Low-Latency Relay Selection for Ocean Monitoring in Buoy-Based Maritime Wireless Networks

Hanni Yu†, Lijia Wang†, Yi Zhang†‡*, and Yinchao Chen†

†Department of Information and Communication Engineering, School of Informatics, Xiamen University, China
‡Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, China
* Corresponding Author: yizhang@xmu.edu.cn

***Abstract*—In recent years, buoys have demonstrated their significant potential in enabling data forwarding services in maritime wireless networks, which play a critical role in ocean monitoring activities. However, limited energy supply, complex and dynamic channels, and restricted bandwidth remain pressing challenges to the buoy-based maritime wireless network. In this paper, we provide a periodical collect-wait-forward process to coordinate the small-size data uploading from the underwater aggregators. A reinforcement learning-based relay selection approach is proposed to minimize the expected long-term end-to-end packet delay by jointly considering the time-varying energy harvesting and the energy consumption of the buoys. The simulation results show that the proposed approach can achieve significant performance gain over other approaches in terms of learning speed, packet delay, and energy consumption.**

***Index Terms*—Maritime wireless network, offshore buoy, energy harvesting, low latency, reinforcement learning.**

I. Introduction

Maritime activities, from shipping and fishing to offshore energy production and oceanographic research, have become essential drivers of global economic development. The major enabler for these activities is the provision of broadband, low-delay, and reliable wireless coverage to the ever-increasing number of vessels, sensors, and actuators [1]. Unlike terrestrial networks, which benefit from stable infrastructure and predictable conditions, *maritime wireless networks* must contend with various environmental factors, including extreme weather, sea state variations, and long-distance signal attenuation. These factors make maritime communication a complex and demanding field, requiring innovative solutions in network design, communication protocols, and resource management. A series of maritime wireless networks have been studied to support low-latency, high-reliability and energy-constrained communication over long distances [2]–[5].

The *offshore buoy*, which is more like a base station (BS) but is closer to the vessels and sensors, has a calculation and forwarding function. Therefore, the buoy-based maritime wireless network has become a popular choice for ocean monitoring programs due to its low-cost, capability for prolonged deployment [6]. For example, by utilizing a buoy as a relay to facilitate communication between a ground BS and an underwater target, a mixed radio frequency underwater radio-acoustic communication relaying system is studied with the assistance of active reconfigurable intelligent surface (RIS) [7]. However, unlike terrestrial communication infrastructures [8], which rely on direct power transmission via established power grids, such standard power supply solutions are unavailable in maritime environments. An alternative way is to harvest the renewable energy from the surrounding environments. For example, the buoy can harvest energy from the ocean wave by taking advantage of the relative motion between the floating buoy and the submerged body [9].

In ocean monitoring systems, sensors periodically upload data of status information, water quality information, plankton information, and so on. Although those applications typically collect only a few bits from sensors, the data uploading still requires frequent activation of the communication module of the buoy relays, which leads to significant energy consumption [10]. It is worth noting that the energy consumption for sending data is very low compared to the activation of the communication module [11]. As a result, some data aggregation methods and periodic activation transmission schemes have been investigated to avoid frequent switching of the communication state [12]–[15]. However, at the same time, these approaches may introduce significant latency in data uploading. Therefore, the trade-off between energy consumption and the uploading latency remains a critical challenge for efficient ocean monitoring.

Considering the challenges of limited energy availability for the buoy's recharging and the small size of the uploading sensing data, we propose a low-latency maritime wireless network for ocean monitoring via the buoy-based mesh network. The connected buoys, which act as fixed infrastructures in the ocean, receive the sensing data from the underwater aggregators and then forward it to their neighboring buoys or directly upload it to the offshore BS. We provide a periodical collect-wait-forward process to coordinate the data uploading among the buoys and the data aggregators. To support long-term data forwarding services, we model the energy harvesting process of the energy-limited buoys and formulate an expected long-term end-to-end packet delay minimization problem to balance the packet delay and the energy consumption. Furthermore, to tackle the large dimensions of both the state space and action space for all buoys, a reinforcement learning (RL)-based

This work was supported by the National Natural Science Foundation of China under Grant No. 62401485. (Corresponding author: Yi Zhang. Email: yizhang@xmu.edu.cn)

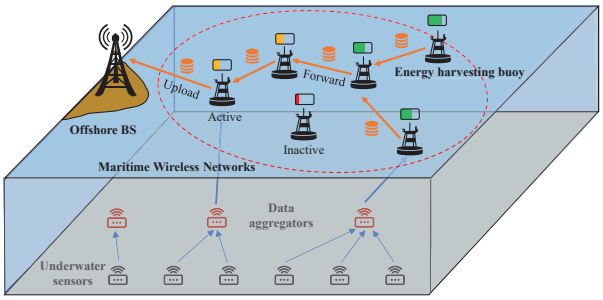


Fig. 1. Buoy-based maritime wireless network for ocean monitoring.

relay selection approach is proposed consisting of an actor-network, two Q-critic networks, and a V-critic network. The BS determines the data forwarding path of the sensing data and broadcasts the relay policy to all buoys. The policy entropy is added to the reward function to improve the exploration efficiency. The simulation results confirmed that our proposed approach can outperform other approaches [16] in terms of learning speed, packet delay, and energy consumption.

II. System Model

The proposed low-latency maritime wireless network for ocean monitoring is shown in Fig. 1. There exists a lot of underwater sensors that periodically upload the sensing data, i.e. hydrological information, to their associated data aggregators $\mathcal{M} = \{1, \cdots, M\}$. These data are packed and forwarded to an offshore BS via the buoy-based mesh network, which consists of a series of connected buoys $\mathcal{N} = \{1, \cdots, N\}$ over a wide sea range. To ensure the stability of the mesh network, the buoys are anchored to the seafloor and therefore act as fixed infrastructures in the ocean. The energy-limited buoys continuously harvest energy from the surrounding environment to support the data forwarding services.

A. Data Uploading Process

In the proposed buoy-based mesh network, each buoy forwards the data to its neighboring buoy or directly uploads to the offshore BS at regular intervals. The periodical collect-wait-forward data uploading process consisting of total K *transmission periods (TPs)* is shown in Fig. 2. Each TP is uniformly divided into $N+1$ transmission cycles and assigned to the connected buoys sequentially to avoid transmission collision. Let $\prec$ denote the transmission order in each TP, we have $i \prec j$ if the transmission cycle reserved for the i-th buoy is earlier than the j-th buoy. Furthermore, we assume $i \prec j$ if $i < j$ with $\forall i, j \in \mathcal{N}$ because the optimization of the transmission order is beyond the scope of this paper. That is, the time interval between $(i-1)\tau$ and $i\tau$ is reserved for the i-th buoy, where τ is the period of a cycle.

Each cycle is further divided into several collection slots by a ratio of $\boldsymbol{\beta} = \{\beta_1, \cdots, \beta_M\}$ to receive the sensing data from the underwater aggregators sequentially, where $\beta_m \in [0, 1]$ with $\forall m \in \mathcal{M}$ and $\sum_{m \in \mathcal{M}} \beta_m = 1$. In this way, each buoy can aggregate the packets and transmit to its next hop, i.e., a neighboring buoy or the BS, at the end of the cycle.

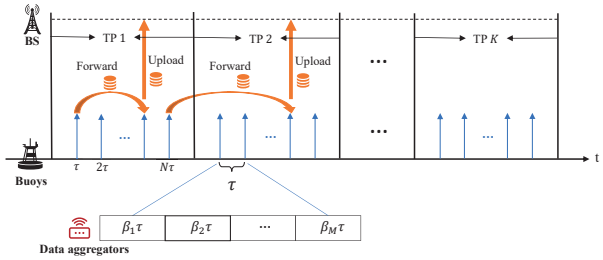


Fig. 2. Data uploading process of the mesh network.

During the k-th TP, the relay selection and transmission power of the connected buoys are denoted by $\boldsymbol{a}^{(k)} = \{a_1^{(k)}, \cdots, a_N^{(k)}\}$ and $\boldsymbol{p}^{(k)} = \{p_1^{(k)}, \cdots, p_N^{(k)}\}$, respectively. The $a_i^{(k)} \in \{0\} \cup \mathcal{N}$ denotes the next hop of the i-th buoy and $p_i^{(k)}$ is its transmission power. Specifically, $a_i^{(k)} = 0$ represents that the i-th buoy can directly transmit to the BS. We denote the residual battery capacity of the buoys as $\boldsymbol{\gamma}^{(k)} = \{\gamma_1^{(k)}, \cdots, \gamma_N^{(k)}\}$. When $\gamma_i^{(k)}$ becomes low, the i-th buoy will switch the state from active to inactive and the remaining packets have to be buffered until the next TP.

B. Channel Model

1) Buoy-to-buoy channel: The buoy-to-buoy channel is quite different from the terrestrial environments due to its dynamic nature and unique oceanic conditions. Let $\boldsymbol{H}^{(k)} = \left[h_{ij}^{(k)}\right]$ denote the channel gain matrix with $i, j \in \mathcal{N}$, where $h_{ij}^{(k)}$ is the channel gain consisting of large-scale and small-scale factors between the i-th and j-th buoys in k-th TP. The log-distance path loss (PL) model, which has been proved to agree well with the large-scale model, can be characterized as a function of the distance between the i-th and j-th buoys d_{ij}:

$$PL(d_{ij}) = PL(d_0) + 10n \log_{10}(d_{ij}/d_0), \tag{1}$$

where d_0 is a reference distance for the antenna far-field to determine $PL(d_0)$ and n is the PL exponent. Note that various distributions are provided to fit the small-scale model, we consider the Rician distribution based on the criteria of Kolmogorov–Smirnov statistics [17].

Therefore, the signal-to-noise ratio (SNR) of the buoy-to-buoy channel is written as

$$\mathrm{SNR}_{ij}^{(k)} = \frac{p_i^{(k)} h_{ij}^{(k)}}{\sigma_j^2}, \tag{2}$$

where $p_i^{(k)}$ is the transmission power of the i-th buoy and σ_j^2 is the noise power at the j-th buoy.

2) Aggregator-to-buoy channel: Considering that the aggregator transmits data to the buoy using acoustic signals, the attenuation of acoustic signals η_{mi} in an underwater environment is given by

$$\eta_{mi} = (\hat{d}_{mi})^{\kappa} q(f)^{\hat{d}_{mi}}, \tag{3}$$

where $\hat{d}_{mi}$ represents the distance between the m-th aggregator and the i-th buoy, $q(f)$ is the absorption coefficient in dB/km calculated by Thorp's formula subject to carrier frequency f

in kHz, and κ is the path loss exponent reflecting propagation geometry. Therefore, the narrow-band SNR is written as

$$\text{SNR}_m^{(k)} = \frac{\hat{p}_m^{(k)}}{\eta_{mi} N_0 W}, \tag{4}$$

where $\hat{p}_m^{(k)}$ denotes the transmit power of the m-th aggregator, N_0 denotes the power spectral density of ambient noise, and W is the available bandwidth.

Since each aggregator transmits once during a cycle τ, the amount of data transmitted from the m-th aggregator to the i-th buoy in k-th TP containing N transmission cycles can be expressed as

$$L_i^{(k)} = N \cdot \beta_m \tau \cdot W \log\left(1 + \text{SNR}_m^{(k)}\right). \tag{5}$$

C. Delay Model

When the i-th buoy chooses the j-th buoy as the next hop, we have $a_i^{(k)} = j$ with $j \neq 0$. The $a_i^{(k)} = i$ means that there is not any relay node selected by the i-th buoy. We introduce $\mathbb{I}(\cdot)$ as the relay selection indicator of the i-th buoy:

$$\mathbb{I}\left(a_i^{(k)}, j\right) = \begin{cases} 0, & \text{if } j \neq a_i^{(k)}, \\ 1, & \text{if } j = a_i^{(k)}, \end{cases} \quad \forall j \in \mathcal{N}. \tag{6}$$

According to the data uploading process, as shown in Fig. 2, we consider the following two relaying cases: *1)* when $i \prec j$, the packet from the i-th buoy can be further forwarded/uploaded by the j-th buoy in the current TP; *2)* when $j \prec i$, the packet has to be buffered at the j-th buoy until the next TP. Using Eq. (6), the amount of data buffered at the i-th buoy before the transmission point in the k-th TP can be represented by

$$\begin{aligned} \rho_i^{(k)} = &\sum_{i \prec j,\, \forall j} \mathbb{I}\left(a_j^{(k-1)}, i\right) \rho_j^{(k-1)} + L_i^{(k-1)} \\ &+ \sum_{j \prec i,\, \forall j} \mathbb{I}\left(a_j^{(k)}, i\right) \rho_j^{(k)}. \end{aligned} \tag{7}$$

We represent the one-hop packet delay experienced from the i-th buoy to its next hop, saying the j-th buoy, as follows:

$$T(i,j) = \begin{cases} 0, & \text{if } j = 0, \\ (j-i)\tau, & \text{if } i \prec j, \\ \left[(N+1) - (i-j)\right]\tau, & \text{if } j \prec i. \end{cases} \tag{8}$$

In this way, the end-to-end average packet delay in the k-th TP can be calculated by

$$D^{(k)} = \frac{\sum_{i \in \mathcal{N}} \left[\rho_i^{(k)} \cdot T(i, a_i^{(k)})\right]}{\sum_{i \in \mathcal{N}} \rho_i^{(k)}}. \tag{9}$$

D. Energy Model

To support long-term data forwarding services, the buoys continuously harvest energy from the surrounding environment. The harvested energy exhibits variability and can be modeled probabilistically to reflect the stochastic nature of maritime environmental conditions. We can assume that the energy harvesting process follows an exponential distribution with an expected value of λ [9].

We assume each buoy has an energy queue to store available energy used for relay transmissions. The energy queue dynamics of the i-th buoy in k-th TP is written as

$$\begin{aligned} &\gamma_i^{(k)} = \\ &\min\left[\max\left(0, \gamma_i^{(k-1)} + \mu_i^{(k-1)} - c_i^{(k-1)} - p_i^{(k-1)}\epsilon\right), B_{\max}\right], \end{aligned} \tag{10}$$

where $B_{\max}$ is the maximum battery capacity of an arbitrary buoy, $\mu_i^{(k)}$ is the energy harvested of the i-th buoy in the k-th TP, and ϵ is the time duration when transmitting an arbitrary packet. The $c_i^{(k)}$ is the energy consumption of activating the communication model, which can be written as

$$c_i^{(k)} = \left(1 - \mathbb{I}(a_i^k, i)\right)\chi, \tag{11}$$

where χ is the energy consumption to activate the communication module once. Accordingly, the average energy consumption of the buoys in the k-th TP will be

$$E^{(k)} = \sum_{i \in \mathcal{N}} \left(c_i^k + p_i^k \epsilon\right). \tag{12}$$

E. Problem Formulation

In the proposed maritime wireless network, the objective is to minimize the expected long-term end-to-end packet delay by scheduling the relay selection $\boldsymbol{a}^{(k)}$, transmission power of the energy-limited buoys $\boldsymbol{p}^{(k)}$, and the ratio of collection slots $\boldsymbol{\beta}$, which is given by

$$\min_{\boldsymbol{a}^{(k)}, \boldsymbol{p}^{(k)}, \boldsymbol{\beta}} \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} D^{(k)} \tag{13}$$

$$\text{s.t.} \quad \sum_{m \in \mathcal{M}} \beta_m = 1 \tag{14}$$

$$p_i^{(k)} \epsilon \leq \gamma_i^{(k)}, \ \forall i, k \tag{15}$$

$$p_i^{(k)} \geq \mathbb{I}\left(a_i^{(k)}, j\right) \cdot \frac{\sigma_j^2}{h_{ij}^{(k)}} \cdot \text{SNR}_{\text{min}}, \ \forall i, j, k \tag{16}$$

$$a_i^{(k)} \in \{0\} \cup \mathcal{N}, \ \forall i, k,$$

$$\beta_m \in [0,1], \ \forall m \in \mathcal{M}$$

The constraint (14) indicates that each cycle consists of several collection slots allocated to the aggregators. The constraint (15) guarantees that the energy consumption for data forwarding does not exceed the residual battery capacity. The constraint (16) derived from (2) ensures reliable data transmission between the i-th buoy and its next hop, that is, $\text{SNR}_{ij}^{(k)} \geq \text{SNR}_{\text{min}}$, where SNR_{min} is the minimum SNR for signal identification.

III. RL-Based Relay Selection

In the proposed buoy-based maritime wireless network, note that the residual battery capacities of the connected buoys vary with not only the data forwarding services but also the energy harvesting process. Due to the large dimensions of both the state space and action space for all buoys, we propose a **RL**-based **R**elay **S**election approach, named by *RLRS*, to minimize

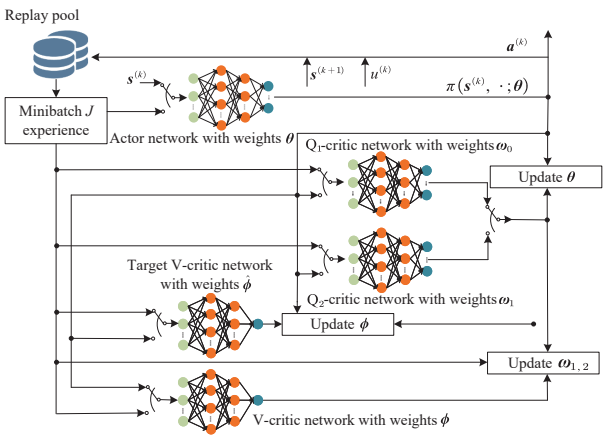


Fig. 3. Illustration of the *RLRS* approach for buoy-based maritime network.

the expected long-term end-to-end packet delay by jointly considering the packet delay and the energy consumption.

A. Network Structure

In this approach, the BS formulates the next hop relay strategy for all buoys based on the channel gain, the packet delay, the number of buffered packets, and the residual battery capacity of each buoy. Given the large dimensions of the state space and action space for all buoys, a soft actor-critic framework is provided in the *RLRS* approach, as illustrated in Fig. 3, consisting of four types of networks: 1) An actor-network derives the optimal relay policy without quantization error; 2) Two Q-critic networks estimate the Q-value corresponding to each state-action pair, guiding the direction to update the actor-network; 3) A V-critic network estimates the state value which helps to mitigate potential overestimations by the Q-critic networks, thereby stabilizing the learning process. 4) A target V-critic network estimates the target V-value for more stable learning. To improve the efficiency of policy exploration, policy entropy is added to the reward function, promoting the exploration of a wider range of potential policies before converging to an optimal solution.

B. State

In the k-th TP, the i-th buoy estimates the channel gain $\boldsymbol{h}_i^{(k)} = \{h_{i1}^{(k)}, \cdots, h_{iN}^{(k)}\}$ with other buoys, measures the position information $\boldsymbol{z}_i^{(k)} = \{z_x^{(k)}, z_y^{(k)}\}$ and the residual battery capacity $\gamma_i^{(k)}$ and sends the status messages to the BS through the feedback channel. Upon receiving the message from all buoys, the BS formulates the maritime wireless network current state $\boldsymbol{s}^{(k)}$ given as

$$\boldsymbol{s}^{(k)} = \left[\boldsymbol{H}^{(k)}, \boldsymbol{Z}^{(k)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\rho}^{(k)}, D^{(k-1)}\right], \tag{17}$$

where $\boldsymbol{H}^{(k)}$ is the estimated channel gain matrix among all buoys with dimension N^2-N, $\boldsymbol{Z}^{(k)} = \{\boldsymbol{z}_1^{(k)}, \cdots, \boldsymbol{z}_N^{(k)}\}$ is the buoys position with dimension $2N$, $\boldsymbol{\gamma}^{(k)} = \{\gamma_1^{(k)}, \cdots, \gamma_N^{(k)}\}$ is the residual batteries capacity, $\boldsymbol{\rho}^{(k)} = \{\rho_1^{(k)}, \cdots, \rho_N^{(k)}\}$ is the number of buffered packets and $D^{(k-1)}$ is the average packet delay of the previous TP.

Algorithm 1: RL-Based Relay Selection (*RLRS*)

Initialize $\boldsymbol{\theta}, \boldsymbol{\omega}_i, \boldsymbol{\phi}$ and $\mathcal{D}$
for $k = 1, 2, 3, \ldots$ **do**
  Estimate channel gain $\boldsymbol{H}^{(k)}$ of all buoys
  Formulate state $\boldsymbol{s}^{(k)}$ via (17)
  Input $\boldsymbol{s}^{(k)}$ to actor network and obtain $\pi\left(\boldsymbol{s}^{(k)}, \cdot; \boldsymbol{\theta}\right)$
  Select relay vector $\boldsymbol{a}^{(k)}$ base on $\pi\left(\boldsymbol{s}^{(k)}, \cdot; \boldsymbol{\theta}\right)$
  Transmit the relay decision to buoys
  Receive the packets from buoys and get $\boldsymbol{Z}^{(k)}$ and $\boldsymbol{\gamma}^{(k)}$
  Measure $\boldsymbol{\rho}^{(k)}$ and $l^{(k)}$
  Calculate $u^{(k)}$ via (18)
  Formulate the next state $\boldsymbol{s}^{(k+1)}$
  Store $\{\boldsymbol{s}^{(k)}, \boldsymbol{a}^{(k)}, u^{(k)}, \boldsymbol{s}^{(k+1)}\}$ in $\mathcal{D}$
  Sample a minibatch $\mathcal{B}$ from $\mathcal{D}$ randomly
  Update the actor-network weights $\boldsymbol{\theta}$ via (20)
  Update the V-critic network weights $\boldsymbol{\phi}$ via (19)
  Update the $Q_{i=1,2}$-critic network weights $\boldsymbol{\omega}_i$ via (21)
  if $k \bmod \chi = 0$ **then**
    $\hat{\boldsymbol{\phi}} \leftarrow \iota\boldsymbol{\phi} + (1-\iota)\,\hat{\boldsymbol{\phi}}$
  end
end

C. Action

As shown in Fig. 3, the BS inputs state $\boldsymbol{s}^{(k)}$ to the actor network with weights $\boldsymbol{\theta}$ and obtains the relay policy probability distribution $\pi\left(\boldsymbol{s}^{(k)}, \cdot; \boldsymbol{\theta}\right)$. Specifically, the structure of the actor-network consists of four fully connected (FC) layers, including an input layer with $N^2 + 3N + 1$ neural nodes, two hidden layers with $f_{1,1}$ and $f_{1,2}$ nodes, and an output layer with N dimension. The relay decision vector $\boldsymbol{a}^{(k)} = \{a_1^{(k)}, \cdots, a_N^{(k)}\}$ is chosen based on the actor-network outputs $\pi\left(\boldsymbol{s}^{(k)}, \cdot; \boldsymbol{\theta}\right)$.

D. Utility

The BS transmits the relay policy to each buoy. Each buoy packages the sensor data from the aggregators, then receives and aggregates packets from other buoys, and finally forwards the packet to the next hop or upload to the BS according to the relay policy. At the end of the TP, using Eqs. (9)(12), the BS measures the utility of the buoy-based network as

$$u^{(k)} = -D^{(k)} - wE^{(k)}, \tag{18}$$

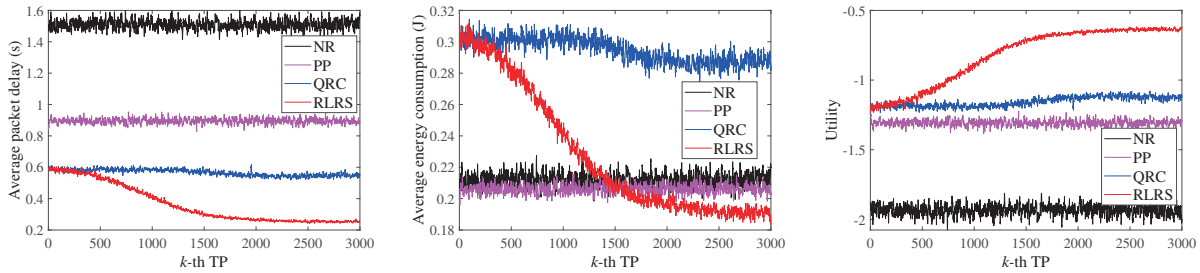
where w denotes the importance of the packet delay and the buoy energy consumption.

E. Update

The overall description of the proposed *RLRS* approach is shown in Algorithm 1. The BS formulates the next state $\boldsymbol{s}^{(k+1)}$ and stores the relay experiences as $\boldsymbol{b}^{(k)} = \left(\boldsymbol{s}^{(k)}, \boldsymbol{a}^{(k)}, \boldsymbol{s}^{(k+1)}, u^{(k)}\right)$ into the replay pool $\mathcal{D}$. A minibatch $\mathcal{B}$ with J experiences are randomly sampled from the replay pool $\mathcal{D}$ to update the weights of the actor-network $\boldsymbol{\theta}$ via policy gradient by

$$\boldsymbol{\theta} \leftarrow \arg\max_{\boldsymbol{\theta}'} \mathbb{E}_{\boldsymbol{e}^{(j)} \in \mathcal{B}} \Big(Q_0\left(\boldsymbol{s}^{(j)}, \pi\left(\boldsymbol{s}^{(j)}, \cdot; \boldsymbol{\theta}'\right); \boldsymbol{\omega}_0\right) - \alpha \ln \pi\left(\boldsymbol{s}^{(j)}, \cdot; \boldsymbol{\theta}'\right)\Big), \tag{20}$$

$$\boldsymbol{\phi} \leftarrow \arg\min_{\boldsymbol{\omega}'} \mathbb{E}_{\boldsymbol{e}^{(j)} \in \mathcal{B}} \left(V\left(\boldsymbol{s}^{(j)}; \boldsymbol{\phi}\right) - \sum_{\boldsymbol{a}' \in \pi\left(\boldsymbol{s}^{(j)}; \boldsymbol{\theta}\right)} \min_{i=0,1} Q_i\left(\boldsymbol{s}^{(j)}, \boldsymbol{a}'; \boldsymbol{\omega}_i\right) - \alpha \ln \pi\left(\boldsymbol{s}^{(j)}, \boldsymbol{a}'; \boldsymbol{\theta}\right) \right)^2 \tag{19}$$



(a) Average packet delay $D^{(k)}$ (b) Average energy consumption $E^{(k)}$ (c) Utility $u^{(k)}$

Fig. 4. Performance of the data uploading process: $N = 5$, $\lambda = 0.14$.

where $\alpha \in [0, 1]$ represents the importance between the Q-value estimated by the critic network and the policy entropy.

The minibatch $\mathcal{B}$ is also used to update the $Q_{i=1,2}$-critic networks with weights $\boldsymbol{\omega}_{i=1,2}$ and V-critic network with weights $\boldsymbol{\phi}$ via the stochastic gradient descent by minimizing the mean square error between the target and estimated Q-value. For example, the Q_1-critic network is structured with four FC layers. The input layer contains N^2+4N+1 neurons, representing the input as the concatenation of state and action, followed by two hidden layers with $f_{2,1}$ and $f_{2,2}$ neurons, respectively. Finally, the output layer with N dimensions is the value of all actions in the given state. For the V-critic network, the structure of V-critic network also consists of four FC layers, including an input layer with $N^2 + 3N + 1$ neural nodes, two hidden layers with $f_{3,1}$ and $f_{3,2}$ nodes and an output layer with 1 dimension is the value of the state. The Q_2-critic network and target V-critic network have the same network structures as the Q_1-critic network and V-critic network, respectively.

The V-critic network is updated by (19) and the $Q_{i=1,2}$-critic network is updated by

$$\boldsymbol{\omega}_i \leftarrow \arg\min_{\boldsymbol{\omega}'} \mathbb{E}_{\boldsymbol{e}^{(j)} \in \mathcal{B}} \left(u^{(j)} + \delta \hat{V}\left(\boldsymbol{s}^{(j+1)}; \hat{\boldsymbol{\phi}}\right) - Q_i\left(\boldsymbol{s}^{(j)}, \boldsymbol{a}^{(j)}; \boldsymbol{\omega}_i\right) \right)^2, \tag{21}$$

where δ is the discount factor to balance the received utility and the future utility in the learning process. To slowly track the weights of the V-critic network, the weights of V-target critic networks $\hat{\boldsymbol{\phi}}$ are updated via the soft update strategy with learning rate ι.

IV. Simulation Results

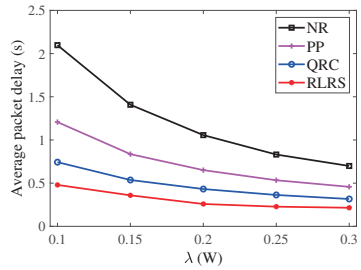
In this section, simulation results are provided to verify the performance of the proposed *RLRS* approach. We consider a maritime wireless network, the BS is located at (0, 0) in meter (m), 5 connected buoys and 5 aggregators are uniformly distributed on a 3km×3km plane. The transmission period of packets TP is set to 300ms, a cycle of the period is set to 60ms, and the time duration when transmitting an arbitrary packet ϵ is set to 10ms. Some key simulation parameters are listed in Table I. We compare four approaches: 1) *No Relay (NR)*, in which the buoys upload the packet to the BS directly; 2) *Proximity Principle (PP)*, in which each buoy forward the packet to its closest neighboring buoy or upload to the BS; 3) *Q-learning-based Routing Control (QRC)*, which adaptively evaluates paths depending on the flow deadline constraints and carefully designs reward function to meet the individual flow deadlines [16]; 4) *RLRS*, the proposed approach. Note that the NR approach may suffer from high latency and low reliability due to the unbalanced traffic and battery power limitations. The PP approach leverages the proximity advantage to reduce transmission but may lead to battery drain of certain buoys performing frequent relays.

TABLE I
List of Key Simulation Parameters

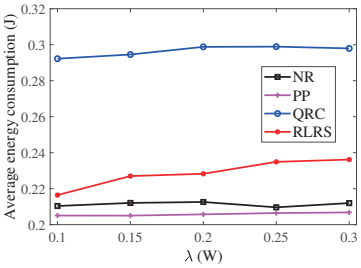
Parameter	Value
Channel bandwidth W	10MHZ
SNR threshold SNR_{min}	20dB
Noise power σ^2	-104dBm [4]
Maximum buoy battery capacity $B_{\max}$	10J
Energy harvesting parameter λ	14W [6]

As shown in Fig. 4, our proposed *RLRS* approach reduces average packet delay by 35.1% from 0.55s to 0.25s with 22.3% less average energy consumption over 3000 TPs compared with *QRC* after 2000 TPs. The is because the policy entropy can find a better balance between exploring new actions and existing experience and the neural network accelerates the learning of buoys relay selection. The *NR* and *PP* approaches could not dynamically adjust relay decisions based on the current buoy status, resulting in highest packet delay.

The performance versus energy harvesting parameter λ and the number of buoys is shown in Fig. 5 and Fig. 6, respectively. In general, the proposed *RLRS* approach can outperform other

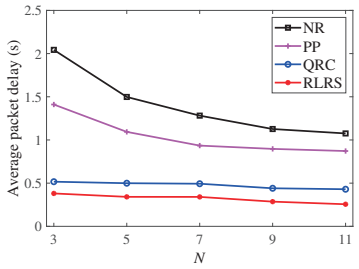


(a) Packet delay $\sum_k^K D^{(k)}/K$

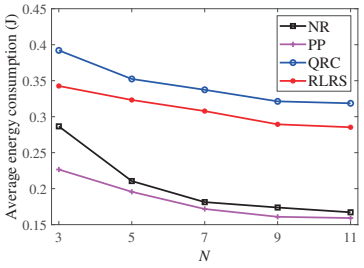


(b) Energy consumption $\sum_k^K E^{(k)}/K$

Fig. 5. Performance versus energy harvesting parameter λ: $K = 3000$.



(a) Packet delay $\sum_k^K D^{(k)}/K$



(b) Energy consumption $\sum_k^K E^{(k)}/K$

Fig. 6. Performance versus the number of buoys: $K = 3000$.

approaches in term of the average packet delay. Specifically, compared with *QRC* in Fig. 5, the proposed *RLRS* approach decreases 36.1% average packet delay and 26.4% average energy consumption when $\lambda = 0.1$. As shown in Fig. 6, with the increase of the buoys, the proposed *RLRS* approach decreases 40.8% average packet delay and 11.1% average energy consumption compared with *QRC* in 11-buoys system.

V. Conclusions

A buoy-based maritime wireless network is studied for ocean monitoring and a periodical collect-wait-forward process is provided to coordinate the data uploading among the buoys and the data aggregators. To support long-term data forwarding services, a RL-based relay selection approach is proposed to minimize the expected long-term end-to-end packet delay. The simulation results confirmed that our proposed approach can outperform other approaches in terms of learning speed, packet delay, and energy consumption.

References

[1] F. S. Alqurashi, A. Trichili, N. Saeed, B. S. Ooi, and M.-S. Alouini, "Maritime communications: A survey on enabling technologies, opportunities, and challenges," *IEEE Internet of Things Journal*, vol. 10, no. 4, pp. 3525–3547, 2023.

[2] C. Liu, Y. Zhang, G. Niu, L. Jia, L. Xiao, and J. Luan, "Towards reinforcement learning in uav relay for anti-jamming maritime communications," *Digital Communications and Networks*, vol. 9, no. 6, pp. 1477–1485, 2023.

[3] Z. Liu, X. Meng, Y. Yang, K. Ma, and X. Guan, "Energy-efficient uav-aided ocean monitoring networks: Joint resource allocation and trajectory design," *IEEE Internet of Things Journal*, vol. 9, pp. 17 871–17 884, 2022.

[4] Z. Wang, B. Lin, Q. Ye, Y. Fang, and X. Han, "Joint computation offloading and resource allocation for maritime mec with energy harvesting," *IEEE Internet of Things Journal*, vol. 11, pp. 19 898–19 913, 2024.

[5] T. Yang, Z. Cui, A. H. Alshehri, M. Wang, K. Gao, and K. Yu, "Distributed maritime transport communication system with reliability and safety based on blockchain and edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, pp. 2296–2306, 2022.

[6] T. Yang, L. Kong, N. Zhao, and R. Sun, "Efficient energy and delay tradeoff for vessel communications in sdn based maritime wireless networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 3800–3812, 2021.

[7] X. Luo, L. Yang, and A.-A. A. Boulogeorgos, "Performance assessment of active-ris-assisted mixed rf-uac systems," *IEEE Transactions on Vehicular Technology*, pp. 1–13, 2024.

[8] Y. Zhang, J.-H. Liu, C.-Y. Wang, and H.-Y. Wei, "Decomposable intelligence on cloud-edge iot framework for live video analytics," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8860–8873, 2020.

[9] A. Hosseini-Fahraji, P. Loghmannia, K. Zeng, X. Li, S. Yu, S. Sun, D. Wang, Y. Yang, M. Manteghi, and L. Zuo, "Energy harvesting long-range marine communication," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, 2020, pp. 2036–2045.

[10] N. Zhang, W. Wang, X. Xin, Y. Liu, H. Shan, and A. Huang, "Low-delay ultra-small packet transmission with in-network aggregation via distributed stochastic learning," *IEEE Transactions on Communications*, vol. 72, no. 5, pp. 2655–2669, 2024.

[11] T. Zhu, J. Li, H. Gao, and Y. Li, "Data aggregation scheduling in battery-free wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 21, pp. 1972–1984, 2022.

[12] H. Harb, A. Makhoul, D. Laiymani, and A. Jaber, "A distance-based data aggregation technique for periodic sensor networks," *ACM Trans. Sen. Netw.*, vol. 13, no. 4, Sep. 2017.

[13] X. Zhou, Y. Li, D. He, C. Zhang, and X. Ji, "Energy-efficient channel allocation based data aggregation for intertidal wireless sensor networks," *IEEE Sensors Journal*, vol. 21, pp. 17 386–17 394, 2021.

[14] J. Wan, J. Wen, K. Wang, Q. Wu, and W. Chen, "Energy-efficient over-the-air computation for relay-assisted iot networks," *IEEE Wireless Communications Letters*, vol. 13, pp. 481–485, 2024.

[15] H. Hong, Y. Zhang, and Y. Xie, "Energy-limited uav visiting planning for age-aware wireless-powered sensor networks," in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, 2023, pp. 1–6.

[16] N. S. Bülbül and M. Fischer, "Reinforcement learning assisted routing for time sensitive networks," *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pp. 3863–3868, 2022.

[17] Q. Zhang, S. Wang, Y. Shi, and K. Yang, "Measurements and analysis of maritime wireless channel at 8 GHz in the south china sea region," *IEEE Transactions on Antennas and Propagation*, vol. 71, pp. 2674–2681, 2023.