop routing among IoUT devices, this process will increase energy consumption and thereby shortens the operational lifespan of battery-powered devices without recharging capabilities [3]. To address this challenge, hierarchical data collection schemes are widely discussed [5]–[7]. Within these schemes, devices are first organized into clusters with designated cluster heads responsible for gathering data from other devices in their respective clusters. These cluster heads then uniformly transmit the collected data to the surface base station. However, these cluster heads often succumb prematurely due to the increased communication overhead, which can lead to potential network failures. In contrast, deploying autonomous underwater vehicles (AUVs) as mobile platforms for collecting data through acoustic links has proven to be a more cost-effective and efficient alternative [8].

The current AUV-assisted data collection schemes often rely on predetermined trajectories, in which devices far from the AUV's trajectory need to forward data to the AUV through other devices, which inevitably leads to redundant energy consumption [9]. Some studies have focused on optimizing AUV trajectories to enhance their autonomy to actively collect data from devices, rather than following a predetermined path, Liu *et al*. modeled the trajectory optimization problem as a traveling salesman problem to find the shortest path [10]. Duan *et al*. optimized the AUV collection trajectory by taking the value of information of the entire IoUT network as an index

[8]. Huang *et al*. proposed a two-stage trajectory optimization mechanism and adopted greedy algorithm to solve it [11]. In general, the aforementioned efforts focus on optimizing the AUV trajectory by minimizing sailing distance, overlooking the effects of the turbulent environment, communication limitations, and device upload demands on the trajectory.

The real-time statuses of IoUT devices vary based on their geographical location and operational tasks, yet limited research addresses this variability. Specifically, devices with a larger backlog of stored data and a higher environmental data generation rate need to be preferentially collected by AUV, because once the data is not collected in time, it will lead to data overflow and the old data will be overwritten by the new data [12]. In this case, the single AUV cannot meet the dynamic upload requirements of IoUT devics in time, and so it is necessary to study the scheduling scheme of multi-AUV data collection to work cooperatively [13]. Furthermore, the current multi-AUV scheduling schemes are often designed for a specific task, and the optimization methods used rely on a large amount of prior data and lack adaptability, which is not suitable for solving the multi-objective optimization data collection problem in complex underwater environments.

Based on the above challenges and analysis, we design a multi-AUV assisted data collection system and optimize the AUV trajectory under environmental and communication constraints to trade off among the AUV energy consumption, the timeliness of data collection, and efficiency of data collection. Our main contributions can be summarized as follows.

- To the best of our knowledge, this is the first multi-AUV data collection work that takes into account the effects of underwater turbulent environments and the dynamic upload requirements of IoUT devices. Through environment and energy-aware trajectory optimization, we achieve trade-offs between multiple optimization objectives such as AUV energy consumption, data collection timeliness, and total data collection rate.
- Since the constrained multi-objective optimization problem formulated is high dimensional NP-hard, it is difficult to deal with. Therefore, we model it as a Markov decision process and design rewards and penalties for corresponding objectives and constraints, and propose a multi-agent independent soft actor-critic (MAISAC) based on decentralized training and decentralized execution to solve it, so as to adapt to the multi-AUV dynamic environment.
- Numerous simulation results indicate that our approach can optimize the trajectories of AUVs based on environmental perception, thereby enhancing data collection efficiency. Confronted with varying environments, communication ranges, and numbers of AUVs, among other escalating dimensions, our approach demonstrates feasibility and scalability.

The rest of this article is organized as follows. Section II reviews relevant cutting-edge work. In Section III, the system model is introduced in detail. In Section IV, the constrained optimization problem is formulated and modeled as MDP, which is solved by MAISAC. In Section V, the performance of our scheme is analyzed from multiple dimensions, and finally the conclusion is drawn in Section VI.

II. Related Works

Early data collection schemes primarily concentrate on collaborative transmission among IoUT devices, Zhang *et al*. proposed a hybrid protocol of selective relay cooperation and dynamic network coding cooperation to improve the data collection efficiency through cooperation between IoUT devices [14]. Han *et al*. classified IoUT devices into different virtual data sets and utilized hierarchical strategies of dynamic layer and static layer to optimize data transmission [15]. The data collection work relying on IoUT device cooperation can be summarized as optimizing data fusion between devices [16], proposing hierarchical collection architecture [15], and designing efficient routing strategies [17]. However, due to the inability to extend the device battery life, schemes relying on data fusion and multi-hop transmission between devices impose an excessive burden, significantly shortening the device's operational lifespan. Due to the flexible and autonomous characteristics of AUVs, Yoon *et al*. innovatively utilized AUVs as mobile relay nodes for the first time to collect data in multi-hop IoUT [9]. Hao *et al*. designed the trajectory of the AUV by predicting the location of the routing void to collect data of nearby IoUT devics. However, the devics far away from the AUV need to pay extra energy to forward data to the AUV [18]. The aforementioned schemes overlook the impact of AUV trajectory optimization on the performance of the data collection system and the efficiency of the collection scheme.

Data collection schemes that consider trajectory optimization primarily focus on the operational cost of AUVs, taking into account factors such as sailing distance and energy consumption during the mission process [19], [20]. Specifically, considering the communication constraints between the AUV and devices, Zhuo *et al*. converted the path planning problem to the traveling salesman problem (TSP) to minimize the AUV's travel time [19]. Similarly, Faigl et al. converted multi AUV data collection into TSP and found the shortest path based on self-organizing mapping, which has certain adapt ability [20]. However, these trajectory optimization studies are conducted in ideal environments, neglecting the impact of the complex underwater environment on AUV trajectory. Mahmoudzadeh *et al*. analyzed the characteristics of ocean turbulence and revealed that these factors would not only interfere with the velocity and direction of AUV, but also affect the energy consumption of AUV [21]. To address the trajectory optimization challenges in both static and dynamic turbulent environments, existing studies have explored a variety of algorithms, including differential evolution, group optimization, and task planning algorithms, and these investigations offer valuable insights [22], [23]. Although the above work discussed the trajectory optimization problem in turbulent environment, they did not consider the requirements of mission-critical IoUT for timeliness and quality of data collection [24]. Liu *et al*. jointly optimized the timeliness and energy efficiency of data collection, and introduced value of information to optimize the trajectory to ensure the timeliness

of delay sensitive data [10]. Fang *et al*. comprehensively considered AUV's trajectory optimization, energy consumption, and resource allocation, and introduced the age of data to ensure the freshness of data [25]. It is neglected that the IoUT devices are different due to differences in distribution, assigned tasks, and settings, which requires the AUVs to preferentially access certain devices. Yu *et al*. discussed the data upload priority of devices when studying unmanned aerial vehicle (UAV) assisted Internet of Things data collection, and adopted reinforcement learning to optimize the trajectory of the UAV, thus achieving a trade-off between the timeliness and the throughput of data collection [26]. Due to the effective learning and optimization of decision-making processes in complex environments offered by reinforcement learning, underwater data collection based on reinforcement learning has emerged as a promising approach. Wang *et al*. proposed a collaborative data collection method for multiple AUVs based on local-global deep Q-learning and data value, which categorizes data into urgent and non-urgent types, achieving hybrid data collection to meet the temporal requirements of different data types [27]. Zhao *et al*. designed a multi-level energy-efficient routing strategy based on reinforcement learning to fulfill the multiple transmission delay demands of various marine applications in multimodal IoUT, enhancing the reliability and network efficiency of data collection [17]. Jiang *et al*. utilized a multi-agent proximal policy optimization reinforcement learning algorithm to guide efficient and energy-saving data collection for AUV swarms in unknown environments based on an uncertainty map of objectives and a digital pheromone mechanism [28].

Based on the above analysis, there is no AUV-assisted IoUT data collection scheme that comprehensively considers environmental disturbance, energy efficiency and device characteristics. In addition, when the task complexity increases, the use of multiple AUVs for collaborative data collection also needs to be studied. Furthermore, traditional methods are not suitable for solving this high-dimensional NP-hard optimization problem. Therefore, based on the proposed MAISAC, this paper designs an environment and energy-aware multi-AUV efficient data collection scheme to fill this gap, which is of great significance to support complex IoUT applications.

III. System Model

A. Multi-AUV Assisted Data Collection System

We consider a multi-AUV assisted data collection system, as shown in Fig. 1, where multiple AUVs navigate in a turbulent ocean environment and determine which target devices they need to visit for data collection based on the status information of IoUT devices broadcast by the surface station. Assume that there are K AUVs and M IoUT devices. K AUVs can be denoted as the set $\boldsymbol{AUVs} = \{AUV_1, AUV_2, \ldots, AUV_K\}$, M IoUT devices can be denoted as the set $\boldsymbol{D} = \{D_1, D_2, \ldots, D_M\}$. For brevity, let $\boldsymbol{K} = \{1, 2, \ldots K\}$ represent the subscript of AUVs, and $\boldsymbol{M} = \{1, 2, \ldots, M\}$ represent the subscript of the devices. The three-dimensional coordinates of the IoUT device i and AUV k are denoted as $\boldsymbol{P}_i = (x_i, y_i, d)$ and $\boldsymbol{P}_k(t) = (x_k(t), y_k(t), z)$,

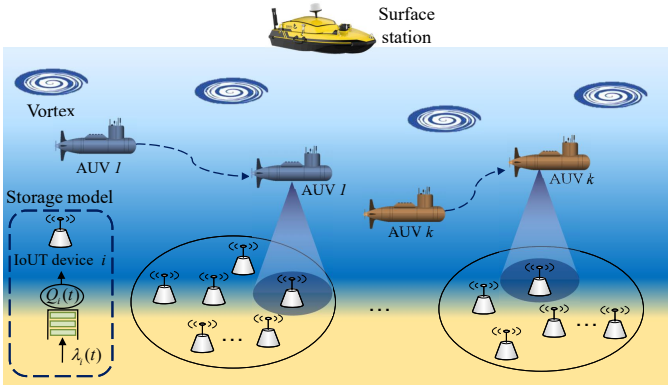


Fig. 1. Illustration of multi-AUV assisted data collection system.

where $0 \leq t \leq T$, T is the given task time. Assume that the state of IoUT device i can be represented by the twin tuple $W_i(t) \triangleq \{Q_i(t), \lambda_i(t)\}$, where $Q_i(t)$ represents the stored data (in bit) of IoUT device i at time t, and $\lambda_i(t)$ represents the data generation rate (in bit/s) of IoUT device i at time t. It is worth noting that the turbulent environment will have an impact on the motion and energy consumption of AUV. Each AUV is equipped with a horizontal acoustic Doppler current profpler (HADCP) for measuring ocean current velocity, which the manufacturer claims is accurate to 1%±5 mm/s and can be used to measure currents on a horizontal line hundreds of meters ahead [12]. With HADCP, the AUV can sense information about the surrounding turbulent field and reduce energy consumption through trajectory optimization.

B. Target Device Selection Model

Before scheduling AUVs to access IoUT devices, it is necessary to determine the target device for each AUV to access based on the data upload urgency of the devices. The current stored data $Q_i(t)$ of device i at time t depends on the status of the previous moment and the data generation rate. The update equation of $Q_i(t)$ is [12]

$$Q_i(t + \Delta t) = Q_i(t) + \lambda_i(t) \cdot \Delta t, \tag{1}$$

where Δt is the update interval, $Q_i(t) \in [0, Q_{\max}]$, $Q_{\max}$ is the maximum amount of data that can be stored by the device, which is constrained by hardware limitations. And $\lambda_i(t)$ is the data generation rate of device i, which follows the Poisson distribution, and we assume that the parameters of the Poisson distribution of each device are different [26]. As the data backlog and data generation rate of devices are different, their data upload urgency is also different. The upload urgency $q_i(t)$ of device i can be defined as

$$q_i(t) = \lambda_i(t) \cdot \frac{Q_i(t)}{Q_{\max}}. \tag{2}$$

As can be seen from (2), the upload urgency of device i not only depends on the current state reflected by the ratio of stored data to storage capacity $Q_i(t)/Q_{\max}$, but also includes the prediction of future state reflected by the data generation rate $\lambda_i(t)$. Considering the actual situation, the priority of AUV k to access device i depends not only on the upload

urgency of device i but also on its relative distance from AUV k. Therefore, we define $q_k^i(t)$ to represent the priority of AUV k to access device i, which is

$$q_k^i(t) = \lambda_i(t) \cdot \frac{Q_i(t)}{Q_{\max}} - a d_{k,i}, \tag{3}$$

where $d_{i,k}$ is the distance between between AUV k and device i , and a is the distance penalty factor. However, the determination of the a value needs to be cautious, as a too large a may prevent the AUV from accessing devices that are located far away but have high upload urgency. AUV k can calculate the willingness to access a particular device based on (3) and select the device with the highest access priority as the target device.

C. Underwater Acoustic Communication Channel Analysis

The AUV collects data from the target device by establishing the underwater acoustic communication channel, so we analyze the underwater acoustic communication channel. In a shallow water acoustic propagation environment, the path loss $A(d_{k,i}, f)$ of an acoustic signal with frequency f between AUV k and device i is

$$A(d_{k,i}, f) = d_{k,i}^{\varsigma} a(f)^{d_{k,i}}, \tag{4}$$

where ς is the spread factor, and $a(f)$ is the absorption coefficient in in dB per km, as calculated using the Thorp formula [29]

$$\begin{aligned} 10\log(a(f)) = \ & 0.11\frac{f^2}{1+f^2} + 44\frac{f^2}{4100+f^2} \\ & + 2.75\times 10^{-4} f^2 + 0.003. \end{aligned} \tag{5}$$

According to [30], underwater environmental noise $N(f)$ is composed of turbulence, ship, wind and thermal noises [11], denoted as $Nt(f), Ns(f), Nw(f)$ and $Nth(f)$, respectively, and therefore we have

$$N(f) = Nt(f) + Ns(f) + Nw(f) + Nth(f). \tag{6}$$

The noise components in (6) can be respectively described as [31]

$$\begin{cases} 10\log N_t(f) = 17 - 30\log f, \\ 10\log N_s(f) = 30 + 20s + \log\left(f^{26}/(f+0.03)^{60}\right), \\ 10\log N_w(f) = 50 + 7.5\vartheta^{1/2} + 20\log\left(f/(f+0.4)^2\right), \\ 10\log N_{th}(f) = -15 + 20\log f, \end{cases} \tag{7}$$

where s is the shipping activity factor, and ϑ is wind speed, $s \in [0,1]$. Thus, the signal-to-noise ratio (SNR) $\gamma(d_{k,i}, f)$ of the channel between device i and AUV k can be given by

$$\gamma(d_{k,i}, f) = \frac{1}{A(d_{k,i}, f) \cdot N(f)}, \tag{8}$$

where $\gamma(d_{k,i}, f)$ is related to distance $d_{k,i}$ and frequency f, which makes it difficult for channel capacity analysis. To simplify, we assume that there is an optimal frequency $f_o(d_{k,i})$ for the given communication distance $d_{k,i}$, and the SNR at this frequency is $\gamma_o(d_{k,i})$. We define a 3-dB frequency range $[f_L(d_{k,i}), f_U(d_{k,i})]$, satisfying $\gamma(d_{k,i}, f_L(d_{k,i})) = \gamma(d_{k,i}, f_U(d_{k,i})) = \gamma_o(d_{k,i}) - 3\mathrm{dB}$ [29]. Assuming that the device and AUV transmits data using a narrow-band signal with a center frequency f_c and bandwidth w falling within the 3 dB frequency range, $\mathrm{SNR}\ \gamma(d_{k,i}, f)$ of the true channel can be replaced by the following switching function

$$\tilde{\gamma}(d_{k,i}) = \begin{cases} \min\left\{\gamma\left(d_{k,i}, f_c - \frac{w}{2}\right), \gamma\left(d_{k,i}, f_c + \frac{w}{2}\right)\right\}, f \in w, \\ 0, \ \text{otherwise.} \end{cases} \tag{9}$$

Assuming that the channel is an additive white Gaussian noise channel, the channel capacity $R(d_{k,i})$ is

$$R(d_{k,i}) = w\log_2\left(1 + \frac{P_{SL}\tilde{\gamma}(d_{k,i})}{w}\right), \tag{10}$$

where $P_{SL}(\ \mathrm{dB\ re}\ \mu\mathrm{Pa})$ is the source level. To convert electrical power P_T in watts to acoustic power P_{SL}, we have the following empirical relationship

$$I_T = \frac{\eta P_T}{2\pi H}, \tag{11}$$

$$P_{SL} = 10\log\frac{I_T}{1\mu\mathrm{Pa}}, \tag{12}$$

where η represents the overall efficiency of the electronic circuitry, which includes the power amplifier and transducer, and H is the water depth. Moreover, I_T denotes the intensity at a reference distance of 1 meter, with the chosen value of $1\ \mu\mathrm{Pa}$ equivalent to be $0.67\times 10^{-22}\ \mathrm{W/cm^2}$.

D. Trajectory Model

Assuming that the set of s_k hovering points of AUV k in the whole process is $\boldsymbol{S}_k$, we set $\boldsymbol{P}_k[\xi], \forall \xi \in \boldsymbol{S}_k = \{1, 2, \ldots, s_k\}$ as the trajectory of AUV k. Then, we define $d_k[\xi]$ as the distance between the two hovering points, which can be calculated as

$$d_k[\xi] = \left\|\boldsymbol{P}_k[\xi+1] - \boldsymbol{P}_k[\xi]\right\|_2, \forall k \in \boldsymbol{K}, \forall \xi \in \boldsymbol{S}_k. \tag{13}$$

Let $h_{k,i}[\xi] = 1$ indicate that AUV k can access device i to collect data at the ξ-th hovering point, otherwise $h_{k,i}[\xi] = 0$. The trajectory design strategy of AUVs can be expressed as $\boldsymbol{H} = \{h_{k,i}[\xi], k \in \boldsymbol{K}, i \in \boldsymbol{M}, \xi \in \boldsymbol{S}_k\}$. In order to ensure that each AUV serves only one device at a time, as well as that a device is only served by one AUV. We have

$$\sum_{k=1}^{K} h_{k,i}[\xi] = 1, \forall k \in \boldsymbol{K}, \forall i \in \boldsymbol{M}, \tag{14}$$

and

$$\sum_{i=1}^{M} h_{k,i}[\xi] = 1, \forall k \in \boldsymbol{K}, \forall i \in \boldsymbol{M}. \tag{15}$$

The time required for AUV k to collect data from device i at the ξ-th hovering point can be denoted as

$$c_{k,i}[\xi] = \frac{Q_i(t_\xi)}{R(d_{k,i}[\xi])}, \tag{16}$$

where $d_{k,i}[\xi]$ is the distance between AUV k and device i at the ξ-th hovering point, and t_ξ represents the time to reach the ξ-th hovering point.

E. Motion Model

We assume that AUV k is moving in a two-dimensional plane with a fixed depth, and its state changes discretely with time, the three-degree-of-freedom motion model is derived without loss of generality. The state of AUV k at time stamp t is $S_k(t) = \{x_k(t), y_k(t), v_{k,x}(t), v_{k,y}(t), a_k(t), \theta_k(t), \omega_k(t)\}$, the elements in the set represent current position, velocity, acceleration, yaw angle, and angular velocity. Since the time stamp is small, the motion within the time stamp can be regarded as uniformly accelerated motion, and the position update equation is

$$x_k(t+1) = x_k(t) + (v_{k,x}(t) + v_{k,x}(t+1))\,\Delta t/2, \quad (17a)$$

$$y_k(t+1) = y_k(t) + (v_{k,y}(t) + v_{k,y}(t+1))\,\Delta t/2. \quad (17b)$$

Similarly, the velocity update equation is

$$v_{k,x}(t+1) = v_{k,x}(t) + a_k(t)\cos(\theta_k(t+1))\,\Delta t, \quad (18a)$$

$$v_{k,y}(t+1) = v_{k,y}(t) + a_k(t)\sin(\theta_k(t+1))\,\Delta t. \quad (18b)$$

Moreover, the angle update equation is

$$\theta_k(t+1) = \theta_k(t) + \omega_k(t)\Delta t. \quad (19)$$

Considering that the angle range is $[-\pi, \pi]$, Eq. (19) is modified as

$$\theta_k(t{+}1) = \begin{cases} \theta_k(t+1) - 2\pi, & \theta_k(t+1) > \pi, \\ \theta_k(t+1) + 2\pi, & \theta_k(t+1) < -\pi, \\ \theta_k(t+1), & \theta_k(t+1) \in [-\pi, \pi]. \end{cases} \quad (20)$$

AUVs are assumed to have the same equipement limitations from the perspectives of steering and acceleration rate, i. e.,

$$\|\omega_k(t)\| \le \zeta_\omega, \quad \forall t, \forall k \in \boldsymbol{K}, \quad (21)$$

$$\|a_k(t)\| \le \zeta_a, \quad \forall t, \forall k \in \boldsymbol{K}, \quad (22)$$

where $\zeta_\omega > 0$ and $\zeta_a > 0$ are steering range limit constant and acceleration limit constant, respectively.

F. Turbulent Ocean Environment Modeling

AUV navigation in the marine environment will be affected by ocean current, wave, wind and other factors, among which the latter two factors can be ignored below 2 m water level, and the AUV movement is mainly affected by ocean current. Due to the Earth's rotation effect, the strength of ocean currents on the horizontal plane is much greater than that on the vertical plane. Considering that this study is conducted on the horizontal plane, the speed of ocean currents on the vertical plane can be ignored [32]. Ocean current models can be characterized by multiple turbulent regions. Which is modeled by two-dimensional Navier-Stokes equations [33]

$$\frac{\partial \varpi}{\partial t} + (\boldsymbol{v}_c \nabla)\,\varpi = \upsilon \Delta \varpi, \quad (23)$$

where $\boldsymbol{v}_c = (V_x, V_y)$ is the velocity of the turbulent field, ϖ and υ are the vorticity of current and the viscosity of the fluid, and ∇ and Δ are the gradient operators and Laplacian operators, respectively. To simplify the Navier-Stokes equation, the numerical equation of the ocean current model is represented by the superposition of several viscous vortex functions, which is described as follows [32]

$$V_x(\boldsymbol{P}(t)) = -\Gamma \cdot \frac{y - y_0}{2\pi \left\|\boldsymbol{P}(t) - \boldsymbol{P}_0\right\|_2^2} \cdot \left(1 - e^{-\frac{\|\boldsymbol{P}(t) - \boldsymbol{P}_0\|_2^2}{\delta^2}}\right), \quad (24)$$

$$V_y(\boldsymbol{P}(t)) = -\Gamma \cdot \frac{x - x_0}{2\pi \left\|\boldsymbol{P}(t) - \boldsymbol{P}_0\right\|_2^2} \cdot \left(1 - e^{-\frac{\|\boldsymbol{P}(t) - \boldsymbol{P}_0\|_2^2}{\delta^2}}\right), \quad (25)$$

and

$$\varpi(\boldsymbol{P}(t)) = \frac{\Gamma}{\pi \delta^2} \cdot e^{-\frac{\|\boldsymbol{P}(t) - \boldsymbol{P}_0\|_2^2}{\delta^2}}, \quad (26)$$

where $\boldsymbol{P}(t)$ and $\boldsymbol{P}_0$ are the current position of AUV and the coordinate vector of Lamb vortex center, δ is the radius of vortex, and Γ is the intensity of vortex [34]. Then the velocity of AUV k under ocean current interference is

$$\boldsymbol{v}_k(\boldsymbol{P}_k(t)) = \boldsymbol{v}_k - \boldsymbol{v}_c(\boldsymbol{P}_k(t)), \quad (27)$$

where $\boldsymbol{v}_c(\boldsymbol{P}_k(t))$ is the water flow velocity at position $\boldsymbol{P}_k(t)$ and $\boldsymbol{v}_k(\boldsymbol{P}_k(t))$ is the relative velocity at position $\boldsymbol{P}_k(t)$ According to the classical computational fluid dynamics (CFD) method [35], the drag force of AUV k hovering and sailing can be expressed as [12]

$$F_k^h = \frac{1}{2}\rho_l \left\|\boldsymbol{v}_c(\boldsymbol{P}_k(t))\right\|_2^2 C_a C_d, \quad (28)$$

and

$$F_k^m = \frac{1}{2}\rho_l \left\|\boldsymbol{v}_k(\boldsymbol{P}_k(t))\right\|_2^2 C_a C_d, \quad (29)$$

respectively, where ρ_l is the density of seawater. C_a and C_d are the drag coefficient and the front area of AUV, respectively.

G. Energy Consumption Model

The energy consumption of AUV comes from the hovering energy consumption during data collection and the motion energy consumption between two hovering points. According to (28), the power consumption of AUV k at the ξ-th hovering point is [29]

$$P_k^h[\xi] = \frac{1}{\zeta} F_k^h[\xi] \left\|\boldsymbol{v}_c(\boldsymbol{P}_k[\xi])\right\|_2^2, \quad (30)$$

where ζ is the electrical conversion efficiency. Since the speed of water flow at each point in the subtrajectory (between two hovering points) is different, this poses a challenge to the analysis of motion energy consumption. To solve this challenge, we calculate the average of the relative velocities at the starting point, end point and midpoint of the subtrajectory to approximate the average relative velocity during the subtrajectory, taking the process of AUV k moving from the ξ th hovering point to the $\xi + 1$-th hovering point as an example, and the average relative velocity is expressed as [29]

$$\tilde{\boldsymbol{v}}_k(\boldsymbol{P}_k[\xi]) = \frac{1}{3}(\boldsymbol{v}_k(\boldsymbol{P}_k[\xi]) + \boldsymbol{v}_k(\boldsymbol{P}_k[\xi_m]) + \boldsymbol{v}_k(\boldsymbol{P}_k[\xi{+}1])), \quad (31)$$

where ξ_m is the midpoint between the ξ-th hovering point and the $\xi{+}1$-th hovering point. According to (29), the resistance of AUV k in the motion of the ξ-th subtrajectory can be expressed as

$$F_k^m = \frac{1}{2}\rho_l \left\|\tilde{\boldsymbol{v}}_k(\boldsymbol{P}_k[\xi])\right\|_2^2 C_a C_d, \quad (32)$$

Therefore, the power consumption required by AUV k in the ξ-th subtrajectory motion can be given by

$$P_k^m[\xi] = \frac{1}{\zeta} F_k^m[\xi] \left\| \tilde{\boldsymbol{v}}_k \left(\boldsymbol{P}_k[\xi] \right) \right\|_2^2 . \tag{33}$$

As a result, the total energy consumption of AUV k is

$$E_k = \sum_{i=1}^{M} \sum_{\xi=1}^{s_k} h_{k,i}[\xi] P_k^h[\xi] c_{k,i}[\xi] + \sum_{\xi=1}^{s_k} P_k^m[\xi] t_k^m[\xi], \tag{34}$$

where $t_k^m[\xi]$ is the time required to navigate from the ξ-th hovering point to the $\xi+1$-th hovering point.

IV. Problem Formulation and Algorithm Design

We formulate the problem and discussed optimization objectives and constraints in this section. Then, the data collection task is modeled as Markov decision process (MDP), and the corresponding reward function is designed. Finally, the MAISAC algorithm is proposed for AUV path planning to jointly optimize multiple objectives.

A. Problem Formulation

The optimization objectives we consider include maximizing the total data throughput and sum data collection rate, minimizing AUV energy consumption, and ensuring the timeliness of data collection. Timeliness of data collection is achieved by considering the upload urgencies of IoUT devices, while the other objectives can be examined using the profit of the multi-AUV data collection system, where profit is defined as revenue minus cost. The total data throughput, meaning the total amount of data collected in the task, is considered a revenue, reflecting the data collection capability of multiple AUVs. The sum data collection rate is also considered a revenue, as it reflects the state of communication links, affecting data collection efficiency. The cost of the system mainly comes from AUV operational energy consumption. The total data $\hat{Q}_k$ collected by AUV k can be expressed as

$$\hat{Q}_k = \sum_{i=1}^{M} \sum_{\xi=1}^{s_k} h_{k,i}[\xi] Q_i \left(t_\xi \right). \tag{35}$$

In the data collection task, the sum of the data collection rate of AUV k at each hover point is

$$\hat{R}_k = \sum_{i=1}^{M} \sum_{\xi=1}^{s_k} h_{k,i}[\xi] R \left(d_{k,i}[\xi] \right). \tag{36}$$

As a result, the profit of the system can be obtained

$$P_r = \sum_{k=1}^{K} \left(\mu \hat{Q}_k + \chi \hat{R}_k - \varsigma E_k \right), \tag{37}$$

where $\mu \in (0,1)$, $\chi \in (0,1)$ and $\varsigma \in (0,1)$ are the contribution or loss weight factors of the corresponding item to profit respectively, which can be dynamically adjusted according to demand.

To maximize the profit defined by (37), it is necessary to optimize the trajectory strategy $\boldsymbol{H}$. The optimization problem can be defined as

$$\max_{\boldsymbol{H}} P_r = \sum_{k=1}^{K} \left(\mu \hat{Q}_k + \chi \hat{R}_k - \varsigma E_k \right), \tag{38a}$$

$$\text{s.t. } h_{k,i}[\xi] = 1, \text{ if } d_{k,i}[\xi] \le R_r, \forall k \in \boldsymbol{K}, \forall i \in \boldsymbol{M}, \tag{38b}$$

$$\sum_{k=1}^{K} h_{k,i}[\xi] = 1, \forall k \in \boldsymbol{K}, \forall i \in \boldsymbol{M}, \tag{38c}$$

$$\sum_{i=1}^{M} h_{k,i}[\xi] = 1, \forall k \in \boldsymbol{K}, \forall i \in \boldsymbol{M}, \tag{38d}$$

$$\|\omega_k(t)\| \le \zeta_\omega, \|a_k(t)\| \le \zeta_a, \forall t, \forall k \in \boldsymbol{K}, \tag{38e}$$

where the optimization objective (38a) is to maximize the total profit of the multi-AUV data collection system. Constraint (38b) indicates that the AUV can collect data only when the distance between the device and the hovering point is within the communication range R_r of the AUV. Constraint (38c) indicates that a device can only communicate with one AUV. Constraint (38d) indicates that an AUV can access only one device simultaneously. Constraint (38e) limits the angular velocity and acceleration of the AUV according to the actual situation.

B. Markov Decision Process Modeling

Since the optimization objective (38a) is NP-hard and it is difficult to find the optimal solution under multiple constraints. To solve it, we first need to convert the optimization objective (38a) into an MDP and then adopt reinforcement learning method to solve it. MDP can be defined by a quintuple [28]

$$\Omega = \{ \boldsymbol{S}, \boldsymbol{A}, \boldsymbol{P}, \boldsymbol{R}, \mu \}, \tag{39}$$

where $\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{R}$ represent state space, action space and reward function, $\boldsymbol{P}$ is state transition probability distribution, μ is discount factor. We pay special attention to $\boldsymbol{S}, \boldsymbol{A}$ and $\boldsymbol{R}$ of AUV k when designing the algorithm

1) **State Space:** In the data collection task, the observation space of AUV k at time t is $\boldsymbol{s}_k(t) \in \boldsymbol{S}$, defined as

$$\begin{aligned} \boldsymbol{s}_k(t) = \{ & \boldsymbol{P}_i, W_i(t), \boldsymbol{P}_k(t), q_k^i(t), \boldsymbol{v}_c \left(\boldsymbol{P}_k(t) \right), d_{k,j}(t), \\ & N_k(t), N_l(t), i \in \boldsymbol{M}, j \in \boldsymbol{K}, j \ne k \}, \end{aligned} \tag{40}$$

where $\boldsymbol{P}_i$ is the location of device $i, W_i(t) \triangleq \{Q_i(t), \lambda_i(t)\}$ describes the state of device i at time t, $\boldsymbol{P}_k(t)$ is the location of AUV k at time $t, q_k^i(t)$ is the priority of AUV k to access device i, $\boldsymbol{v}_c \left(\boldsymbol{P}_k(t) \right)$ is the turbulent velocity of AUV k at the location of time t, $d_{k,j}(t)$ is the relative distance between AUV k and AUV j, $N_k(t)$ is the cumulative number of out-of-bounds of AUV k. $N_l(t)$ is the the number of devices with data overflow.

2) **Action Space:** In the process of data collection, the AUV k makes action $\boldsymbol{a}_k(t) \in \boldsymbol{A}$ at time t by observing state $s_k(t)$, which is

$$\boldsymbol{a}_k(t) = \{ i_k(t), a_k(t), \omega_k(t) \}, \tag{41}$$

where $i_k(t)$ indicates that AUV k selects the device i_k as target device at time t according to the calculation of the access priority of all devices.

3) **Reward Function:** In reinforcement learning, agents rely on rewards to evaluate learning strategies. To solve (38a), we need to design corresponding reward functions to guide AUVs to make reasonable decisions to optimize trajectories in complex environments and improve the efficiency of data collection. The reward $\boldsymbol{r}_k(t) \in \boldsymbol{R}$ received by the AUV k at time t consists of the following parts

$$r_k^{(1)}(t) = \begin{cases} R\left(d_{k,i_k}(t)\right), & \text{if } d_{k,i_k}(t) \le R_r, \\ 0, & \text{otherwise}, \end{cases} \tag{42}$$

$$r_k^{(2)}(t) = \begin{cases} -\frac{1}{2\zeta}\rho_l C_a C_d \left(\left\|\boldsymbol{v}_c(\boldsymbol{P}_k(t))\right\|_2^2\right)^2, & \text{if AUV is hovering}, \\ -\frac{1}{2\zeta}\rho_l C_a C_d \left(\left\|\boldsymbol{v}_k(\boldsymbol{P}_k(t))\right\|_2^2\right)^2, & \text{if AUV is moving}, \end{cases} \tag{43}$$

$$r_k^{(3)}(t) = \sum_{i=1}^{M}\sum_{\xi=1}^{s_t} h_{k_i i}[\xi] Q_i\left(t_\xi\right), \tag{44}$$

$$r_k^{(4)}(t) = \sum_{j=1, j\neq k}^{K} \left(d_{k,j}(t) - d_{safe}\right), \tag{45}$$

$$r_k^{(5)}(t) = -d_{k,i_k}(t) - N_k(t) - N_l(t), \tag{46}$$

where $r_k^{(1)}(t)$ is a reward item used to encourage AUV k to establish high-quality communication links. When the distance $d_{k,i_k}(t)$ between AUV k and target device i_k is less than R_r, AUV k will get a reward, which is the data collection rate between AUV k and target device i_k, and its magnitude is related to $d_{k,i_k}(t)$. $r_k^{(2)}(t)$ is to punish the excessive energy consumption of AUV k in the hovering and moving stages, and the power consumption of AUV k at time t is related to the velocity of AUV k and the turbulent velocity of the location. $r_k^{(3)}(t)$ represents the total amount of data collected by AUV k during the task period, which is regarded as a reward to encourage AUV k to collect more data. To prevent collisions between AUVs, we use $r_k^{(4)}(t)$ to punish AUVs when the distance between AUV k and AUV j is less than the safe distance d_{safe}. $r_k^{(5)}(t)$ is the system auxiliary reward, the greater the penalty is when AUV k is farther away from the target device i_k, the more times AUV k crosses the boundary, the more the penalty is, and the more the number of devices that fail to be visited by AUV k in time and cause data loss, the more the penalty is. Therefore, the total reward available for AUV k at time t can be weighted by

$$\boldsymbol{r}_k(t) = \sum_{l=1}^{5} \omega^{(l)} r_k^{(l)}(t), \tag{47}$$

where $\omega^{(l)}$ is the weight coefficient of each reward, which can be adjusted according to the application needs.

C. Multi-Agent Independent Soft Actor-Critic

Traditional reinforcement learning methods cannot adapt to the multi-AUV dynamic data collection environment considered in this paper. Considering that soft actor-critic (SAC) can naturally balance exploration and utilization compared with other popular reinforcement learning methods such as proximal policy optimization and deep Q-network, it can realize efficient learning in a wide range of tasks [36]. So we extend the SAC algorithm to MAISAC using decentralized training and decentralized execution (DTDE) to train the AUVs in parallel and independently, enabling them to perform their own tasks in the unknown dynamic environment. In MAISAC, AUV k has two action value functions Q_k^1 and Q_k^2, and a policy function π_{θ_k}. To tackle the challenge of Q value overestimation, we employ a pair of critic networks denoted as Θ_k^1 and Θ_k^2, along with their corresponding target networks $\tilde{\Theta}_k^1$ and $\tilde{\Theta}_k^2$. Opting for the network exhibiting a lower Q value serves to alleviate the overestimation issue. Consequently, the loss functions of Q_k^1 and Q_k^2 are

$$L_{Q_k^1}\left(\Theta_k^1\right) = E_{\left(\boldsymbol{s}_t^k, \boldsymbol{a}_t^k, \boldsymbol{r}_t^k, \boldsymbol{s}_{t+1}^k\right)\sim\mathcal{D}_k}\left[\frac{1}{2}Q_{\Theta_k^1}\left(\boldsymbol{s}_t^k, \boldsymbol{a}_t^k\right) - \left(\boldsymbol{r}_t^k + \gamma V_{\tilde{\Theta}_k^1}\left(\boldsymbol{s}_{t+1}^k\right)\right)\right]^2, \tag{48}$$

$$L_{Q_k^2}\left(\Theta_k^2\right) = E_{\left(\boldsymbol{s}_t^k, \boldsymbol{a}_t^k, \boldsymbol{r}_t^k, \boldsymbol{s}_{t+1}^k\right)\sim\mathcal{D}_k}\left[\frac{1}{2}Q_{\Theta_k^2}\left(\boldsymbol{s}_t^k, \boldsymbol{a}_t^k\right) - \left(\boldsymbol{r}_t^k + \gamma V_{\tilde{\Theta}_k^2}\left(\boldsymbol{s}_{t+1}^k\right)\right)\right]^2, \tag{49}$$

where $\mathcal{D}_k$ denotes the replay buffer, whereas $V_{\tilde{\Theta}_k^1}(\cdot)$ and $V_{\tilde{\Theta}_k^2}(\cdot)$ are the state value functions parameterized by $\tilde{\Theta}_k^1$ and $\tilde{\Theta}_k^2$, respectively. To prevent AUV k from becoming trapped in local optimal policy, we introduce entropy regularization and represent $V_{\tilde{\Theta}_k^1}\left(\boldsymbol{s}_{t+1}^k\right)$ and $V_{\tilde{\Theta}_k^2}\left(\boldsymbol{s}_{t+1}^k\right)$ as follows

$$V_{\tilde{\Theta}_k^1}\left(\boldsymbol{s}_{t+1}^k\right) = \min_{j=1,2} Q_{\tilde{\Theta}_k^j}\left(\boldsymbol{s}_{t+1}^k, \boldsymbol{a}_{t+1}^k\right) - \partial_k \log\left(\pi_{\theta_k}\left(\boldsymbol{a}_{t+1}^k | \boldsymbol{s}_{t+1}^k\right)\right), \tag{50}$$

$$V_{\tilde{\Theta}_k^2}\left(\boldsymbol{s}_{t+1}^k\right) = \min_{j=1,2} Q_{\tilde{\Theta}_k^j}\left(\boldsymbol{s}_{t+1}^k, \boldsymbol{a}_{t+1}^k\right) - \partial_k \log\left(\pi_{\theta_k}\left(\boldsymbol{a}_{t+1}^k | \boldsymbol{s}_{t+1}^k\right)\right), \tag{51}$$

where ∂_k is the regularization coefficient, determining the weight placed on entropy in the policy. Subsequently, the loss function for the policy can be derived from the simplified KL divergence

$$L_{\pi_{\theta_k}}(\theta_k) = E_{\boldsymbol{s}_t^k\sim\mathcal{D}_k, \boldsymbol{a}_t^k\sim\pi_{\theta_k}}\left[\partial_k \log\left(\pi_{\theta_k}\left(\boldsymbol{a}_t^k \mid \boldsymbol{s}_t^k\right)\right) - \min_{j=1,2} Q_{\Theta_k^j}\left(\boldsymbol{s}_t^k, \boldsymbol{a}_t^k\right)\right]. \tag{52}$$

To address the issue of non-differentiability when sampling actions from Gaussian distribution $\mathcal{N}$ the reparameterization trick is introduced, allowing the policy function to be expressed as $\boldsymbol{a}_t^k = f_{\theta_k}\left(\phi_t; \boldsymbol{s}_t^k\right)$, where ϕ_t represents a noise random variable. By considering two action value functions simultaneously, the policy's loss function is

$$L_{\pi_{\theta_k}}(\theta_k) = E_{\boldsymbol{s}_t^k\sim\mathcal{D}_k, \phi_t\sim\mathcal{N}}\left[\partial_k \log\left(\pi_{\theta_k}\left(f_{\theta_k}\left(\phi_t; \boldsymbol{s}_t^k\right) | \boldsymbol{s}_t^k\right)\right) - \min_{j=1,2} Q_{\Theta_k^j}\left(\boldsymbol{s}_t^k, f_{\theta_k}\left(\phi_t; \boldsymbol{s}_t^k\right)\right)\right], \tag{53}$$

Algorithm 1 MAISAC Algorithm

Initialize the training environment, including the replay buffer $\mathcal{D}_k$, network parameters, and entropy regularization Θ_k^1, Θ_k^2, $\tilde{\Theta}_k^1$, $\tilde{\Theta}_k^2$, θ_k, ∂_k of AUV k.
for each episode i **do**
  Reset the training environment and total reward.
  for each time step t **do**
    Sample an action for AUV k according to the policy:
    $\boldsymbol{a}_t^k \sim \pi_{\theta_k}\left(\boldsymbol{a}_t^k \mid \boldsymbol{s}_t^k\right)$;
    Collect the next state from environment:
    $\boldsymbol{s}_{t+1}^k \sim \mathcal{P}\left(\boldsymbol{s}_{t+1}^k \mid \boldsymbol{s}_t^k, \boldsymbol{a}_t^k\right)$;
    Calculate reward $\boldsymbol{r}_t^k$ by Eq. (42) $\sim$ (47);
    Store sampling tuple $\left(\boldsymbol{s}_t^k, \boldsymbol{a}_t^k, \boldsymbol{r}_t^k, \boldsymbol{s}_{t+1}^k\right)$ into $\mathcal{D}_k$.
    Extract N batches tuple of data from $\mathcal{D}_k$.
    $\Theta_k^j \leftarrow \Theta_k^j - \lambda_{\Theta_k^j} \nabla_{\Theta_k^j} J_{\Theta_k^j}\left(\Theta_k^j\right), \quad j = 1, 2.$
    $\theta_k \leftarrow \theta_k - \lambda_{\theta_k} \nabla_{\theta_k} J_{\theta_k}\left(\theta_k\right).$
    $\partial_k \leftarrow \partial_k - \lambda_{\partial_k} \nabla_{\partial_k} J_{\partial_k}(\partial_k).$
    $\tilde{\Theta}_k^j \leftarrow \kappa \Theta_k^j + (1-\kappa)\tilde{\Theta}_k^j, \quad j = 1, 2.$
  end for
end for

To automatically adjust the entropy regularization term, the goal of reinforcement learning can be reformulated as a constrained optimization problem

$$\max_{\pi_{\theta_k}} E_{\pi_{\theta_k}}\left[\sum_t \boldsymbol{r}_t^k\right] \text{ s.t. } E_{\boldsymbol{s}_t^k \sim \mathcal{D}_k, \boldsymbol{a}_t^k \sim \pi_{\theta_k}}\left[-\log\left(\pi_{\theta_k}\left(\boldsymbol{a}_t^k | \boldsymbol{s}_t^k\right)\right)\right] \geq H_0. \tag{54}$$

More intuitively, the objective is to maximize the expected total reward while ensuring that the entropy mean exceeds H_0. By simplifying (54), we can derive the loss function for

$$L(\partial_k) = E_{\boldsymbol{s}_t^k \sim \mathcal{D}_k, \boldsymbol{a}_t^k \sim \pi_{\theta_k}}\left[-\partial_k \log\left(\pi_{\theta_k}\left(\boldsymbol{a}_t^k \mid \boldsymbol{s}_t^k\right)\right) - \partial_k H_0\right]. \tag{55}$$

Eq. (55) implies that if the policy entropy is below the desired value H_0, the training target $L\left(\partial_k\right)$ will raise the value of ∂_k. Consequently, it will amplify the significance of the corresponding term in the policy entropy during the process of minimizing the loss function $L_{\pi_{\theta_k}}\left(\theta_k\right)$. Conversely, if the policy entropy exceeds H_0, $L\left(\partial_k\right)$ will lower ∂_k, thereby directing the policy training towards prioritizing value improvement.

V. Simulation Results and Discussions

In this section, we first introduce the setting of the simulation experiment in detail, including the environment and algorithm parameters. Secondly, we conduct a large number of experimental results to demonstrate the process of multi-AUV data collection tasks and verify the feasibility of proposed MAISAC algorithm. Finally, we compare the MAISAC with the baseline algorithm in different task environments, and investigate the effect of learning rate on the performance of our algorithm, which proves the superiority of MAISAC.

A. Experiment Settings

1) Environment Parameters: The size of the experiment site is $L \times L$, the initial $L = 120$m, and the depth is -50m.

TABLE I
Simulation Environment Parameters.

Parameters	Values
Value range of L	$[60\text{m}, 240\text{m}]$
The number of IoUT devices	$[30, 120]$
Device transmit power	30 mW
Transmit frequency f	20kHz
Transmit bandwidth w	1kHz
Spreading factor ς	1.5
Shipping activity factor s	0.5
AUV sailing height z	-10 m
Value range of the communication radius R_r	[6m,8m]
Capacity of data buffer $Q_{\max}$	2Mbits
Maximum package data buffer C	5000
Maximum velocity of AUV $v_{\max}$	2 m/s
Locations of vortex centers	$(30\text{m}, 23\text{m}), (63\text{m}, 45\text{m}), (75\text{m}, 77\text{m})$
Strength of the vortex Γ	8
Radius of the vortex δ	48 m
Number of training epochs ε	450
Crash distance d_{safe}	5 m
discount factor γ	0.99
Learning rate λ	0.0003
Target entropy H_0	2.0
Soft updating rate τ	0.01
Value range of penalty weight a	$[0, 0.42]$

The experimental environment is divided into turbulence-free environment and turbulent environment. When the turbulent environment is considered, there are several vortex distributions, whose centers are distributed at (30m,23m), (63m,45m) and (75m,77m), with an intensity Γ of 8 and a radius δ of 48m. Initially, 45 IoUT devices are randomly distriuted in the environment, and their $\lambda_i(t)$ are randomly selected from the set $\{3, 5, 8, 12\}$. In the initial case, there are two AUVs at a certain height to collect data of devices with 6m as the communication radius. In addition, motion parameters, system communication parameters, etc., are detailed in Table. I.

2) Algorithm Parameters: In the training stage, the total duration of each epoch is $T = 1000$s, the step $\Delta T = 1$s, the network learning rate λ is set to 3×10^{-4}, and the discount factor γ is assigned to 0.99. To facilitate network updates, the soft update coefficient τ is set to 0.01, and the initial value of the entropy α is 0.2. In addition, the initial value of penalty alpha is 0.03, which affects the strategy of AUV in selecting target device, which will be discussed in detail later, and other algorithm parameters are summarized in Table. I.

B. Simulation Results and Analysis

Based on the DTDE framework, multiple AUVs are trained with MAISAC algorithm to optimize their respective policies. At the beginning of each epoch, the position of AUVs and the status of devices will be reset. Each AUV selects the target device according to (3). Then, the AUV navigates to the target device for data collection. Once the data of the device is collected, the AUV will reselect the target device and repeat the above process until the task is completed. In order to study the influence of environment on simulation results and to demonstrate the adaptability of our proposed algorithm, experiments are conducted respectively in turbulence-free environment and turbulent environment, where the AUV

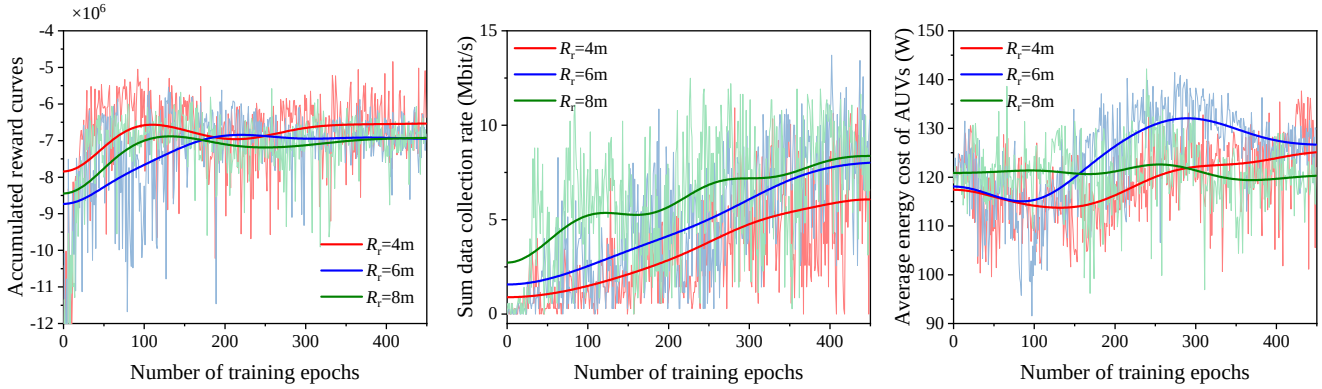


Fig. 2. Training process of the multi-AUV data collection in the turbulence-free environment via MAISAC: (a) Accumulated reward curves. (b) Sum data collection rate. (c) Average energy cost curves.

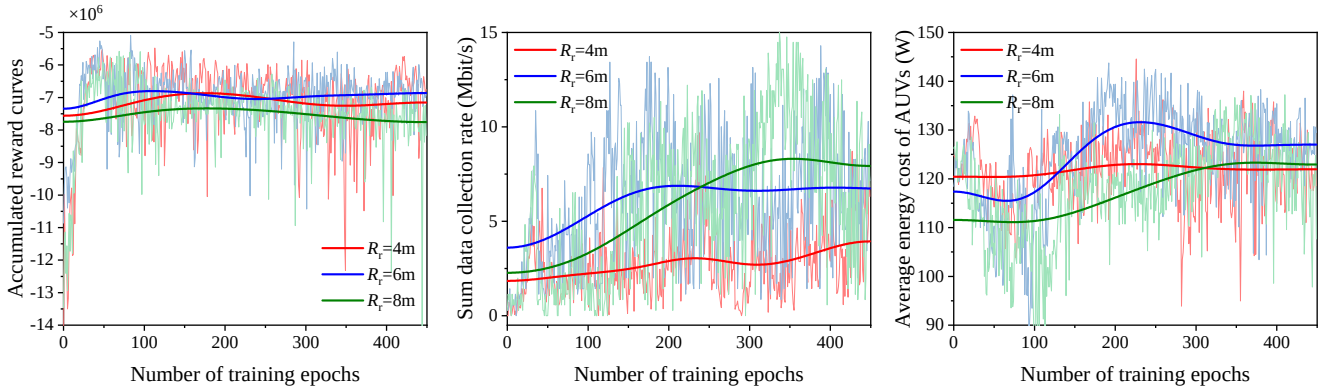


Fig. 3. Training process of the multi-AUV data collection in the turbulent environment via MAISAC: (a) Accumulated reward curves. (b) Sum data collection rate. (c) Average energy cost curves.

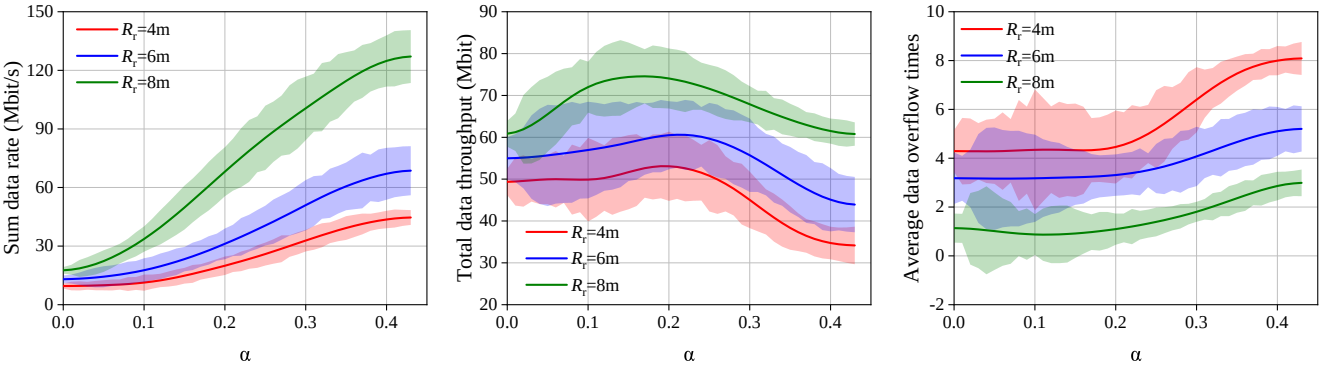


Fig. 4. The change of performance indicators versus α. (a) Sum data rate; (b) Total data throughput. (c) Average data overflow times.

communication range R_r is gradually expanded from 4m to 8m. The experimental results are shown in Figs. 2 and 3.

Upon the cumulative reward curves in Figs. 2(a) and 3(a), the policies of AUVs are not perfect in the initial training stage, resulting in a large number of trials and errors. Through extensive interaction with the environment, the reward curves gradually converge, indicating that the data collection policies of AUVs eventually converge to the level of experts. Considering the environmental impact, the cumulative reward and sum data rate in the turbulent ocean environment are generally lower than that in the turbulence-free environment, and the fluctuation is more obvious. It is worth noting that our proposed algorithm successfully achieves the tradeoff of multiple optimization objectives. To illustrate this, taking the case of $R_r = 6$m in Fig. 3 as an example, AUVs takes reaching the target devices as the highest priority goal at the cost of energy before 300 epoches, as can be seen from the rising energy consumption curve in Fig. 3(c). After this phase, AUVs recalibrate their policies and turn their attention to improving energy efficiency while continuing to improve the sum data rate. In the subsequent training process, our algorithm can still maintain a balance between the optimization objectives,

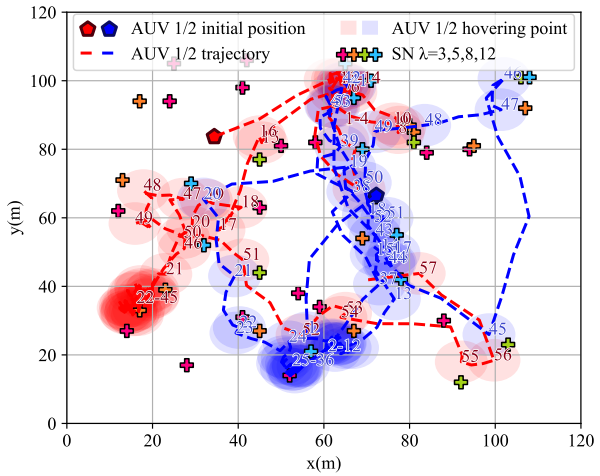


Fig. 5. Trajectories of AUVs for the data collection task in the turbulence-free environment.

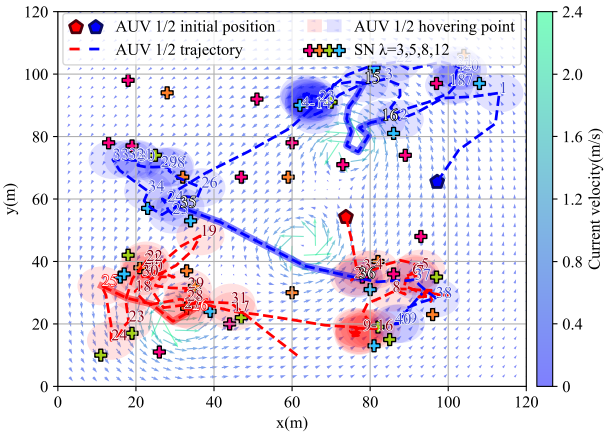


Fig. 6. Trajectories of AUVs for the data collection task in the turbulent environment.

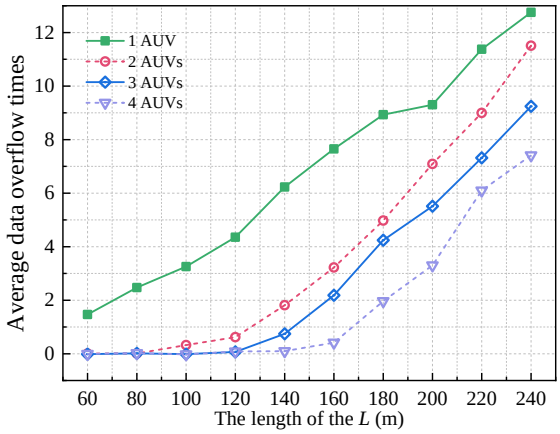


Fig. 7. The average data overflow time versus the number of devices and the number of AUVs.

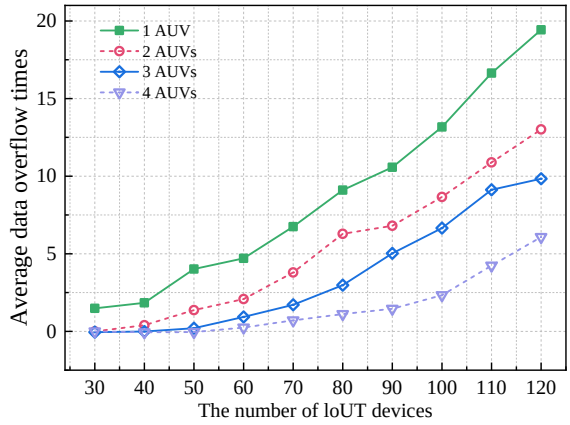


Fig. 8. The average data overflow time versus the number of devices and AUVs.

and this robustness is particularly significant when cumulative reward fluctuations are observed.

The distance penalty term a in (3) affects AUV's decision to select target device. To evaluate the impact of a on algorithm performance, we use the trained expert policy to evaluate the relationship between $a \in [0, 0.42]$ and three performance indicators: sum data rate, total data throughput and average time of data overflow. To ensure accuracy, we performed 40 Monte Carlo samples for each a value, and the results are shown in Fig. 4. Figs. 4(a) and 4(c) show that a is positively correlated with the data rate and the average time of data overflow, which means that the sum data rate increases with the increase of a. At the same time, when a increases, AUV prefers to serve the nearest devices, which indirectly leads to the increase of the average time of data overflow. In addition, Figs. 4(b) and 4(c) reveal that when $a < 0.2$, the total data throughput is positively correlated with a while the average time of data overflow remains relatively stable. When $a > 0.2$, the two indicators deteriorate rapidly. In order to achieve a trade-off between multiple optimization objectives, we set a to 0.2.

We use the trained expert policies to guide AUVs to perform 600s data collection tasks in turbulence-free environment and turbulent environment, respectively. The motion trajectories of AUVs in the two scenarios are shown in Figs. 5 and 6, respectively.

As can be seen from the trajectories of AUVs in Fig. 5, each AUV plans its data collection trajectory by evaluating the distance and the device's upload urgency. For devices with low data generation rate, their data upload urgency is low, and so the AUV will preferentially visit the devices closer to itself to serve as many devices as possible, which can be seen from the dense hover points in the figure. For devices with high data upload urgency, each AUV will choose the optimal path to ensure that these devices' data can be collected in a timely manner. The figure shows that devices with high data generation rate are accessed, which confirms our analysis. In addition, since our proposed MAISAC algorithm carries out the environment-aware trajectory design, AUVs can be competent for the data collection task in turbulent environment as shown in Fig. 6. When conducting data collection tasks in turbulent environments, AUVs optimize their trajectories to avoid dangerous vortex areas due to higher water velocity around vortices, which increases the energy consumption of

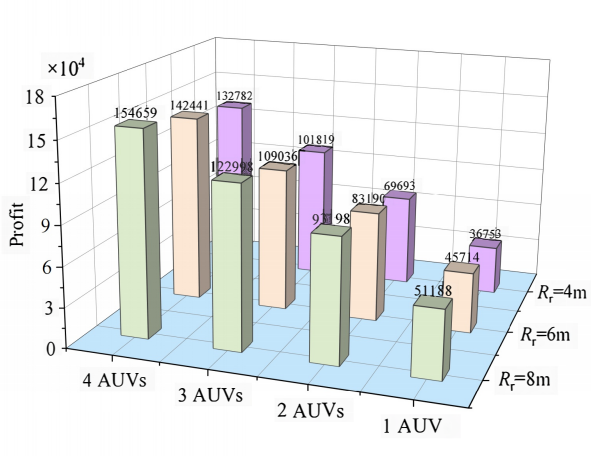


Fig. 9. The average data overflow time versus the number of devices and AUVs.

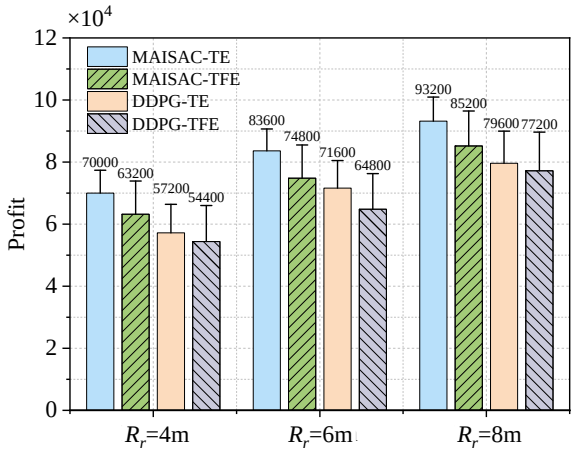


Fig. 10. Profit versus different algorithms under different conditions.

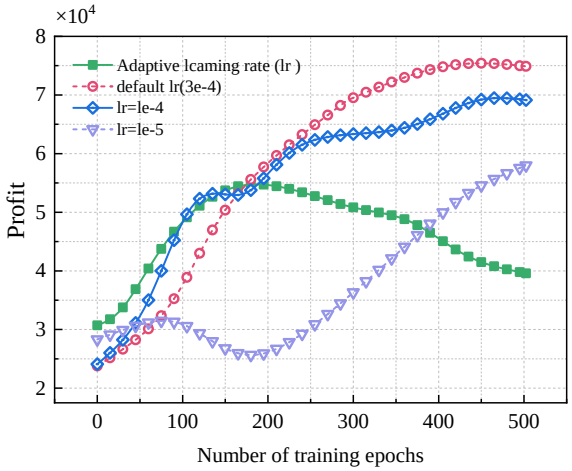


Fig. 11. Impact of the learning rate on the convergence performance of MAISAC.

AUVs. The above analysis proves that our proposed MAISAC not only has high data collection efficiency, but also has the ability of environment-aware trajectory design, and has excellent performance and adaptability.

Subsequently, in order to evaluate the timeliness of data collection of our proposed scheme under different task scenarios, the changes of the average data overflow times with the site size and the number of devices are shown in Fig. 7 and Fig. 8, respectively. On the one hand, the average data overflow times increases with the size of the site, as shown in Fig. 7. As the average distance between devices increases, AUVs cannot timely respond to the upload requirements of the target devices. When the number of AUVs increases, the data overflow phenomenon is effectively alleviated. In particular, the collaboration of four AUVs to collect data can ensure that no data overflow occurs when the site side length is less than 160m. On the other hand, the average data overflow times increases with the increase of devices, as shown in Fig. 8. The reason is that the number of devices that generate upload demands at the same time is increasing, and the ability of AUVs to handle devices at the same time is limited. When the number of participating AUVs gradually increases, this phenomenon can be significantly improved. For example, when the number of devices is 120, the data overflow time of four AUVs is reduced from about 19 to 7 times compared with that of a single AUV. Based on the above analysis, it can be concluded that multi-AUV collaboration can effectively improve the timeliness of data collection and be competent for complex tasks.

According to Eq. (38a), the higher the total data collection rate, the larger the total data collection volume, and the smaller the total energy consumption, the greater the profit. Fig. 9 shows the profits of different numbers of AUVs in different communication ranges. It can be concluded that with constant communication coverage radius, the more AUVs, the greater the profits. This is because the system's ability to respond to multiple tasks at the same time becomes stronger, and AUVs can choose the optimal target device to collect data. In the case of the same number of AUVs, the larger the communication coverage radius, the greater the profit, because as the communication radius increases, the AUV can improve the data collection rate and reduce energy consumption by optimizing the hovering point.

To further highlight the advantages of MAISAC proposed in this paper, we compare the total profit obtained by MAISAC-based optimization scheme with that obtained by DDPG-based optimization scheme in turbulent environment (TE) and turbulence-free environment (TFE) respectively, and the results are shown in Fig. 10. It can be seen that the profits obtained by the two algorithms in the turbulent environment are respectively smaller than those in the turbulence-free environment, which is in line with the actual situation. The turbulent environment increases the energy consumption of AUV, interferes with the motion of AUV, increases the instability of the environment, reduces the learning efficiency of reinforcement learning to a certain extent, and ultimately results in the decline of profits. In addition, it can be found that the performance of the proposed MAISAC algorithm is better than that of the DDPG algorithm in any environment, and the error is smaller.

In reinforcement learning, the setting of key hyperparameters such as learning rate is very important, which seriously affects the performance of the algorithm. Considering that MAISAC deals with dynamic environments with multiple AUVs, we optimize the learning rate. Generally speaking, too high a learning rate leads to instability of the model and even gradient explosion or gradient disappearance. On the contrary, network parameters are updated slowly and convergence is slow. In Fig. 11, we compare the performance of algorithms using adaptive learning rate and several fixed learning rates. It can be seen that MAISAC using adaptive learning rate will gradually adjust the learning rate according to the training process, so as to achieve a higher level of convergence.

VI. Conclusion

In this paper, an environment and energy aware multi-AUV assisted IoUT data collection scheme is proposed, aiming at efficient data collection of IoUT devices with different states in turbulent environment. This problem is a multi-objective optimization problem, that is, for IoUT devices, their data must be collected in time to prevent overflow, and for AUVs, they must plan reasonable paths to avoid the vortex areas and reduce energy consumption, and ultimately make the system the most profitable. For this high-dimensional NP-hard problem, we formulate constrained optimization problem and propose a MAISAC algorithm to train each AUV to make the best decision based on DTDE. The simulation results show that the proposed scheme can complete the data collection task well both in turbulence-free and turbulent environments, and can also take into account the timeliness of data collection, the sum data collection rate, the total data throughput, AUV energy consumption and operation safety when the task conditions become complicated. In addition, our proposed MAISAC algorithm has high adaptability.

In the near future, we plan to investigate the data collection task in large-scale IoUT. A limited number of AUVs may not respond timely to the data upload requests from numerous sensor nodes. Therefore, it is beneficial to explore the hierarchical data collection mechanism, this involves initially clustering the nodes and having each cluster head collect data within its cluster, followed by the AUVs collecting data from the cluster heads. Moreover, it is essential to study the simultaneous wireless data and power transfer, allowing AUVs to collect data while simultaneously charging nodes with low battery levels to extend the lifespan of the data collection system.

References

[1] S. Guan, J. Wang, C. Jiang, R. Duan, Y. Ren, and T. Q. S. Quek, "MagicNet: The maritime giant cellular network," *IEEE Commun. Mag.*, vol. 59, no. 3, pp. 117–123, Mar. 2021.

[2] R. H. Jhaveri, K. M. Rabie, Q. Xin, M. Chafii, T. A. Tran, and B. M. ElHalawany, "Guest editorial: Emerging trends and challenges in Internet-of-underwater-things,"*IEEE Internet Things Mag.*, vol. 5, no. 4, pp. 8–9, Dec. 2022.

[3] Z. Wang, Z. Zhang, J. Wang, C. Jiang, W. Wei, and Y. Ren, "AUV-assisted node repair for IoUT relying on multi-agent reinforcement learning," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 4139-4151, Jul. 2023.

[4] R. Zhang, X. Ma, D. Wang, F. Yuan, and E. Cheng, "Adaptive coding and bit-power loading algorithms for underwater acoustic transmissions," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 9, pp. 5798–5811, Apr. 2021.

[5] Q. Guan, F. Ji, Y. Liu, H. Yu, and W. Chen, "Distance-vector-based opportunistic routing for underwater acoustic device networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3831–3839, Jan. 2019.

[6] H. Harb, A. Makhoul, and R. Couturier, "An enhanced K-means and ANOVA-based clustering approach for similarity aggregation in underwater wireless device networks," *IEEE Sens. J.*, vol. 15, no. 10, pp. 5483–5493, Jun. 2015.

[7] G. Han, X. Long, C. Zhu, M. Guizani, Y. Bi, and W. Zhang, "An AUV location prediction-based data collection scheme for underwater wireless device networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6037–6049, Apr. 2019.

[8] R. Duan, J. Du, C. Jiang, and Y. Ren, "Value-based hierarchical information collection for AUV-enabled Internet of Underwater Things," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9870–9883, May. 2020.

[9] S. Yoon, A. K. Azad, H. Oh, and S. Kim, "AURP: An AUV-aided underwater routing protocol for underwater acoustic device networks," *Devices*, vol. 12, no. 2, pp. 1827–1845, Feb. 2012.

[10] 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, Sep. 2022.

[11] M. Huang, K. Zhang, Z. Zeng, T. Wang, and Y. Liu, "An AUV-assisted data gathering scheme based on clustering and matrix completion for smart ocean," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9904–9918, Apr. 2020.

[12] Z. Fang, J. Wang, J. Du, X. Hou, Y. Ren, and Z. Han, "Stochastic optimization-aided energy-efficient information collection in Internet of underwater things networks," *IEEE Internet Things J.*, vol. 9, no. 3, pp. 1775–1789, Jun. 2022.

[13] G. Han, X. Long, C. Zhu, M. Guizani, and W. Zhang, "A high-availability data collection scheme based on multi-AUVs for underwater device networks," *IEEE Trans. Mob. Comput.*, vol. 19, no. 5, pp. 1010–1022, Mar. 2020.

[14] Y. Zhang, Y. Chen, S. Zhou, X. Xu, X. Shen, and H. Wang, "Dynamic node cooperation in an underwater data collection network," *IEEE Sens. J.*, vol. 16, no. 11, pp. 4127–4136, Jun. 2016.

[15] G. Han, Z. Zhou, Y. Zhang, M. Martínez-García, Y. Peng, and L. Xie, "Sleep-scheduling-based hierarchical data collection algorithm for gliders in underwater acoustic sensor networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9466–9479, Sep. 2021.

[16] M. Chaudhary, N. Goyal, A. Benslimane, L. K. Awasthi, A. Alwadain, and A. Singh, "Underwater wireless device networks: Enabling technologies for node deployment and data collection challenges," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 3500–3524, Nov. 2023.

[17] Z. Zhao, C. Liu, X. Guang, and K. Li, "MLRS-RL: An energy-efficient multilevel routing strategy based on reinforcement learning in multimodal UWSNs," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11708–11723, Feb. 2023.

[18] K. Hao, Y. Ding, C. Li, B. Wang, Y. Liu, X. Du, and C. Q. Wang, "An energy-efficient routing void repair method based on an autonomous underwater vehicle for UWSNs," *IEEE Sens. J.*, vol. 21, no. 4, pp. 5502–5511, Oct. 2021.

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

[20] J. Faigl, and G. A. Hollinger, "Autonomous data collection using a self-organizing map," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 5, pp. 1703–1715, Mar. 2018.

[21] S. Mahmoudzadeh, D. M. W. Powers, A. M. Yazdani, K. Sammut, and A. Atyabi, "Efficient AUV path planning in time-variant underwater environment using differential evolution algorithm," *J. Mar. Sci. Appl.*, vol. 17, no. 4, pp. 585–591, Sep. 2018.

[22] S. Mahmoudzadeh, D. M. W. Powers, and A. Atyabi, "UUV's hierarchical DE-based motion planning in a semi dynamic underwater wireless device network," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 2992–3005, Jun. 2019.

[23] P. Yao, Z. Zhao, and Q. Zhu, "Path planning for autonomous underwater vehicles with simultaneous arrival in ocean environment," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3185–3193, Sep. 2020.

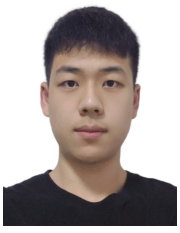
[24] X. Hou, J. Wang, Z. Fang, X. Zhang, S. Song, X. Zhang, and Y. Ren, "Machine-learning-aided mission-critical Internet of underwater things," *IEEE Netw.*, vol. 35, no. 4, pp. 160–166, Oct. 2021.

[25] Z. Fang, J. Wang, C. Jiang, Q. Zhang, and Y. Ren, "AoI-inspired collaborative information collection for AUV-assisted internet of underwater things," *IEEE Internet Things J.*, vol. 8, no. 19, pp. 14559–14571, Oct. 2021.

[26] Y. Yu, J. Tang, J. Huang, X. Zhang, D. K. C. So, and K. K. Wong, "Multi-objective optimization for UAV-assisted wireless powered IoT networks based on extended DDPG algorithm," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6361–6374, Jun. 2021.
[27] J. Wang, S. Liu, W. Shi, G. Han and S. Yan, "A Multi-AUV Collaborative Ocean Data Collection Method Based on LG-DQN and Data Value," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 9086-9106, Mar. 2024.
[28] B. Jiang, J. Du, C. Jiang, Z. Han and M. Debbah, "Underwater Searching and Multiround Data Collection via AUV Swarms: An Energy-Efficient AoI-Aware MAPPO Approach," *IEEE Internet Things J.*, vol. 11, no. 7, pp. 12768-12782, Apr. 2024.
[29] 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, Dec. 2023.
[30] P. Abichandani, S. Torabi, S. Basu, and H. Benson, "Mixed integer non-linear programming framework for fixed path coordination of multiple underwater vehicles under acoustic communication constraints," *IEEE J. Ocean. Eng.*, vol. 40, no. 4, pp. 864–873, Jan. 2015.
[31] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 11, no. 4, pp. 34–43, Oct. 2007.
[32] B. Garau, A. Alvarez, and G. Oliver, "AUV navigation through turbulent ocean environments supported by onboard H-ADCP," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Orlando, FL, USA, 2006, pp. 3556–3561.
[33] L. Shi, R. Zheng, S. Zhang, and M. Liu, "Cooperative estimation to reconstruct the parametric flow field using multiple AUVs," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-10, Nov. 2021.
[34] S. Shuai, and M. H. Kasbaoui, "Accelerated decay of a lamb-oseen vortex tube laden with inertial particles in eulerian-lagrangian simulations," *J. Fluid Mech.*, vol. 936, pp. A8, Feb. 2022.
[35] M. M. Bhatti, M. Marin, A. Zeeshan, and S. I. Abdelsalam, "Recent trends in computational fluid dynamics," *Front. Phys.*, vol. 8, pp. 593111, Oct. 2020.
[36] J. Duan, Y. Guan, S. E. Li, Y. Ren, Q. Sun and B. Cheng, "Distributional Soft Actor-Critic: Off-Policy Reinforcement Learning for Addressing Value Estimation Errors," in *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6584-6598, Nov. 2022.

Zekai Zhang was born in Nanjing, Jiangsu, China, in 2000. He received the B.S. degree in Electronic Engineering from North University of China, Shanxi, China, in 2021. He is currently pursuing the M.S. degree in Electronic Information at Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. His research interests include robot simulation technology, multi-agent cooperation and industrial applications.

Jingzehua Xu was born in Xuzhou, Jiangsu, China in 2001. He received his B.S. degree in Marine Science, and B.E. degree in Electronic Science and Technology from Zhejiang University, Hangzhou, China in 2023. He is currently pursuing the M.S. Degree in Electronic Information from Tsinghua Shenzhen International Graduate School, Tsinghua University, China. His main research interests include deep reinforcement learning and underwater robots. Besides, he is also the outstanding graduate in Zhejiang University.



Guanwen Xie was born in 2002. He is currently pursuing the B.E. degree in Ocean Engineering and Technology at Ocean College from Zhejiang University, and will pursue the M.S. degree in Electronic Information from Tsinghua Shenzhen International Graduate School, Tsinghua University, China. His main research interest is applying reinforcement learning to underwater robots. Besides, he is also the outstanding graduate in Zhejiang University.

Jingjing Wang (S'14-M'19-SM'21) received his B.S. degree in Electronic Information Engineering from Dalian University of Technology, Liaoning, China in 2014 and the Ph.D. degree in Information and Communication Engineering from Tsinghua University, Beijing, China in 2019, both with the highest honors. From 2017 to 2018, he visited the Next Generation Wireless Group chaired by Prof. Lajos Hanzo, University of Southampton, UK. Dr. Wang is currently a professor at School of Cyber Science and Technology, Beihang University. His research interests include AI enhanced next-generation wireless networks, UAV swarm intelligence and confrontation. He has published over 100 IEEE Journal/Conference papers. Dr. Wang was a recipient of the Best Journal Paper Award of IEEE ComSoc Technical Committee on Green Communications & Computing in 2018, the Best Paper Award of IEEE ICC and IWCMC in 2019.

Zhu Han (S'01-M'04-SM'09-F'14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively. From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an assistant professor at Boise State University, Idaho. Currently, he is a John and Rebecca Moores Professor in the Electrical and Computer Engineering Department as well as in the Computer Science Department at the University of Houston, Texas. Dr. Han's main research targets on the novel game-theory related concepts critical to enabling efficient and distributive use of wireless networks with limited resources. His other research interests include wireless resource allocation and management, wireless communications and networking, quantum computing, data science, smart grid, carbon neutralization, security and privacy. Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. Dr. Han was an IEEE Communications Society Distinguished Lecturer from 2015-2018, AAAS fellow since 2019, and ACM distinguished Member since 2019. Dr. Han is a 1% highly cited researcher since 2017 according to Web of Science. Dr. Han is also the winner of the 2021 IEEE Kiyo Tomiyasu Award (an IEEE Field Award), for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."

Yong Ren (M'11-SM'16) received his B.S, M.S and Ph.D. degrees in electronic engineering from Harbin Institute of Technology, China, in 1984, 1987, and 1994, respectively. He worked as a post doctor at Department of Electronics Engineering, Tsinghua University, China from 1995 to1997. Now he is a professor of Department of Electronics Engineering, and the director of the Complexity Engineered Systems Lab (CESL) in Tsinghua University. He holds 60 patents, and has authored or co-authored more than 300 technical papers in the behavior of computer network, P2P network and cognitive networks. He has serves as a reviewer of IEICE Transactions on Communications, Digital Signal Processing, Chinese Physics Letters, Chinese Journal of Electronics, Chinese Journal of Computer Science & Technology, Chinese Journal of Aeronautics and so on. His current research interests include complex systems theory and its applications to the optimization and information sharing of the Internet, Internet of Things and ubiquitous network, cognitive networks and Cyber Physical Systems.