Journal of Computers Vol. 35 No. 1, February 2024, pp. 51-63
doi: 10.53106/199115992024023501004

MR-SFAMA-Q: A MAC Protocol Based on Q-Learning for Underwater Acoustic Sensor Networks

Wei-Kai Sun[1*], Xiao-Mei Wang[1], Bin Wang[1],

Jia-Sen Zhang[2], Hai-Yang Du[2]

[1] Information System Engineering College, PLA Strategic Support Force Information Engineering University, Zhengzhou, China
[2] School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou, China
swk19950112@163.com

Received 11 May 2023; Revised 24 August 2023 Accepted 9 October 2023

Abstract. In recent years, with the rapid development of science and technology, many new technologies have made people's exploration of the ocean deeper and deeper, and due to the requirements of national defense and marine development, the underwater acoustic sensor network (UASN) has been paid more and more attention. Nevertheless, the underwater acoustic channel has the properties of considerable propagation delay, limited bandwidth, and unstable network topology. In order to improve the performance of the medium access control (MAC) protocol in UASN, we propose a new MAC protocol based on the Slotted-FAMA of Multiple Reception (MR-SFAMA) protocol. The protocol uses the Q-Learning algorithm to optimize the multi-receiver handshake mechanism. The current state is judged according to the received node request, and the Q-table is established. Through the multi-round interaction between the node and the environment, the Q-table is continuously updated to obtain the optimal strategy and determine the optimal data transmission scheduling scheme. The reward function is set according to the total back-off time and frame error rate, which can reduce the packet loss rate during network data transmission while reducing the delay. In addition, the matching asynchronous operation and uniform random back-off algorithm are used to solve the problem of long channel idle time and low channel utilization. This new protocol can be well applied to unstable network topology. The simulation results show that the protocol performs better than Slotted-FAMA and MR-SFAMA regarding delay and normalized throughput.

Keywords: underwater acoustic sensor network, medium access control, multiple reception, Slotted-FAMA, Q-Learning

1 Introduction

UASN has a wide range of applications in defense and industry fields. Its development level is related to the realization of the full utilization of marine resources and the military game between countries. At present, the international situation is complex and changeable, and the demand for UASN is increasing day by day. In order to gain the initiative in the complex underwater environment battlefield, underwater combat will inevitably transition from standalone combat to system-supported group combat, in which the UASN plays an important role.

MAC is one of the core protocols of UASN, which determines the channel usage and distribution mode in UASN. Compared with terrestrial wireless sensor networks, MAC protocol for UASN faces the following difficulties in design [1]:

(1) The propagation delay of the underwater acoustic link is extended. In underwater environments, acoustic waves are used as a propagation medium, resulting in a propagation delay in the water five orders of magnitude higher than land. Therefore, the MAC protocol suitable for land radio frequency networks has the problem of low throughput and high delay in UASN.

(2) The available bandwidth resources of underwater acoustic links are limited. In wireless communication, bandwidth division is the division of signal frequency. In the underwater environment, the available frequency band range is small because the high-frequency signal has a high power attenuation rate. It will be further decreased with the increase in transmission distance.

(3) The Multipath effect and Doppler effect in the underwater environment make the underwater acoustic link

* Corresponding Author

unstable, and the node drift or the movement of the underwater autonomous vehicle also directly affects the on-off change of the network link.

The competitive MAC protocol based on the handshake mechanism has played a significant advantage in the performance of UASNs, which can effectively alleviate the above problems. The MR-SFAMA protocol proposed in Reference [2] has an ideal effect. MR-SFAMA protocol achieves multiple data by requiring the sink node to receive as many RTS (Request To Send) control packets as possible in one-time slot, thus overcoming the disadvantages of Slotted-FAMA [3], such as low throughput and poor fairness. However, in terms of data transmission scheduling calculated by receiving nodes, new problems are introduced. Due to the unstable network topology of UASN and the significant change of acoustic velocity with the environment, the probability of data transmission collision is greatly enhanced. Therefore, we use the Q-Learning algorithm to optimize the data transmission scheduling scheme, adopt the time-slot asynchronous operation, and apply a new random back-off algorithm [4, 5]. We call the new protocol MR-SFAMA-Q (SFAMA of Multiple Reception based on Q-Learning). The status of our work is that we have used the Q-Learning algorithm to optimize the data transmission scheduling scheme in the protocol and designed a suitable asynchronous time slot scheme and random back-off algorithm. The simulation results show that compared with Slotted-FAMA and MR-SFAMA, MR-SFAMA-Q significantly improved throughput and delay. In summary, our main technical achievements are as follows:

(1) The Q-Learning algorithm is used to optimize the data transmission scheduling scheme of the multiple reception mechanism handshake protocol. The learning environment of communication nodes in UASN is mapped, and the state set, action set, and value function are defined. The optimal strategy is obtained by establishing a Q-table to obtain the optimal data transmission scheduling scheme;

(2) In order to improve the channel utilization, the total back-off time and the frame error rate are used as the evaluation indexes of the reward function. The total back-off time can intuitively reflect the quality of the back-off mechanism, and the inclusion of the frame error rate can effectively reduce the packet loss rate during network data transmission;

(3) The time slot asynchronous operation and the new random back-off algorithm are used to adapt to the protocol, which solves the problem of long channel idle time and low channel utilization.

The rest of this paper is organized as follows: Section 2 mainly analyzes the work related to MAC protocol for UASN based on the existing literature. Section 3 presents the model of UASN and some hypotheses. MR-SFAMA-Q protocol and related issues are described in Section 4. Section 5 covers the verification and performance comparison analysis of the simulation platform. A brief conclusion is given in section 6.

2 Related Works

In this section, the research related to the topic of this paper is proposed, and the related work is compared to distinguish it from the work of this paper.

2.1 MAC Protocols for UASN

As one of the core protocols of UASN, MAC protocol is responsible for using and allocating channels, which guarantees efficient communication of UASN. Whether the MAC protocol can use the limited frequency spectrum reasonably and efficiently directly affects the performance of UASN.

Generally, the MAC protocols are divided into two basic mechanisms [6] according to the allocation policies: fixed allocation MAC and competitive MAC. The fixed allocation MAC divides the channels from time, frequency, and space perspectives. Because UASN has the characteristics of narrow bandwidth, time extension, and complex synchronization, it is not suitable for using a fixed MAC allocation mechanism [7]. Therefore, most studies are designed based on competitive MAC mechanisms. Competing MAC is a network node that sends data by preempting or reserving the channel. The mainstream competitive MAC protocols include random competition MAC protocol and MAC protocol based on handshake reservation.

The most easily implemented competitive MAC protocol is ALOHA [8], first developed by the University of Hawaii in the United States. It has low communication delay and good communication performance when the communication network load is small. Based on this, the UW-ALOHA [9] protocol considers the possible distribution of underwater nodes and selects the binary index and Poisson's escape scheme, which is better applied in USAN.

To solve the problem of hidden terminals, the node applying CSMA listens to the channel before sending data packets. If the channel is idle, the node will send data packets. In ALOHA-CS [10], a new contention window accommodates variable propagation delays. The window size is between two and five times the maximum propagation delay. As soon as the node senses that the channel is idle, it transmits data, and the unsuccessful transmission is assigned a random back-off time. This provides a particular avoidance of communication collision. However, because of the longer delay and narrower bandwidth in the underwater acoustic channel, new problems of hidden terminal and exposed terminal are brought.

The competitive MAC protocol based on the handshake mechanism plays a significant advantage in the performance of UASNs. MACA protocol is the first underwater acoustic MAC protocol using an RTS/CTS handshake mechanism [11]. Bharghavan [12] improved it and proposed the MACAW protocol, which adopted an adaptive back-off algorithm. It added an automatic request retransmission mechanism and adopted RTS-CTS-DS-DATA-ACK control packet mode. It effectively solves the problem of high collision rate and packet loss rate.

In addition, Molins and Stojanovic [13] proposed an improved Slotted-FAMA protocol based on FAMA, which set the RTS and CTS (Clear To Send) control packet length and divided the time into time slots for the characteristics of UASN to reduce the impact of propagation delay on performance.

The Slotted-FAMA protocol only allows one pair of sending-receiving nodes to access the channel in one data transmission cycle, dramatically affecting the performance of UASN, such as throughput and delay. Zhang [14] proposed a MAC protocol for UASN based on data link based on Slotted-FAMA, named SFAMA-DT, which improves the channel utilization rate by forming data packet sequences of multiple transmission pairs during each round of simultaneous handshake. The problem of multiple RTS attempts of Slotted-FAMA in high-traffic environments is overcome, significantly reducing the relative proportion of time wasted due to control packet propagation delay.

Lin [2] proposed the MR-SFAMA protocol, which allows the sink nodes to receive multiple RTS control packets in one-time slot. By controlling the scheduling time of packets sent by its neighbor nodes, the possibility of collision is effectively reduced, and the throughput is better improved. However, its data transmission scheduling algorithm does not consider the adverse factors such as the everchanging topology structure of UASN and the immense change of acoustic velocity with the environment, which significantly enhances the probability of data transmission collision, resulting in the reduction of network data transmission throughput, unstable work efficiency, and other problems. To solve this problem, we use the Q-Learning algorithm to optimize the data transmission scheduling scheme, adopt the time-slot asynchronous operation, and use the new random back-off algorithm [15]. The improved MAC protocol has better performance and stability.

2.2 Summary

As one of the core protocols of UASN, MAC protocol is responsible for the use and allocation of channels, which is the guarantee of efficient communication of UASN. Whether the MAC protocol is able to use the limited frequency spectrum reasonably and efficiently directly affects the performance of UASN.

Table 1. MAC protocols for UASN

Protocol	Topology	Synchronization	Advantage	Disadvantage
UW-ALOHA	Distributed	Yes	The principle is simple and easy to implement.	It is unsuitable for large UASNs, and the throughput is challenging to improve.
ALOHA-CS	Distributed	No	It can use long propagation delay to improve network throughput.	Underwater listening mechanism leads to high energy consumption of nodes.
MACAW	Multi-hop	No	Using fewer control packets to solve the exposed terminal problem.	The handshake time is longer, leading to the node waiting time extension.
Slotted-FAMA	Distributed	Yes	Reasonable control of packet length and division of time slots reduce the impact of propagation delay on protocol performance.	Frequent control packet interaction leads to low channel utilization.
MR - SFAMA	Centralized	Yes	The multi-receiver mechanism is used to improve network throughput.	It is not suitable for UASN with topology changes.

This section reviews the contention-based MAC protocol, mainly discussing the random contention MAC protocol based on the carrier sensing mechanism and the handshake mechanism. In practical applications, hidden terminals and exposed terminals will bring some challenges to the random contention protocol. The protocol based on the ALOHA variant does not consider the channel state before sending data, resulting in a higher collision probability. The protocol based on the CSMA variant listens to the channel before sending data, which reduces the collision probability. However, due to the long propagation delay in UASN, it may lead to ultralong listening time. In addition, underwater carrier detection is costly and needs to be more suitable. The competition MAC protocol based on handshake can not only effectively avoid the exposed terminal problem but also ensure the stability of information transmission. However, its frequent control packet interaction significantly affects the throughput and delay of the underwater acoustic communication network. The multi-receiver protocol based on handshake is the most efficient solution at present. Based on the existing research, we discuss how to design an excellent data reception scheduling scheme to apply to UASN with topology changes so as to improve the performance of the MAC protocol.

3 System Model and Assumptions

This section will introduce the UASN system architecture and Q-learning algorithm.

3.1 System Model

UASN system is mainly composed of UASN, a Data Transmission Network, and a Management Control Center, among which UASN includes underwater sensor nodes, underwater sink nodes, and surface relay nodes [16], as shown in Fig. 1. The ideal working mode of UASN is described as follows: underwater sensor nodes and underwater sink are arranged in a designated area to form a UASN. Each underwater sensor node collects data within the network, which is first transmitted to the underwater sink node (AUV). After the mobile AUV receives the data sent by the sensor node within the transmission range, it is transmitted to the surface relay node through multiple hops. AUV moves according to the preset path. The surface relay node then sends the data or processed information to the control center through the transmission network. The network adopts the Static routing protocol. The data transmission path in a cluster is set as all child nodes transmit to the central node, and the central node transmits to the sea buoy node.

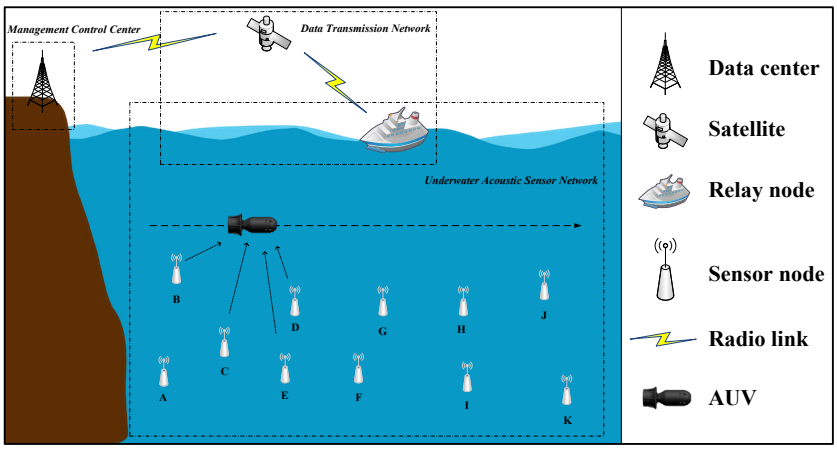

Fig. 1. UASN system

We make the following assumptions:

(1) The control packet transmission process can calculate the propagation delay between each underwater sensor node and the sink node.

(2) Underwater sensor nodes can save recent data transmission situations locally.
(3) Except for data packet collision, all nodes can receive data packets successfully.

3.2 Reinforcement-Learning Technique

Reinforcement learning is one of the paradigms and methodologies of machine learning. Learning strategies are used to maximize reward in the agent interacting with the environment. As shown in Fig. 2, reinforcement learning mainly comprises agents, environment, states, actions, and rewards. After an agent performs an action, the environment transitions to another state, and the environment rewards the conversion. The agent then performs a new action according to the reward strategy based on the new state and environment feedback. When performing an inevitable step, the evaluation of the current agent in this state is mainly represented by the value function, including the state value function and the state-action value function (action-value function for short). Reinforcement learning is a general paradigm using the Bellman equation, which can be simplified as a Markov Decision Process (*MDP*) [17].

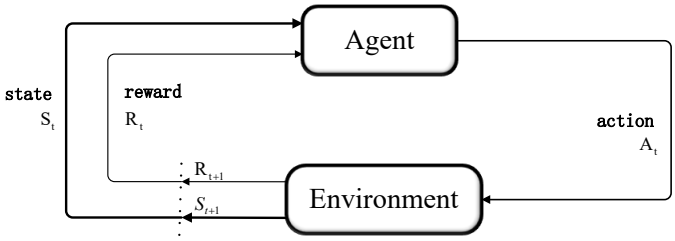

Fig. 2. Reinforcement-Learning technique

$$MDP = (S, A, P^{a_t}_{s_t s_{t+1}}, R)\,. \tag{1}$$

In the formula, S represents the state space set, S={s_1, s_2, …, s_n }, where s_i represents the state of time step i. A denotes the set of action spaces, A={a_1, a_2, …, a_n }, where a_i represents the action of time step i. $Pa_t\ s_t s_{t+1}$ is the state transition probability, representing the probability distribution of moving to another state s_{t+1} after executing an action in the current state s_t. R is the reward function, representing the reward obtained by moving to another state s_{t+1} after executing an action in state s_t.

In fact, solving reinforcement learning is equivalent to optimizing the Bellman equation. The state value function can be divided into two parts: the immediate reward R_t and the discount value of the future state $\gamma V(s_{t+1})$, whose Bellman equation is:

$$V(s_t) = R_t + \gamma \sum_{s_{t+1} \in S} P_{s_t s_{t+1}} V(s_{t+1})\,. \tag{2}$$

Similarly, the Bellman equation of the action value function is:

$$Q(s_t, a_t) = R_t + \gamma \sum_{s_{t+1} \in S} P^{a_t}_{s_t s_{t+1}} Q(s_{t+1}, a_t)\,. \tag{3}$$

The objective of reinforcement learning is to solve the optimal strategy of the *MDP*, and the value function is the expression of the optimal strategy (the optimal strategy is the strategy that maximizes the value function). The optimal value function can represent the optimal strategy.

$$\pi^*(s_t) \to \begin{cases} V^*(s_t) = \max\limits_{\pi} V(s_t) \\ Q^*(s_t, a_t) = \max\limits_{\pi} Q(s_t, a_t) \end{cases}. \tag{4}$$

In other words, if the optimal value function is known, the optimal strategy of *MDP* can be obtained. Therefore, the optimal strategy can be obtained by maximizing.

$$\pi^*(a_t \mid s_t) = \begin{cases} 1, & a_t^* = \underset{a_t \in A}{\operatorname{argmax}}\, Q(s_t, a_t) \\ 0, & \text{else} \end{cases}. \tag{5}$$

Q-learning is a model-free and non-fixed strategy algorithm, which belongs to the Temporal Difference algorithm; when updating the Q-value, the target uses the maximum value of the action value function and is independent of the strategy used when selecting the action [18].

3.3 Analyses and Discussions

The algorithm we designed dynamically adjusts the multi-receive data scheduling strategy by learning from the current communication process to avoid data conflicts in the channel and improve the fairness of indirect channel access of nodes. After the nodes start working, all sending and receiving nodes establish contact through RTS control packets. According to the number of control packets received by the receiving node and the receiving order, it is mapped into a state, and each state establishes a Q-table, which is used to store the order and timing of the sending nodes to send data packets after that. After several rounds of interaction, the data scheduling strategy corresponding to the maximum Q value in the Q-table is selected to improve the network throughput. Because there is an interaction process between the node and the communication environment, the Q-table is constantly updated with the change in the environment. Hence, the algorithm is suitable for UASN with changing network topology.

4 MR-SFAMA-Q Protocol

In this section, we will introduce the MR-SFAMA-Q protocol in detail, including the mechanism of the MR-SFAMA-Q protocol, the design of the reward function, and the convergence property.

4.1 The Protocol Overview

In UASN, each underwater sink node is mapped as an agent of reinforcement learning, and the communication process of the whole network is the learning environment of the agent.

Fig. 3 is an example, and the protocol workflow is as follows:

1) To establish a connection, three asynchronous nodes send RTS control packet requests to the sink node at the beginning of the time slot.

2) Sink receives the first RTS moment as the time-slot start time and receives three RTS control packets in a time-slot, which are node1, node3, and node2. The optimal data transmission scheduling scheme is selected according to the Q-Learning algorithm, and the scheme is sent to all nodes along with CTS in the next time slot.

3) The first node of the data transmission scheduling scheme sends data immediately in the next time slot, and the other two nodes wait for some time to send data, respectively. The node that failed to receive the CTS control packet enters the back-off state. The waiting time of the two nodes is shown in (6) and (7).

$$waitTime\_1 = Delay_T^1 + Delay_{\Pr op}^1 - Delay_T^2. \tag{6}$$

$$waitTime\_2 = Delay_T^2 + Delay_{\Pr op}^2 + waitTime\_1 - Delay_T^3. \tag{7}$$

Where *waitTime_1* and *waitTime_2* are the waiting times of two nodes, respectively, $Delay_T^n$ represents the transmission delay of the nth packet, *and Delay* $_{Prop}^{n}$ represents the propagation delay of the *nth* packet.

4) After receiving all packets or maximum waiting time, the sink broadcasts an ACK packet radio success, and the node that fails to receive the packet enters the back-off.

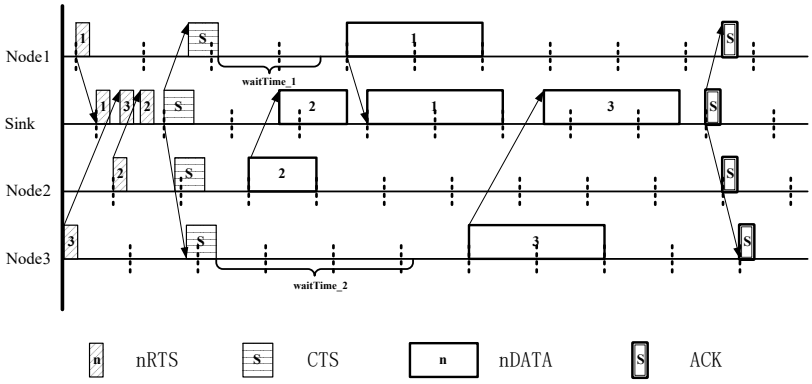

Fig. 3. The procedure of MR-SFAMA-Q protocol

4.2 Transmission Mechanism

This part introduces the working process of the Q-Learning algorithm for data transmission scheduling in MAC protocol.

In the scheme, the corresponding agent is each underwater node in the network, and the learning environment corresponds to the communication process of the whole network. The agent's state $s_t(s_t \in S)$ is the RTS control packet received by the sink node in a time slot (including the total number of RTS control packets and the receiving order). A set of Q-tables is established for all possible data transmission orders, and a data transmission policy is selected as the action $a_t(a_t \in A)$ based on the Q-Learning algorithm.

The environment will generate a reward *Rt* based on the action feedback of the agent. Definition $Q(s_t,a_t)$represents the average reward expectation of the underwater node at time *t* when selecting action a_t in the state s_t. According to the Bellman equation, the Q-table is updated by (8) as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[\mathrm{R}_t + \gamma \max Q(s_{t+1}, a_t) - Q(s_t, a_t)]. \quad (8)$$

Where $\alpha \in (0,1]$ represents the learning rate, its value determines the speed at which the model training can obtain the optimal solution. If the learning rate is too large, the optimal solution may be missed, and the model cannot converge. On the contrary, it will affect the training efficiency of the model. $\gamma \in [0, 1]$ represents the discount rate and maxQ(s_{t+1}, a_t)represents the impact of long-term decisions on current behavior. Represents the maximum expected value corresponding to the new state $s_{t+1}(s_{t+1} \in S)$ entered by the environment under the action of the current action in the Q-table.

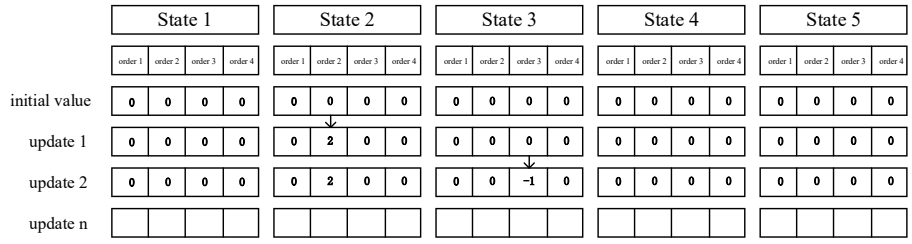

Fig. 4. Updating process of Q-table

The updating process of the Q-table is shown in Fig. 4. According to the different sending requests received by the receiving node, it is represented as a state (i.e., the total number of RTS control packets received and the receiving order). Each state maintains a Q-table, and an action (i.e., data transmission sequence) corresponds to a Q value. After multiple rounds of interaction, the Q-table is updated, and the action with the most significant Q value is performed. The number of states in the Q-table is a crucial parameter in the Reinforcement Learning process regarding memory capacity that should be investigated.

Q-learning can be used to deduce the optimal action strategy of underwater nodes without knowing the system model. When the optimal data transmission sequence strategy is selected, the Q-value is updated according to (8) after this round of data transmission. If $Q^*(s_t\ a_t)$ represents the Q value obtained when the optimal strategy π^* is followed, the optimal strategy can be deduced:

$$\pi^* = \arg\max_{a \in A} Q^*(s_t, a_t) . \tag{9}$$

Q-learning can be used to deduce the optimal action strategy of underwater nodes without knowing the system model. When the optimal data transmission sequence strategy is selected, the Q-value is updated according to (8) after this round of data transmission. If $Q^*(s_t\ a_t)$ represents the Q value obtained when the optimal strategy π^* is followed, the optimal strategy can be deduced:

In order to avoid falling into the suboptimal solution state, the strategy ε-greedy is adopted to adjust in time according to the selected action and feedback reward. The algorithm compromises exploration and utilization based on probability, that is, to explore with the probability of exploration rate ε and to utilize with the probability of 1-ε. When ε is large, the model has better flexibility, can explore the potential higher reward faster, and the convergence speed is fast. When ε is small, the model has better stability and more opportunities to take advantage of the current best reward, but the convergence rate is slow. Based on this, the strategy of underwater nodes performing action is designed as follows:

$$a = \begin{cases} random(Q(s_t, a_t) > 0), & rand \le \varepsilon \\ \max(Q(s_t, a_t)), & rand > \varepsilon \end{cases} . \tag{10}$$

Where rand = random[0,1].

The q-learning algorithm for the data transmission sequence is shown in Algorithm 1.

Algorithm 1. Q-Learning algorithm for data transmission sequence

```
Initialize Q(s_t, a_t), ∀s_t∈S, a_t∈A, arbitrarily, and Q(terminal-state, ·)=0
Receive RTS from nodes
Repeat (for each episode):
    Initialize S
    Repeat (for each step of the episode):
        Choose from s_t using policy ε-greedy from Q
        Take action a_t the observer, s_{t+1}
        Q(s_t, a_t)←Q(s_t, a_t) + α[r + γmaxQ(s_{t+1}, a_t) – Q(s_t, a_t)]
        s_t←s_{t+1}
    Broadcast Q(s_t, a_t) using RTS
    until it is terminal
```

4.3 Asynchrony and Back-off

Relevant studies have shown that, in the ground environment, data packet reception is time-synchronous, and the propagation delay can be ignored. Therefore, data packets from different nodes can be successfully received at the receiver with only a tiny protection time. Under this condition, the channel utilization rate is high. However, if an asynchronous operation is applied, a large number of collisions will occur because packets will collide at the receiving end due to the short time-slot duration. However, in the underwater environment, the time slot length requires more time to complete the control packet and the packet interaction to accommodate the longer

propagation delay [19]. The long propagation delay results in a long idle time on the receiver node, which reduces channel utilization. However, the idle time is often enough to avoid overlapping reception, so the protocol using asynchronous operations is less prone to conflict [20].

The node will enter the back-off state when the data collision occurs and transmission fails. The traditional time-slot protocol back-off algorithm is defined as follows:

$$time\_back-off = \text{int}[\text{CW} \times Random()] \times slot\_time . \tag{11}$$

CW indicates the size of the node back-off window, and slot_time indicates the time-slot length.

In this way, the starting time of each node's slot frame cannot be changed, which is unfavorable to the above asynchronous operation and leads to network convergence failure. Therefore, this protocol adopts a uniform random back-off algorithm, as shown in the Fig. 5. Using this scheme, for each collision, the nodes randomly delay the start time of the next slot according to a uniform distribution.

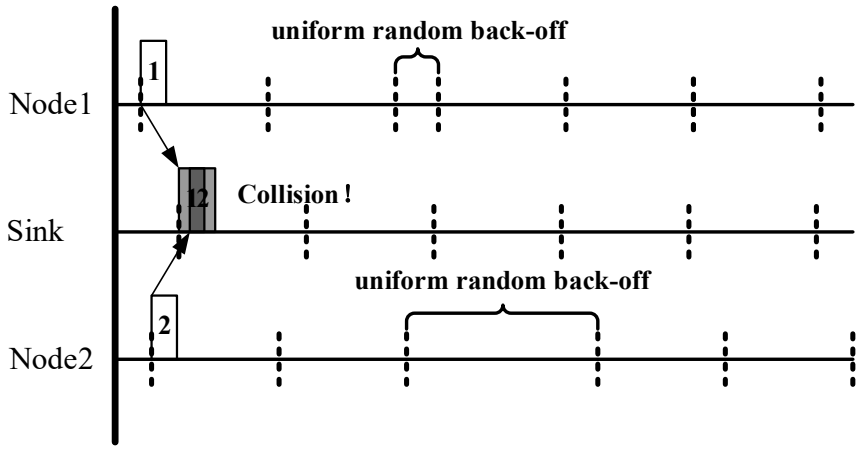

Fig. 5. Uniform random back-off algorithm

4.4 Design of Reward Function

The MAC back-off mechanism in UASN based on Q-Learning aims to improve the channel utilization rate and ensure the fairness of node competition. The design of the reward function is an essential aspect of Q-learning, which directly affects the model's learning efficiency and convergence effect. In order to evaluate the feasibility of action execution, various indexes are introduced into the reward function, including the total time of back-off and frame error rate. Then the reward function of choosing action at in state s_t at time t is defined as follows:

$$R_t = \beta_1 \times r_t(back-off) + \beta_2 \times r_t(FER) . \tag{12}$$

Where $\beta_1, \beta_2 \in (0,1]$.

The total back-off time refers to the time from triggering the back-off mechanism to occupying the channel to send data before each successful data transmission, which can be expressed as:

$$T_t(back-off) = \sum_n back-off\_time_n . \tag{13}$$

Where n indicates the number of conflicts, $back\text{-}off\_time_n$ represents the back-off time of the nth conflict.

The total back-off time can reflect the back-off mechanism. Here, the difference between the total back-off

time at time *t* and the total back-off time at time *t-1* is used as the reward value:

$$r_t(back-off) = T_t(back-off) - T_{t-1}(back-off). \quad (14)$$

However, only using the total back-off time to evaluate the index may increase the packet loss rate, so the frame error rate [21] is also included in the evaluation index. Specifically:

$$r_t(FER) = \ln \frac{\mathrm{FER}_{t-1}}{\mathrm{FER}_t}. \quad (15)$$

In summary, the reward function is defined as follows:

$$R_t = \beta_1 \times [T_t(back-off) - T_{t-1}(back-off)] + \beta_2 \times \ln \frac{\mathrm{FER}_{t-1}}{\mathrm{FER}_t} \quad . \quad (16)$$

4.5 Convergence Property

Convergence is also an essential problem in reinforcement learning algorithms. Watkins and Dayan use a stochastic process and fixed-point theory to give [22]:

1) The learning process is Markov;

2) All state-action pairs can be accessed indefinitely;

3) The learning rate α must meet four value conditions at the same time: $0 \leqslant \alpha \leqslant 1$, $\sum_{t=0}^{\infty} \alpha_t = \infty$ $\sum_{t=0}^{\infty} \alpha_t^2 < \infty$

the learning process can converge to the optimal action-value function $Q^*(s_t, a_t)$. Therefore, we can see that the back-off algorithm satisfies all convergence conditions.

5 Evaluations

5.1 Simulation Settings

The NS-3 discrete event network simulator is used for simulation and verification. The ideal channel model provided in the UAN module is adopted. In order to evaluate the proposed solution, the configuration of simulation network parameters and learning parameters is shown in Table 2. The MAC protocol proposed in this paper mainly solves the problem that the traditional multiple reception handshake protocol has low throughput and poor performance in an unstable topology network. Therefore, the AUV in this paper adopts the simplest linear motion movement model: the roundtrip movement between two points, and the underwater sensor nodes are randomly distributed around the AUV trajectory.

Table 2. Simulation parameters table

Parameter	Value
Node layout range	600m×600m
Depth of nodes	70m
Packet generation rate	80bit/s
Carrier center frequency	12kHz
The velocity of AUV	0.2m/s-2m/s
α	0.87
γ	0.9
ε	0.1
β_1	0.73
β_2	0.27

5.2 Performance Evaluation Criteria

In order to illustrate the performance of the MAC protocol, we use time delay and normalized throughput to evaluate the performance of the MAC protocol in UASN.

1) Delay. *The delay* reflects the timeliness of data transmission. It refers to the time that a packet goes through in the channel from the sending node to the sink, which can be divided into processing delay $Delay_{Proc}$, queue waiting time T_{Qwait}, transmission delay $Delay_T$, and propagation delay $Delay_{Prop}$ according to different stages [23].

$$Delay = Delay_{\Pr oc} + T_{Qwait} + Delay_T + Delay_{\Pr op} . \quad (17)$$

2) Normalized throughput. The throughput represents the number of bits the node successfully sends data frames per unit time [24]. The normalized throughput is the value that normalizes the network's throughput to a value from 0 to 1. It is the most representative performance indicator that reflects the working efficiency of the algorithm and network performance.

5.3 Simulation Results

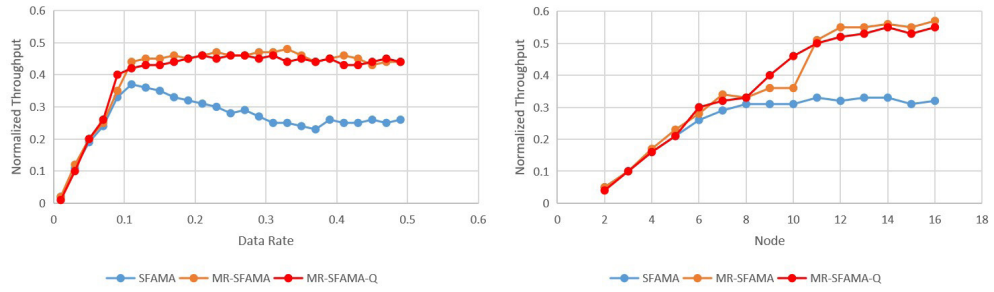

Fig. 6. Normalized throughput performance of 3 MAC protocols without AUV

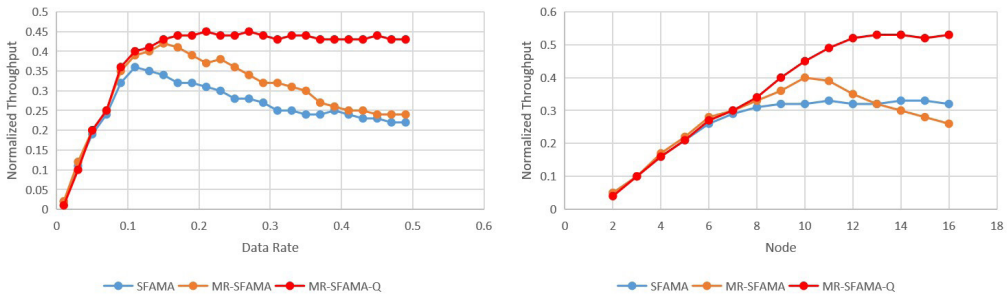

Fig. 7. Normalized throughput performance of 3 MAC protocols with AUV

Fig. 6 and Fig. 7 show the normalized throughput performance of MR-SFAMA-Q compared with Slotted-FAMA and MR-SFAMA without and with AUV, respectively. In the scenario without AUV, all nodes are fixed. It can be found that the throughput of MR-SFAMA-Q is similar to that of traditional multiple reception protocols and sometimes even worse. In the scenario with AUV， when the data rate is less than 0.15, the throughput of the three protocols increases with the data rate increase. When the data rate is more significant than 0.15, the throughput of Slotted-FAMA and MR-SFAMA decreases slowly. This is because the traditional handshake mechanism increases the channel conflict rate under the condition of unstable network topology. In addition, the

throughput of MR-SFAMA-Q increases with the increase in the number of nodes due to the efficient convergence of the Reinforcement-Learning algorithm adopted in this paper.

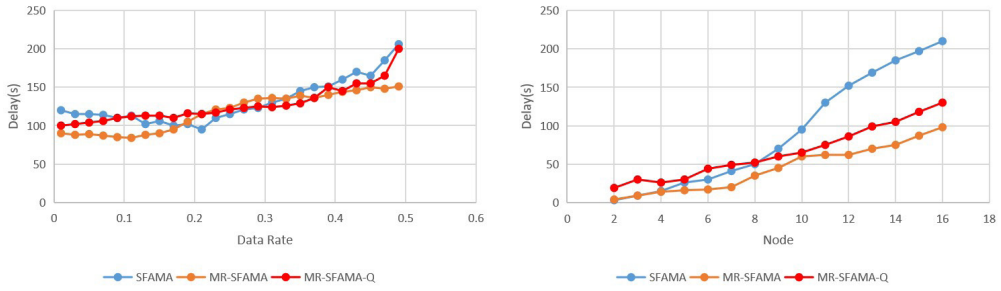

Fig. 8. Delay performance of the three MAC protocols without AUV

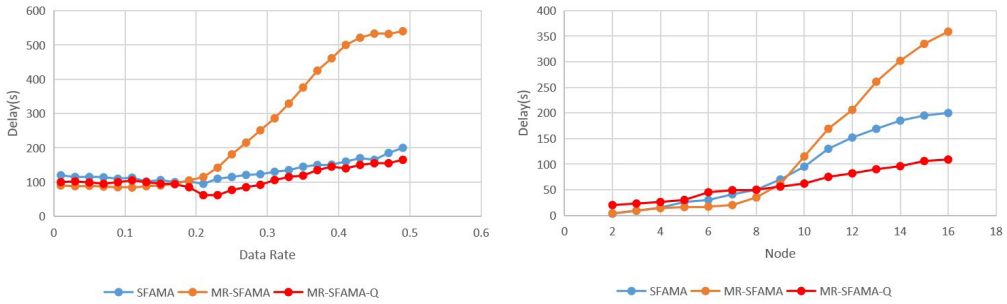

Fig. 9. Delay performance of the three MAC protocols with AUV

Fig. 8 and Fig. 9 show the comparison of the delay performance of the three MAC protocols without and with AUV. MR-SFAMA-Q achieves significant performance improvement in UASN with network topology changes. For example, when the node data transmission rate increases, the delay of MR-SFAMA increases significantly, which is caused by more data collisions. In addition, we can see that as the number of nodes increases, the MR-SFAMA-Q delay of the MR-SFAMA-Q protocol increases more slowly than the other two protocols. However, in the fixed network topology scenario, the traditional multiple reception protocol performs better because the traditional multiple reception protocol has no learning and computing process, and the response is faster.

6 Conclusion

In order to improve the data transmission performance of UASN, we propose a MAC protocol called MR-SFAMA-Q. The reinforcement learning framework based on the Q-Learning algorithm is introduced. According to the different reception requests received by the receiving node, it is expressed as a state. Each state maintains a Q-table. The data scheduling strategy corresponding to the maximum Q value in the Q-table is selected to optimize the data transmission scheduling scheme of multiple receiving nodes, which is suitable for mobile acoustic networks with unstable topology. The total back-off time and frame error rate are used as the criteria for setting the reward function to reduce the network delay and reduce the collision rate. The results show that compared with Slotted-FAMA and MR-SFAMA, the protocol has better performance in terms of throughput and delay. On the one hand, future work will do something within its power from the methodology perspective. On the other hand, it will explore the impact of reinforcement learning on network energy consumption and solve the problem of excessive energy consumption in the process of peers.

References

[1] X. Geng, Y.R. Zheng, Exploiting Propagation Delay in Underwater Acoustic Communication Networks via Deep Reinforcement Learning, IEEE Transactions on Neural Networks and Learning Systems 34(12)(2023) 10626-10637.

[2] L. Wen, MR-SFAMA: A novel MAC protocol for underwater acoustic sensor networks, in: Proc. 2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2015.

[3] L. Qian, S. Zhang, M. Liu, A slotted floor acquisition multiple access based MAC protocol for underwater acoustic networks with RTS competition, Frontiers of Information Technology & Electronic Engineering 16(3)(2015) 217-226.

[4] S.H. Park, P.D. Mitchell, D. Grace, Reinforcement Learning Based MAC Protocol (UW-ALOHA-Q) for Underwater Acoustic Sensor Networks, IEEE Access 7(2019) 165531-165542.

[5] S.H. Park, P.D. Mitchell, D. Grace, Reinforcement Learning Based MAC Protocol (UW-ALOHA-QM) for Mobile Underwater Acoustic Sensor Networks, IEEE Access 9(2020) 5906-5919.

[6] S. Jiang, State-of-the-Art Medium Access Control (MAC) Protocols for Underwater Acoustic Networks: A Survey Based on a MAC Reference Model, IEEE Communications Surveys & Tutorials 20(1)(2018) 96-131.

[7] T. Melodia, H. Kulhandjian, L.-C. Kuo, E. Demirors, Advances in Underwater Acoustic Networking, in: S. Basagni, M. Conti, S. Giordano, I. Stojmenovic, Eds., Mobile Ad Hoc Networking, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013 (pp. 804-852).

[8] L.G. Roberts, ALOHA Packet System With And Without Slots and Capture, ACM SIGCOMM Computer Communication Review 5(2)(1975) 28-42.

[9] Z. Peng, Z. Zhou, J.-H. Cui, Z.J. Shi, Aqua-Net: An underwater sensor network architecture: Design, implementation, and initial testing, in: Proc. OCEANS 2009, 2009.

[10] F. Guerra, P. Casari, M. Zorzi, World ocean simulation system (WOSS): A simulation tool for underwater networks with realistic propagation modeling, in: Proc. of the 4th International Workshop on Underwater Networks WUWNet 2009, 2009.

[11] H.-H. Ng, W.-S. Soh, M. Motani, MACA-U: A Media Access Protocol for Underwater Acoustic Networks, in: Proc. IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference, 2008.

[12] V. Bharghavan, A. Demers, S. Shenker, L. Zhang, MACAW: A media access protocol for wireless LAN's, ACM SIGCOMM Computer Communication Review 24(4)(1994) 212-225.

[13] M. Molins, M. Stojanovic, Slotted FAMA: a MAC protocol for underwater acoustic networks, in: Proc. OCEANS 2006 - Asia Pacific, 2006.

[14] S. Zhang, L. Qian, M. Liu, Z. Fan, Q. Zhang, A Slotted-FAMA based MAC Protocol for Underwater Wireless Sensor Networks with Data Train, Journal of Signal Processing Systems 89(1)(2017) 3-12.

[15] W. Zhang, W. Huang, Y. Chen, X. Xu, A Q-Learning and Data Importance Rating-Based MAC Protocol for Dynamic Clustering Underwater Acoustic Networks, in: Proc. 2022 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2022.

[16] E. Cayirci, H. Tezcan, Y. Dogan, V. Coskun, Wireless sensor networks for underwater survelliance systems, Ad Hoc Networks 4(4)(2006) 431-446.

[17] R. Sutton, A. Barto, Reinforcement Learning: An Introduction, Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1998.

[18] Y.-M. Shi, Z. Zhang, Research on Path Planning Strategy of Rescue Robot Based on Reinforcement Learning, Journal of Computers 33(3)(2022) 187-194.

[19] F. Ahmed, H.-S. Cho, A Time-Slotted Data Gathering Medium Access Control Protocol Using Q-Learning for Underwater Acoustic Sensor Networks, IEEE Access 9(2021) 48742-48752.

[20] G. Han, X. Wang, N. Sun, L. Liu, A Collision-free MAC protocol based on quorum system for underwater acoustic sensor networks, in: Proc. 2020 16th International Conference on Mobility, Sensing and Networking (MSN), 2020.

[21] G. Wang, J. Wu, Y.R. Zheng, An Accurate Frame Error Rate Approximation of Coded Diversity Systems, Wireless Personal Communications 86(3)(2016) 1377-1386.

[22] C. Cortes, V. Vapnik, Support-Vector Networks, Machine Learning 20(3)(1995) 273-297.

[23] X. Guo, M.R. Frater, M.J. Ryan, Design of a Propagation-Delay-Tolerant MAC Protocol for Underwater Acoustic Sensor Networks, IEEE Journal of Oceanic Engineering 34(2)(2009) 170-180.

[24] A. Stok, E.H. Sargent, Comparison of diverse optical CDMA codes using a normalized throughput metric, IEEE Communications Letters 7(5)(2003) 242-244.