IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 May 2022, accepted 24 June 2022, date of publication 5 July 2022, date of current version 11 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3188705

RESEARCH ARTICLE

Reinforcement Learning-Based Power Control for MACA-Based Underwater MAC Protocol

FAISAL AHMED[1], (Student Member, IEEE), JUNHO CHO[2], ETHUNGSHAN SHITIRI[1], (Member, IEEE), AND HO-SHIN CHO[1], (Senior Member, IEEE)

[1]School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, South Korea
[2]Korea Research Institute for Defense Technology Planning and Advancement (KRIT), Jinju-si 52851, South Korea

Corresponding author: Ho-Shin Cho (hscho@ee.knu.ac.kr)

This work was supported by the Project "Development of Distributed Underwater Monitoring and Control Networks" funded by the Ministry of Oceans and Fisheries, South Korea.

ABSTRACT A major amount of the energy of battery-powered sensors is spent during packet transmissions. This issue has led to the development of power-control-based multiple-access collision avoidance (MACA) protocols that can reduce the packet transmission power and conserve energy. However, the reduction in transmission power renders the packets susceptible to collisions. To reduce these collisions while maintaining high energy efficiency, we propose a power control protocol that utilizes reinforcement learning to choose the optimal transmission power. The total reward is determined by the occurrence of a collision, amount of transmission power used, frequency of DATA packet retransmissions, and update of the interference range. A key feature of the proposed protocol is that it enables sensors to prevent collisions without any prior knowledge of interferences, thus eliminating the need for additional signaling. Simulation results under varying average traffic loads indicate that the proposed protocol can improve network throughput by up to 20% compared to benchmark protocols, while minimizing network energy consumption with a similar gain and reducing collisions per packet by more than 35%. These results demonstrate that the proposed protocol is effective.

INDEX TERMS Collisions, interference, medium access control, power control, Q-learning, underwater acoustic sensor networks.

I. INTRODUCTION

Over the past few decades, underwater networks have been significantly developed for practical applications such as marine resource exploration, pollution monitoring, military surveillance, and oceanic environment observations. In particular, underwater wireless acoustic sensor networks (UWASNs) have been proven to be ideal candidates for such applications [1], [2]. Because radiofrequency waves are easily absorbed in an underwater environment owing to the high conductivity of water, acoustic waves are used for communication in UWASNs [3]. However, UWASNs have a limited bandwidth, which prompts the need for resource-efficient medium access control (MAC) protocols. Owing to this limitation and the unique characteristics of underwater acoustic channels, such as a low propagation speed (approximately 1500 m/s in seawater), high bit error rate, low channel capacity, and high dynamics of channel quality, the design of MAC protocol for UWASNs is challenging [1], [2], [4].

One of the primary functions of MAC protocols is to mitigate data packet collisions at the receiver. Among the aforementioned challenges, a low propagation speed is a key factor that renders existing terrestrial MAC protocols unsuitable for UWASNs [3], [4]. Moreover, energy consumption is another major issue as the sensors of UWASNs are battery-powered. Replacing these batteries is challenging and costly because of the harsh nature of water. Thus, energy is a valuable resource. These problems related to the batteries directly affect the lifetime of the network. As such, extending the lifetime of a network is necessary to ensure that it operates smoothly for a long period, thereby reducing the frequency of battery replacements [5]. Importantly, most of the energy consumed

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

by the sensor nodes occurs during packet transmissions [5]. The rate at which this energy is consumed is referred to as the transmission power. Therefore, efficient consumption of energy during packet transmissions is critical for extending the lifetime of the network. Thus, this study focuses on energy-efficient MAC protocols as they are suitable for UWASNs.

MAC protocols are mainly classified as contention-free and contention-based protocols [6]. For underwater MAC protocols, contention-based protocols are preferred as they can fully utilize the underwater channel bandwidth [7]. Contention-based MAC protocols are categorized into handshaking and random-access protocols. The former dominates the latter because of its ability to solve the hidden node problem. Notably, the hidden node problem is a major cause of collision in sensor networks.

Herein, we focus on a well-known group of contention-based MAC protocols for wireless sensor networks, termed multiple-access collision avoidance (MACA)-based protocols. These protocols have the ability to solve the hidden node problem and offer high energy efficiency. Power control MAC protocols [8] and collision avoidance power control (CAPC) MAC protocols [9] are well-known MACA-based energy-efficient MAC protocols. The energy consumption of a network is minimized by allocating the maximum transmit power to the control packets, such as request-to-send (RTS)-clear-to-send (CTS) packets, and the minimum possible transmit power to the DATA and acknowledgement packets. Although transmission power control saves the energy of the sensors, it renders the packets more vulnerable to interference and aggravates the hidden node problem [10]. This problem is commonly referred to as the large interference range collision (LIRC) problem [10] and, if not addressed, can introduce more collisions in the network.

The LIRC problem is illustrated in Figure 1. Let us consider three sensors: a *sender* (S), a *receiver* (R), and an *interferer* (I). In this setting, two ranges are relevant: the *transmission range* (*TR*) and *interference range* (*IR*). *TR* is defined as the range within which a DATA packet can be successfully received and correctly decoded by R (from S). *IR* is defined as the range within which R can be interfered by I, which is outside the *TR* of both S and R. More specifically, a transmission from I received by R can be treated as interference when R cannot correctly decode the received signal because it is below a certain decoding threshold. The solid line circles represent the *TR* of the RTS and CTS, which are transmitted at the maximum transmission power by S and R, respectively. Notably, because I is beyond this *TR*, it will neither receive the RTS nor CTS and therefore cannot enter a sleep state. At this moment, if I has a DATA packet to transmit, it first performs a control packet (that is, RTS) transmission in order to capture the channel, as it assumes that the channel is free. However, R receives a DATA packet from S that is transmitted with a controlled transmission power, and the RTS transmission from I may cause a collision. This is because, the controlled power of the DATA packet could be lower than that of the RTS packet transmitted by I with the maximum transmission power, thus leading to low signal strength. Therefore, the signal-to-noise-interference ratio (SINR) of the DATA packet decreases, making it vulnerable to interference. As mentioned previously, this induced interference is termed the LIRC problem [10]. In Figure 1, the area in the stripes represents the *IR* of R.

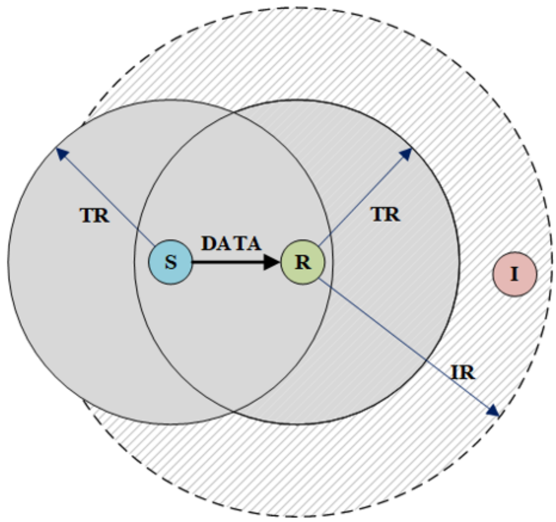

FIGURE 1. LIRC problem of MACA-based MAC protocol. ***S***, ***R***, and ***I*** denote the sender, receiver, and the interferer node, respectively. ***TR*** is the transmission range and ***IR*** is the interference range.

Current solutions that address the LIRC problem require the distance of a potential interferer [9] to realize optimal performance. However, acquiring the distance information of a potential interferer can be challenging owing to propagation delay and can incur high signaling overhead costs. The overhead cost is further amplified when there are more than one interferes. To overcome this issue, we focus on reinforcement learning algorithms and, in particular, the Q-learning algorithm, in which a learner's objective is to maximize rewards by interacting with its environment [11]. It is widely applied in wireless sensor networks because it does not require prior knowledge of the environment or training data to determine the optimal solution. Moreover, it can operate distributively and has low complexity and computational requirements [6].

In this study, we propose a Q-learning-based power control protocol that can optimize the transmission power under multiple interferers for an underwater MACA-based MAC protocol. The proposed protocol enables the sensors to adaptively determine the optimal transmission power level such that the transmitted packets do not undergo collisions. Most importantly, this is achieved without any prior information regarding the interferers. A preliminary version of the current study was presented at the 2021 IEEE Region 10 Symposium (TENSYMP 2021) [12].

The remainder of this paper is organized as follows. Section II provides a discussion on the related works. The system model is described in Section III. After that, a detailed description of the proposed protocol is presented in Section IV. Section V and Section VI discuss about interference range estimation and energy consumption analysis, respectively. Section VII provides the performance evaluation. Finally, in Section VIII we present the conclusion.

II. RELATED WORKS

Power control is a common but important mechanism for both terrestrial wireless sensor networks and UWASNs because it facilitates energy saving and throughput improvement. However, in sensor networks, LIRC problems induced by power control cannot be ignored, as collisions are inevitable. In fact, the LIRC problem is significantly more severe in UWASNs than in terrestrial sensor networks [10]. Several techniques have been discussed in previous studies to address this problem.

In [10], a two-level power control (TLPC) MAC protocol was proposed, where two different transmission power levels were used. RTS/CTS and ACK were sent with a minimum transmission power, while DATA packets were sent with the maximum transmission power. Specifically, RTS/CTS was sent with a transmission power that was equal to one-tenth of the maximum transmission power. Thus, the induced interference range at the receiver could be limited to the sender–receiver distance, thereby minimizing the LIRC problem. Moreover, a grid topology was considered, in which the inter-node distance between the sensors was fixed.

The authors of [9] proposed a CAPC MAC protocol to solve the LIRC problem for terrestrial wireless ad hoc networks. The CAPC MAC protocol employed the maximum transmission power for RTS/CTS transmission and the minimum transmission power for DATA packet transmission to maintaining the required SINR at the receiver. Thus, the LIRC problem was addressed by limiting the interference range to the RTS/CTS transmission range. As mentioned above, the senders require the distance information from the interferers. In this context, the distance is measured using a signaling mechanism that has a low accuracy.

Similar to the method in [9], for the MACA-based power control (MACA-PC) MAC protocol [13] and dual busy tone multiple access (DBTMA) with a power control protocol [14], the authors used the maximum transmission power for RTS/CTS transmission and the minimum required transmission power for DATA packets and ACK transmission. However, a different approach was adopted to avoid the LIRC problem: a notification signal was transmitted with the maximum transmission power along with the DATA packet transmission.

In [8], [15], and [16], the senders periodically transmitted the DATA packets with the maximum transmission power instead of the minimum transmission power for the DATA packets and the maximum transmission power for the RTS/CTS packets. More specifically, the senders increased the transmission power of the DATA packets to its maximum value in a fixed interval within a fixed amount of time. Consequently, the interfering sensors could detect the transmission and deferred their future transmission.

In [17], the authors proposed an adaptive range-based power control (ARPC) MAC protocol in which a range cover mechanism was implemented. The basic idea was to cover the interfering sensors with RTS/CTS transmission so that the LIRC problem could be avoided. Four such mechanisms were proposed for adapting the transmission power of the sensors. Moreover, an adaptive algorithm was employed to select one of the proposed mechanisms based on the distance between the sender and receiver.

Although these studies have been pivotal in solving the LIRC problem, they present some limitations. For example, the TLPC is designed for a fixed grid topology. In practice, the nodes are randomly deployed. Hence, TPLC may perform poorly, thus increasing the energy consumption of the network. CAPC, MACA-PC, DBTMA, and ARPC suffer from high signaling overhead and low energy efficiency. In [8], [15] and [16], concurrent transmission was reduced, which resulted in low throughput, in addition to that, energy consumption was increased due to periodic DATA transmission power increase.

To address the aforementioned gaps, we propose a MAC protocol that does not require the distance information of the interferers and can be deployed in a practical network setup. Aided by Q-learning, the sensors can calculate the optimal transmission power that can mitigate the LIRC problem. Hence, no additional signaling exchange is required, resulting in low overheads. Consequently, the throughput and energy efficiency increase.

III. SYSTEM MODEL

We considered a data-gathering cluster-based network, as shown in Fig. 2. There are three clusters in this network, and each cluster consists of a cluster header (CH) that is located in the center of the cluster. Moreover, multiple sensors are uniformly distributed within a one-hop range of the CH. The sensors collect DATA from the environment and deliver them directly to the CHs. Similarly, the sink gathers DATA from the CHs in a one-hop fashion and then forwards them to the surface buoy. It is assumed that all the sensors have the same battery lifetime, computation ability, and transmission power. Because the sensors are randomly deployed, hidden node problem exists in the network. In this study, we consider both intra-cluster and inter-cluster interferences. In addition, the sensors are considered to be static; therefore their movements, such as those caused by water current and waves, are ignored.

Moreover, we consider the traditional handshaking-based MAC protocol for channel access that employs RTS/CTS/DATA/ACK. Sensors and CHs exchange RTS/CTS with maximum power such that all the other sensors in the transmission range switch to sleep states.

Intuitively, the LIRC problem can be solved by controlling the size of *IR*. This can be achieved by assigning the optimal transmission power for the DATA packet transmissions. This

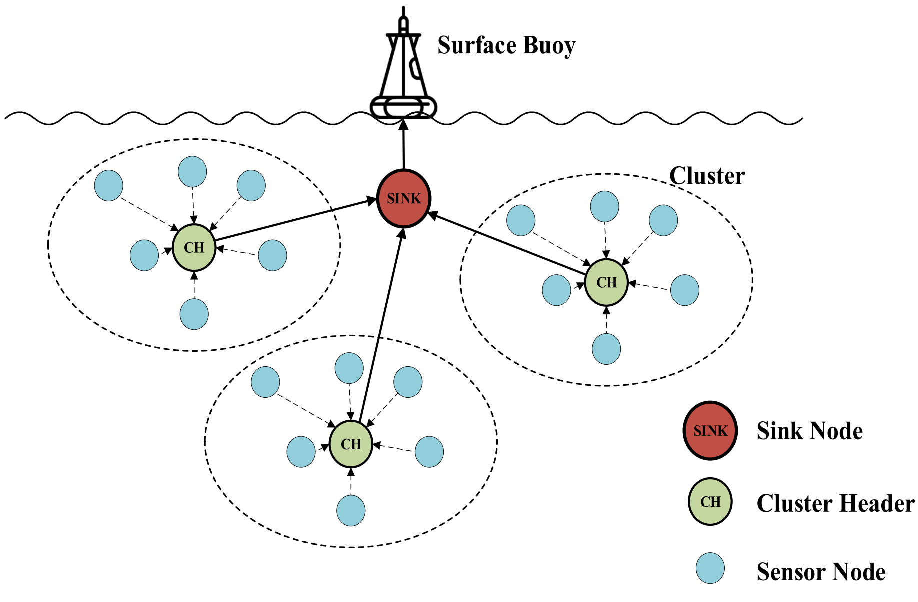

FIGURE 2. Underwater multi-cluster network.

transmission power must satisfy the following condition:

$$SINR_{\mathrm{DATA}} \geq SINR_{\mathrm{TH}}, \tag{1}$$

where $SINR_{\mathrm{DATA}}$ is the signal-to-interference-plus-noise-ratio of the DATA packet, and $SINR_{\mathrm{TH}}$ is the minimum $SINR$ required to successfully decode a received packet. (1) can be reformulated as follows:

$$SINR_{\mathrm{DATA}} = \frac{P_{\mathrm{RX}}}{N+I} \geq SINR_{\mathrm{TH}}, \tag{2}$$

where P_{RX} is the received power level at the receiver, I is the interference signal, and N is the ambient noise.

Let S and R represent the sender and the receiver, respectively. Let us assume that P_{DATA} is the DATA packet transmission power of S. Therefore, according to [18] and [19], we can define P_{RX} as

$$P_{\mathrm{RX}} = \frac{P_{\mathrm{DATA}}}{A(D_{\mathrm{SR}}, f)}, \tag{3}$$

where $A(D_{\mathrm{SR}}, f)$ is the attenuation function considering the acoustic pathloss for a center frequency f and distance D_{SR} between S and R.

Since, the interference occurs beyond the RTS/CTS transmission range, the maximum value of the interference signal can be expressed as

$$I = \frac{P_{\mathrm{MAX}}}{A(D_{\mathrm{IR}}, f)}, \tag{4}$$

where P_{MAX} is the maximum allowable transmission power (defined as a system parameter) and $A(D_{\mathrm{IR}}, f)$ is the attenuation between I and R that are separated by a distance D_{IR}.

According to [18] and [19], the noise N can be approximated as

$$N = 50 - 18\, log(f). \tag{5}$$

Finally, inputting (3), (4), and (5) in (2), we can rewrite (2) as

$$\frac{\frac{P_{\mathrm{DATA}}}{A(D_{\mathrm{SR}}, f)}}{50 - 18\, log(f) + \frac{P_{\mathrm{MAX}}}{A(D_{\mathrm{IR}}, f)}} \geq SINR_{\mathrm{TH}}. \tag{6}$$

When multiple interferes exists, the following condition must be met according to (6) for decoding the received signal:

$$\frac{\frac{P_{\mathrm{DATA}}}{A(D_{\mathrm{SR}}, f)}}{50 - 18\, log(f) + \sum \frac{P_{\mathrm{MAX}}}{A(D_{\mathrm{IR}}, f)}} \geq SINR_{\mathrm{TH}}. \tag{7}$$

IV. PROPOSED Q-LEARNING-BASED POWER CONTROL

In the proposed scheme, the optimal transmission power for DATA packet transmissions is determined using Q-learning. Each sensor in the network acts as an agent of Q-learning and maintains a Q-table. Moreover, the initial Q values are set to 0 and updated according to the reward obtained after executing an action using below equation [11] as follows:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha[r_t + \gamma\ \max_a Q_t(s',a) - Q_t(s,a)], \tag{8}$$

where $Q_{t+1}(s,a)$, $Q_t(s,a)$, r_t, a, s, α, γ is the updated Q value, old Q value, reward, action, state, learning rate ($\in [0,\ 1]$), and discount factor ($\in [0,\ 1]$), respectively.

We define actions as selections of power levels for the DATA packet transmissions. For example, at a time step t, if a sensor j selects a power level P_{DATA}, then the defined action is expressed as $a_{j,t} = P_{\mathrm{DATA}}$. If each sensor employs a discrete power level for P_{DATA}, then, $P_{\mathrm{DATA}} \in \{P_0, P_1, \ldots, P_{M-1}\}$, where P_0 is the minimum transmission power level and M is the number of power levels. The step size between consecutive power levels is a constant denoted as

$$\Delta = P_i - P_{i-1}, \tag{9}$$

where $i = 1, 2, \ldots, M-1$.

Furthermore, we define a state as the receiver of DATA packets. Because each cluster has only one CH that acts as the sole DATA packet receiver, the Q-learning method applied here is a single-state Q-learning method. Several recently published studies, such as, [20]–[22], and [23], related to both underwater and terrestrial MAC protocols have adopted single-state Q-learning because of its low computational complexity. Moreover, in the absence of states, it is relatively easier to design protocols by relying only on action–reward pairs [24].

In the case of single-state Q-learning, only the immediate reward considered as γ is set to 0, given that no future state exists in the system. The rule for updating the Q values is different from (8) and is described as

$$Q_{t+1}(a) = (1-\alpha)Q_t(a) + \alpha r_t. \tag{10}$$

Once a sensor executes an action, it rates the quality of its action by using the total reward r_t. This reward consists of four reward factors, namely, a success/collision-related reward factor (r_1), an efficiency-related reward factor (r_2), a retransmission-related reward factor (r_3), and an interference range-related reward factor (r_4).

The parameter r_1 indicates whether collision has occurred. An ACK is not returned when a collision occurs, thereby indicating that the DATA transmission power is insufficient. In this case, the reward is -1. Conversely, an ACK is returned when there is no collision, indicating that the DATA transmission power is sufficient to avoid collision. In this case, the reward is +1. We can express r_1 as

$$r_1 = \begin{cases} +1, & \text{if ACK is received} \\ -1, & \text{if ACK is not received.} \end{cases} \tag{11}$$

The parameter r_2 represents the goodness of the selected transmission power and is expressed as

$$r_2 = 1 - \frac{P_{\mathrm{DATA}}}{P_{\mathrm{MAX}}}. \tag{12}$$

A sensor receives a lower reward if a higher power is allocated to transmit the DATA packet. This is because the total energy consumption of the network increases with the allocation of a higher DATA packet transmission power. Conversely, a sufficiently low transmission power provides a higher reward.

The parameter r_3 represents the occurrence frequency of the DATA packet retransmissions in the network. A high number of DATA packet retransmissions indicates that, at the receiver, the condition of *SINR* for decoding the received DATA packet is not satisfied. Consequently, the throughput of the network decreases. Therefore, an extremely low number of retransmissions is required. The parameter r_3 can be expressed as follows:

$$r_3 = 1 + \frac{1-\psi}{\psi_{\mathrm{MAX}} - 1}, \quad \psi_{\mathrm{MAX}} > 1 \tag{13}$$

where ψ is the number of transmission of the same DATA packet, and ψ_{MAX} represents the maximum allowable number of retransmission of DATA packets.

r_4 can be written as

$$r_4 = \frac{P_{\mathrm{RX}}}{P_{\mathrm{MAX}}}. \tag{14}$$

A higher value of r_4 indicates a lower range of interference. This can improve both the throughput and energy efficiency of the network by lowering the collision probability.

Therefore, the total reward r_t for learning the minimum transmission power robust to the interference is as follows:

$$r_t = \sum_{i=1}^{4} \beta_i r_i, \tag{15}$$

where β_i is the weighting parameter of r_i, and $\sum_{i=1}^{4} \beta_i = 1$. The four reward factors can be weighted differently depending on the purpose of the network. For example, the settings $\beta_1 > \beta_2$ indicate that more priority is given to collision avoidance rather than energy consumption minimization. This can improve the network throughput at the cost of higher energy consumption. Similarly, the converse is true when $\beta_2 > \beta_1$.

To obtain r_t corresponding to a chosen P_{DATA}, the sensors can update the Q value according to (10). Overtime, the Q value converges and facilitates the determination of the optimal transmission power that is robust to interferences (that is, the goal of this collective reward is to allocate the minimum power that avoids collision). Notably, in this proposed protocol, the term convergence in an optimal transmission power occurs when the Q value of the selected optimal transmission power approaches 1. This is because the reward factors are designed to have a maximum value of 1. In addition, these factors are scaled by the weighting parameter β_i, where $\sum \beta_i = 1$.

In the proposed protocol, ε-greedy [25] is applied to introduce exploration while sufficiently exploiting the single-state Q-learning technique.

A. EXAMPLE

An example of the proposed protocol is shown in Figure 3. When S receives the CTS, it allocates P_{DATA} (action) by using a Q-table (or an arbitrary P_{DATA} if it is the first time). After sending the DATA packet, S listens for the ACK within the ACK timeout duration. Despite the interference(s), if R decodes the DATA successfully, it sends an ACK back to S. Upon receiving the ACK, S updates its Q-table based on the current r_t with the corresponding r_1(=1, ACK received), r_2, r_3, and r_4. However, if the DATA collides at the receiver due to the interference(s), R does not send back the ACK. At the end of the ACK timeout duration, S updates the Q-table based on the current r_t with the corresponding r_1(=-1, ACK not received), r_2, r_3, and r_4 values.

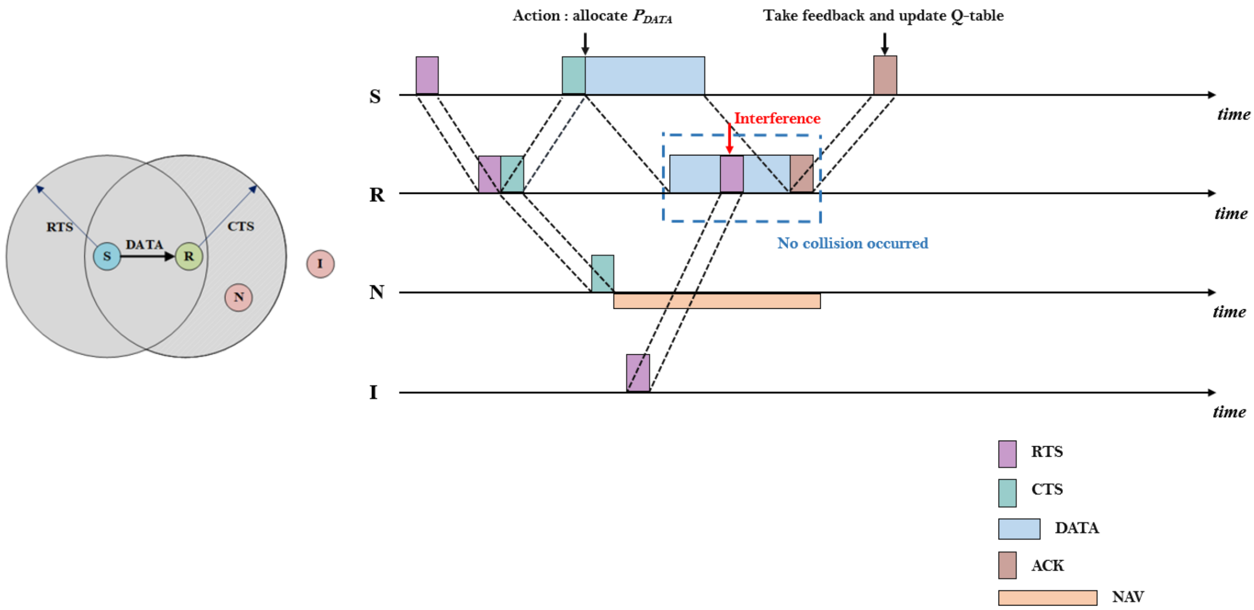

FIGURE 3. Operation of the proposed scheme.

V. INTERFERENCE RANGE ESTIMATION

In this section, the effect of the reduced power level on *IR* is discussed. In general, *IR* is larger than *TR* [8], [13]. Moreover, *IR* increases with a decrease in the P_{DATA} [9]. Specifically, LIRC is the result of the expansion of *IR* owing to the usage of the controlled transmission power such that sensors that are out of *IR* are originally covered by the expanded interference range afterward. From [10], *IR* is estimated as follows:

$$IR_{P_{DATA}} = \frac{W(\frac{P_{\text{MAX}}.A(D_{\text{SR}},f).SINR_{\text{TH}}.\ln a_f}{P_{\text{DATA}}})}{\ln a_f}, \tag{16}$$

where $W(\cdot)$ denotes the Lambert W function [26], and a_f is the frequency (measured in kilohertz) dependent absorption coefficient (measured in decibels per kilometer), which can be expressed empirically using Thorp's underwater channel model [27].

VI. ENERGY CONSUMPTION ANALYSIS

According to [28], the energy consumption to transmit a packet can be expressed as

$$E_{\text{TX}} = P_{\text{TX}}.T_{\text{TX-PACKET}}, \tag{17}$$

where $T_{\text{TX-PACKET}}$ is the time duration to transmit a packet.

In addition, the energy consumption to receive a packet can be defined as

$$E_{\text{RX}} = P_{\text{RX}}.T_{\text{RX-PACKET}}, \tag{18}$$

where $T_{\text{RX-PACKET}}$ is the time duration to receive a packet.

In the proposed protocol, the transmit energy consumption is considered according to (17) and can be expressed as

$$E_{\text{TX}} = [P_{\text{MAX}}.(T_{\text{TX-RTS}} + T_{\text{TX-CTS}})] + [P_{\text{DATA}}.(T_{\text{TX-DATA}} + T_{\text{TX-ACK}})], \tag{19}$$

where, $T_{\text{TX-RTS}}$, $T_{\text{TX-CTS}}$, $T_{\text{TX-DATA}}$, $T_{\text{TX-ACK}}$ is the time duration to transmit RTS, CTS, DATA, and ACK packets, respectively.

The received energy consumption can be described according to (18) as follows:

$$E_{\text{RX}} = P_{\text{RX}}.(T_{\text{RX-RTS}} + T_{\text{RX-CTS}} + T_{\text{RX-DATA}} + T_{\text{RX-ACK}}), \tag{20}$$

where, $T_{\text{RX-RTS}}$, $T_{\text{RX-CTS}}$, $T_{\text{RX-DATA}}$, $T_{\text{RX-ACK}}$ is the time duration to receive RTS, CTS, DATA, and ACK packets, respectively.

Therefore, the total energy consumption of the network can be written as

$$E_{\text{TOTAL}} = E_{\text{TX}} + E_{\text{RX}}. \tag{21}$$

VII. PERFORMANCE EVALUATION

A. SIMULATION MODEL

We evaluate the performance of the proposed protocol by using computer simulations in MATLAB. The simulation parameters are listed in Table 1. To reflect the underwater channel characteristics, Thorp's empirical underwater acoustic channel model [29] is applied. Additionally, the following interference scenarios are assumed. First, it is assumed that when an interference occurs between a control packet and a DATA packet, the DATA packet cannot be decoded if the condition given in (7) is not satisfied even for partial overlapping [30]. Second, it is assumed that the interference signal and the receiving signal are uncorrelated and that the interference signals can be considered as additive terms [30]. To determine the values of the data rate and TX/RX powers, we use the specifications of the commercially available underwater Teledyne Benthos ATM-903 modem.

Each sensor is randomly deployed in one of the three clusters and forms a star topology network centered around the sink, as shown in Figure 2. The three clusters are deployed to consider inter-cluster interference scenarios. Moreover, a specific simulation scenario, such as varying average traffic load with a fixed number of sensors in a cluster, is considered. For example, in cluster-1, 12 sensors generate traffic, which follows a Poisson distribution, at a rate of 0.005 to 0.015 packets/s. Each simulation is repeated 3000 times to improve the reliability of the results.

TABLE 1. System parameters and values.

Parameters	Values
Number of clusters	3
Cluster radius	6 km
Propagation speed	1500 m/s
Transmission range of CH	6 km
Transmission range of sensors	6 km
Data packet size	6000 bits
RTS/CTS packet size	200 bits
Bit rate	2400 bps
$SINR_{th}$	20 dB
$\psi_{N_{MAX}}$	4
P_{MAX}	20 W
α	0.1
ϵ	0.1
Simulation time	21,600 s

B. PERFORMANCE MATRICES

Five performance metrics, namely, – network throughput, network energy consumption, network energy efficiency, collisions per packet, and packet delivery ratio (PDR), are considered for performance evaluation.

1) NETWORK THROUGHPUT, ρ

Let N_R denote the total number of DATA packets that are successfully received throughout T for which the network was active. Then, the network throughput can be express as

$$\rho = \frac{N_{\mathrm{R}}.L_{\mathrm{D}}}{T}[\text{bits/sec}] \tag{22}$$

where L_{D} denotes the DATA packet size.

2) NETWORK ENERGY CONSUMPTION, σ

The network energy consumption is defined as the total amount of energy consumed by the entire network throughout T. In the proposed scheme, it is calculated according to (21).

3) NETWORK ENERGY EFFICIENCY, θ

By considering σ, network energy efficiency can be expressed as

$$\theta = \frac{N_{\mathrm{R}} \cdot L_{\mathrm{D}}}{\sigma}[\text{bits/j}] \tag{23}$$

4) COLLISIONS PER PACKET, τ

The collisions per packet is defined as the ratio of the number of collided DATA packets to the number of transmitted DATA packets. It is expressed as

$$\tau = \frac{Coll_{\mathrm{TOTAL}}}{N_{\mathrm{T}}} \tag{24}$$

where $Coll_{\mathrm{TOTAL}}$ and N_{T} represent the total number of collided DATA packets and total number of transmitted DATA packets throughout T, respectively.

5) PACKET DELIVERY RATIO (PDR), μ

The PDR can be express as

$$\mu = (\frac{N_{\mathrm{R}}}{N_{\mathrm{T}}}) \times 100 \tag{25}$$

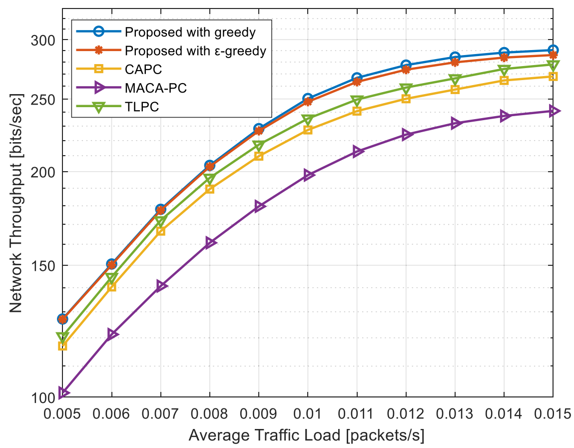

FIGURE 4. Network throughput vs average traffic load.

C. SIMULATION RESULTS

The proposed protocol is compared with the following benchmarks: CAPC [9], TLPC [10], and MACA-PC [13]. In addition, it is also evaluated by implementing greedy algorithm [31].

Figure 4 shows the network throughput for the various average traffic loads. An upward performance is observed for all the five protocols. The performance of the proposed protocol with the greedy algorithm is superior to that of the proposed protocol with the ϵ-greedy algorithm, TLPC, CAPC, and MACA-PC, with a throughput improvement of approximately 2%, 4.5%, 8%, and 20%, respectively, when average traffic load is 0.015. Moreover compared to the proposed protocol with the ϵ-greedy algorithm, the proposed protocol with the greedy algorithm shows a slightly better performance with an average traffic load of 0.010. This is because the proposed protocol with greedy algorithm selects the transmission power based on the maximum Q value until the end of the simulation period. Conversely, the proposed protocol with the ϵ-greedy algorithm explores although it finds a transmission power corresponding to the maximum Q value. Therefore, a low power level is occasionally selected for DATA packet transmission, which cannot satisfy the $SINR_{TH}$ requirements. Hence, a small drop in the throughput can be observed. Likewise, both the proposed protocol with the greedy algorithm and that with the ϵ-greedy algorithm adopt the DATA transmission power according to the surrounding interference environment. Thus, packet losses are minimized. Conversely, TLPC, MACA-PC, and CAPC cannot allocate sufficient power to a DATA transmission that is robust to multiple interferences, resulting in higher packet losses. When packet losses occur, the latency for successful packet delivery increases, thus decreasing the network throughput.

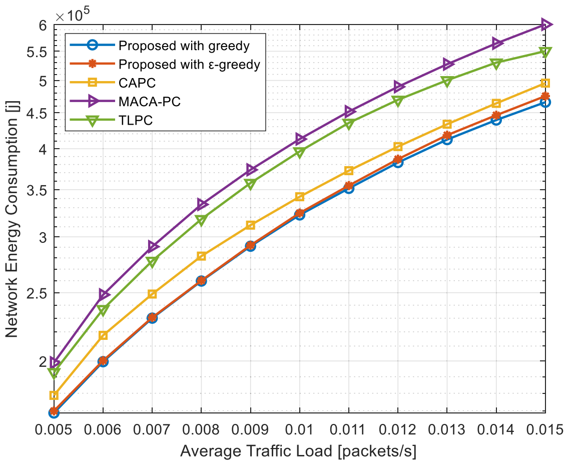

FIGURE 5. Network energy consumption vs average traffic load.

Figure 5 shows the network energy consumption for the various average traffic loads. The proposed protocol with the greedy algorithm has lower energy consumption than TLPC, CAPC, and MACA-PC by, 15%, 6%, and 22%, respectively, for an average traffic load of 0.015. However, the proposed protocol with greedy algorithm has a slightly lower energy consumption of 2% than that with the ϵ-greedy algorithm. This is because the former does not explore; therefore, the DATA packet transmission power is selected based on the maximum Q value. Thus, the number of retransmissions is minimized. In the case of the proposed with ϵ-greedy algorithm, sometimes insufficient power levels are chosen because of the exploration. This increases the number of retransmissions; consequently, the energy consumption increases. MACA-PC has the highest energy consumption among all the protocols. This is because, in this protocol, before transmitting DATA packets with optimal transmission powers, a short tone packet with a maximum power level is always transmitted. Thus, the energy consumption in the network increases. In addition, MACA-PC does not consider multiple interferences when determining the optimal transmission power. Thus, when multiple interferences occur, the *SINR* cannot satisfy the threshold requirements. Moreover, the energy consumption increases owing to an increase in the number of retransmissions. The reason why TLPC and CAPC have high energy consumption is similar to the explanation for MACA-PC. However, in TLPC and CAPC, no tone packet is sent before DATA packet transmissions.

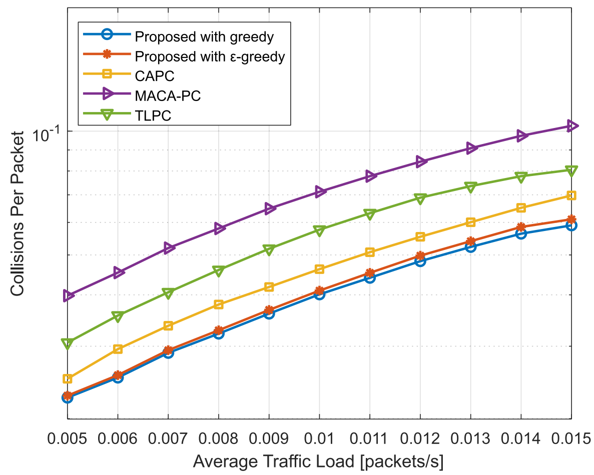

FIGURE 6. Collisions per packet vs average traffic load.

Figure 6 shows the average number of collisions per packet for the various average traffic loads. The value increases with an increase in the average traffic load. Notably, all the protocols exhibit similar characteristics. This is because, with an increasing average traffic load, CHs are subjected to numerous collisions owing to an increase in the number of multiple interferences. These interferences are introduced because of an increase in the number of DATA transmissions in a network with power control. The proposed protocol with greedy algorithm has a 27%, 15%, and 37% lower average number of collisions per packet compared to TLPC, CAPC, and MACA-PC, respectively, when average traffic load is 0.015. This is because both the proposed protocols with greedy and ϵ-greedy algorithms have effectively learned the optimal transmission power that can overcome multiple interferences, thus lowering the average number of collisions per packet. Conversely, the results of TLPC, CAPC, and MACA-PC reveal that assigning insufficient DATA transmission power can increase packet losses and subsequently the number of average collisions per packet. Moreover, the highest average number of collisions per packet is observed in MACA-PC. This is because the interference range in MACA-PC is set to twice the transmission range according to [32]. Notably, if the interference range doubles the transmission range size, the collision probability increases because more sensors are covered by the interference range.

Figure 7 shows the PDR, which decreases as the average traffic load increases. It is also observed that when the weighting parameter $\beta_1 > \beta_2$, the PDR is higher than that when $\beta_1 < \beta_2$. This is because, $\beta_1 > \beta_2$ implies that a higher priority is given to the collision avoidance reward factor (that is, r_1) rather than the energy efficiency reward factor (that is, r_2). Therefore, the network throughput increases; however, this leads to a decrease in energy efficiency.

Figure 8 shows the network energy efficiency obtained by varying the average traffic load. The efficiency decreases with an increase in the average traffic load, as shown in Figure 7.

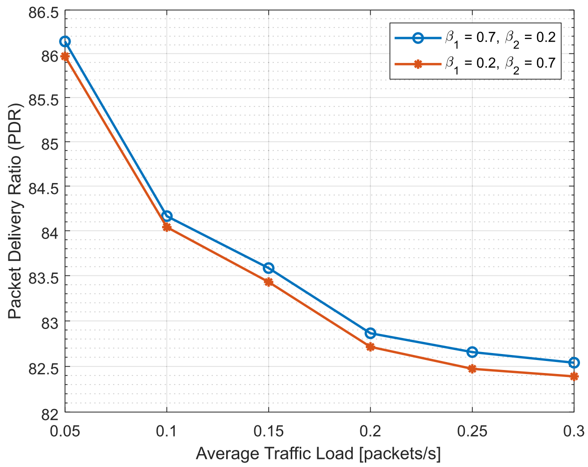

FIGURE 7. Packet delivery ratio vs average traffic load.

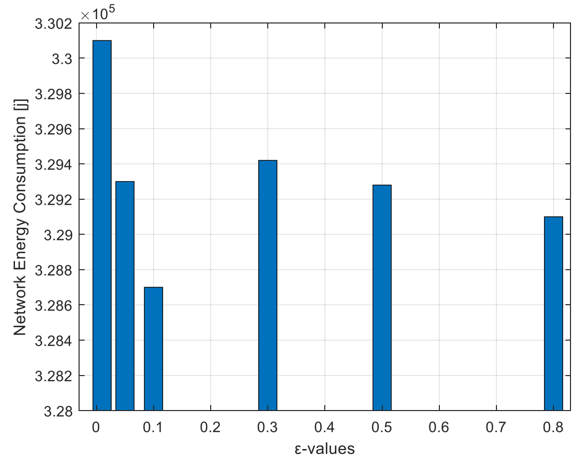

FIGURE 9. Network energy consumption vs ϵ-values.

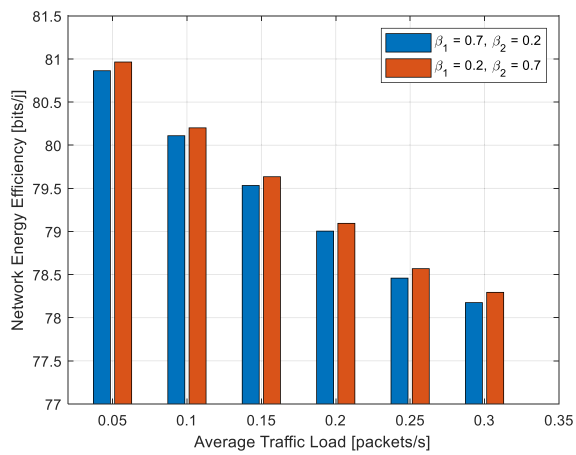

FIGURE 8. Network energy efficiency vs average traffic load.

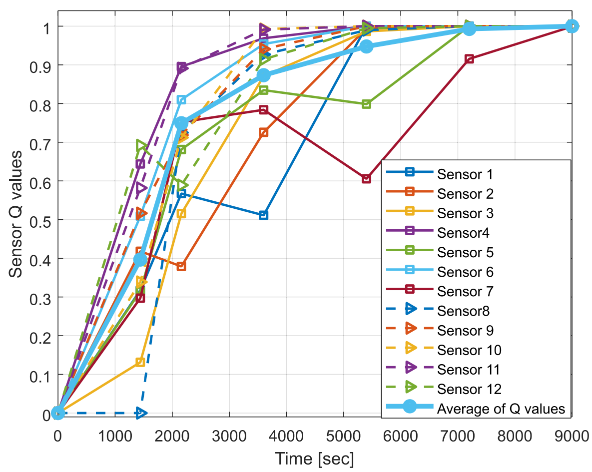

FIGURE 10. Convergence of Q values vs simulation time.

When $\beta_1 < \beta_2$, more weight is given to the energy-efficient reward factor than when $\beta_1 > \beta_2$, resulting in higher energy efficiency.

Figure 9 shows the network energy consumption for varying ϵ-values. Notably, the network energy consumption is the lowest when the ϵ value is set to 0.1. This is because, for an extremely low ϵ-value, there is insufficient exploration, thereby making the convergence extremely slow. Conversely, for a high ϵ-value, an unnecessarily large amount of exploration is introduced, which increases the energy consumption. The second lowest network energy consumption occurs when $\epsilon = 0.8$. On average, the discrepancy between $\epsilon = 0.1$ and $\epsilon = 0.8$ is 400 W.

Figure 10 shows the convergence of the Q values of the sensor with increasing simulation time. Among the three clusters, only one (consisting of 12 sensors) is considered for this particular simulation. It can be observed that the sensor behavior is quite similar in terms of convergence. Most of the sensors converge at 5400 s, whereas very few of them converge within the range from 7000 to 9000 s. On average, approximately 3500 s are required to achieve 90% convergence. Moreover, on average, almost full convergence is achieved in 7000 s, which is one-third of the total simulation time. The result presented in Figure 10 demonstrates the adaptability and robustness of the proposed protocol.

VIII. CONCLUSION

To alleviate the LIRC problem, which is intrinsic to MACA-based underwater MAC protocols, we demonstrated that the determination of the optimal transmit power using Q-learning improves the network's performance. In particular, we used a collective reward system to update the Q values. The proposed protocol is tested under a practical multiple-interference scenario. Our findings demonstrated that the proposed protocol is robust to collisions, as it can reduce collisions by 35%, and energy-efficient, as it consumed 20% lesser energy, which is a valuable resource for sensors. Moreover, the proposed protocol can determine the optimal transmission power of the sensors in the multi-cluster network that is adapted to the strength of the surrounding interferences. In addition, under different traffic load conditions, the proposed protocol consistently outperformed existing benchmark schemes.

REFERENCES

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: Research challenges," *Ad Hoc Netw.*, vol. 3, no. 3, pp. 257–279, Mar. 2005.

[2] J. Heidemann, W. Ye, J. Wills, A. Syed, and Y. Li, "Research challenges and applications for underwater sensor networking," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, vol. 1, Apr. 2006, pp. 228–235.

[3] J. Partan, J. Kurose, and B. N. Levine, "A survey of practical issues in underwater networks," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 11, no. 4, pp. 23–33, 2007, doi: 10.1145/1347364.1347372.

[4] S. Jiang, "State-of-the-art medium access control (MAC) protocols for underwater acoustic networks: A survey based on a MAC reference model," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 96–131, 1st Quart., 2018.

[5] K. Chen, M. Ma, E. Cheng, F. Yuan, and W. Su, "A survey on MAC protocols for underwater wireless sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1433–1447, 3rd Quart., 2014.

[6] F. Ahmed and H.-S. Cho, "A time-slotted data gathering medium access control protocol using Q-learning for underwater acoustic sensor networks," *IEEE Access*, vol. 9, pp. 48742–48752, 2021.

[7] P. Casari and M. Zorzi, "Protocol design issues in underwater acoustic networks," *Comput. Commun.*, vol. 34, no. 17, pp. 2013–2025, Nov. 2011, doi: 10.1016/j.comcom.2011.06.008.

[8] J.-Z. Zhang, H. Yang, X. Zhou, and X.-J. Zhou, "A power control MAC protocol for ad hoc networks," in *Proc. 4th Int. Mobile Multimedia Commun. Conf.*, New York, NY, USA, 2008, pp. 36–47.

[9] K.-P. Shih and Y.-D. Chen, "CAPC: A collision avoidance power control MAC protocol for wireless ad hoc networks," *IEEE Commun. Lett.*, vol. 9, no. 9, pp. 859–861, Sep. 2005.

[10] Y.-D. Chen, C.-Y. Lien, Y.-S. Fang, and K.-P. Shih, "TLPC: A two-level power control MAC protocol for collision avoidance in underwater acoustic networks," in *Proc. MTS/IEEE OCEANS-Bergen*, Jun. 2013, pp. 1–6.

[11] C. J. Watkins and P. Dayan, "Technical note: Q-learning," *Mach. Learn.*, vol. 8, no. 3–4, pp. 279–292, May 1992.

[12] J. Cho, F. Ahmed, E. Shitiri, and H.-S. Cho, "Power control for MACA-based underwater MAC protocol: A Q-learning approach," in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Aug. 2021, pp. 1–4.

[13] L. Qian, S. Zhang, M. Liu, and Q. Zhang, "A MACA-based power control MAC protocol for underwater wireless sensor networks," in *Proc. IEEE/OES China Ocean Acoust. (COA)*, Jan. 2016, pp. 1–8.

[14] Z. J. Haas and J. Deng, "Dual busy tone multiple access (DBTMA)–A multiple access control scheme for ad hoc networks," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 975–985, Jun. 2002.

[15] H. Yan, J. Li, G. Sun, and H.-H. Chen, "An optimistic power control MAC protocol for mobile ad hoc networks," in *Proc. IEEE Int. Conf. Commun.*, vol. 8, Jun. 2006, pp. 3615–3620.

[16] M. Zawodniok and S. Jagannathan, "A distributed power control MAC protocol for wireless ad hoc networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, vol. 3, Mar. 2004, pp. 1915–1920.

[17] K.-P. Shih, Y.-D. Chen, and C.-C. Chang, "Adaptive range-based power control for collision avoidance in wireless ad hoc networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2007, pp. 3672–3677.

[18] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," in *Proc. 1st ACM Int. workshop Underwater Netw. (WUWNet)*, New York, NY, USA, 2006, pp. 41–47.

[19] D. E. Lucani, M. Stojanovic, and M. Medard, "On the relationship between transmission power and capacity of an underwater acoustic communication channel," in *Proc. OCEANS-MTS/IEEE Kobe Techno-Ocean*, Apr. 2008, pp. 1–6.

[20] S. H. Park, P. D. Mitchell, and D. Grace, "Reinforcement learning based MAC protocol (UW-ALOHA-Q) for underwater acoustic sensor networks," *IEEE Access*, vol. 7, pp. 165531–165542, 2019.

[21] Y. Chu, P. D. Mitchell, and D. Grace, "ALOHA and Q-learning based medium access control for wireless sensor networks," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2012, pp. 511–515.

[22] S. H. Park, P. D. Mitchell, and D. Grace, "Reinforcement learning based MAC protocol (UW-ALOHA-QM) for mobile underwater acoustic sensor networks," *IEEE Access*, vol. 9, pp. 5906–5919, 2021.

[23] C. Savaglio, P. Pace, G. Aloi, A. Liotta, and G. Fortino, "Lightweight reinforcement learning for energy efficient communications in wireless sensor networks," *IEEE Access*, vol. 7, pp. 29355–29364, 2019.

[24] S. Barrachina-Muñoz, A. Chiumento, and B. Bellalta, "Stateless reinforcement learning for multi-agent systems: The case of spectrum allocation in dynamic channel bonding WLANs," 2021, *arXiv:2106.05553*.

[25] S. Kosunalp, Y. Chu, P. D. Mitchell, D. Grace, and T. Clarke, "Use of Q-learning approaches for practical medium access control in wireless sensor networks," *Eng. Appl. Artif. Intell.*, vol. 55, pp. 146–154, Oct. 2016.

[26] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert*W* function," *Adv. Comput. Math.*, vol. 5, pp. 329–359, Dec. 1996.

[27] L. M. Brekhovskikh, L. M. Brekhovskikh, and Y. Lysanov, *Fundamentals of Ocean Acoustics*, 3rd ed. New York, NY, USA: Springer, Mar. 2003.

[28] G. Xing, Y. Chen, L. He, W. Su, R. Hou, W. Li, C. Zhang, and X. Chen, "Energy consumption in relay underwater acoustic sensor networks for NDN," *IEEE Access*, vol. 7, pp. 42694–42702, 2019.

[29] X. Lurton, *An Introduction to Underwater Acoustics: Principles and Applications* (Geophysical Sciences Series). New York, NY, USA: Springer, 2002.

[30] H. A. Leinhos, "Capacity calculations for rapidly fading communications channels," *IEEE J. Ocean. Eng.*, vol. 21, no. 2, pp. 137–142, Apr. 1996.

[31] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[32] A. Kamerman and L. Monteban, "WaveLAN-II: A high-performance wireless LAN for the unlicensed band," *Bell Labs Tech. J.*, vol. 2, no. 3, pp. 118–133, May 1997.