Journal of Smart Internet of Things (JSIoT)
VOL. 1 (2023), No.01 | 05-13 | 2023
DOI: 10.2478/jsiot-2023-0001

Article

A Deep Reinforcement-Learning-Based Relay Selection for Underwater Sensors Network

Muhmmad Waleed Aftab[1], Sajjad Hussain[1], Aftab Husain[1], Umar Ali Khan[1], and Hamza Kundi*[1]

[1] Electrical Engineering Department, Gomal University, D.I.Khan, Pakistan

*Corresponding author: Hamza Kundi (email: artificiali19@gmail.com)

Received xx March 2023; Accepted x July 2023; Published xx July 2023

Abstract: Due to their limited frequency range and fast fading channels, underwater sensor networks (USNs) are vulnerable to collisions of packets. In this paper, we propose a deep reinforcement learning-based relay selection scheme with shortest latency (DRL-SL) for USNs that enables to choose the relay based on the state that comprised of the bit error rate (BER) of the previous transmission, and the jamming power measured by the relay node. The DRL-SL-based relay selection scheme completed in two phases. In the first phase, a deep neural network based learning is performed and second phase is the real-time interaction with the underwater sensor network. Numerical results give the bound on how efficiently the system performs in terms of bit error rate, energy use, and node utility. According to the numerical results, the proposed DRL-SL based relay selection scheme can enhance relay performance in comparison to the benchmark underwater relay techniques.

Keywords: Collision avoidance, energy efficiency, deep learning, reinforcement learning, underwater wireless sensor network.

Copyright © 2023 Journal of Smart Internet of Things (JSIoT) published by Future Science for Digital Publishing and Sciendo . This is an open access article license CC BY (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

The Underwater sensor networks (USNs) aims to perform a variety of tasks, including ocean sampling, environmental control, deep-sea investigation, pollution monitoring, and aided navigation, has recently become quite popular [1], [2]. In USNs, each sensor node (SN) is randomly placed throughout a region to gather data and transmit it to the final node (FN). After that, FN transfers the data it has gathered to the floating node, which then sends it through cable or terrestrial connectivity to the monitoring centers. Researchers are making great endeavor to reliably transfer the data gathered nodes to the respective FN as the further transmission is wired and reliable. High bit error rates (BER) are experienced when a signal was transmitted via an underwater transmission medium (UTM) due to a long propagation delay. However, the USNs' use of sound waves as a transmission medium is what causes the significant propagation delay [3], [4]. Harsh UTM effects,

including a multipath fading, high BER, significant attenuation, and lengthy propagation delays, are experienced by the acoustic signal. An acoustic signal's transportation loss is caused by the signal's frequencies and the distance among nodes [5]–[8]. Since cooperative communication is the most trustworthy method for transmitting gathered information in an energy-limited sensors network, and in the recent times it grabbed attention of the researchers.

Researchers need to focus on choosing a routing system that is extremely reliable and consumes less energy. The transmission capacity between USNs is noticeably lower than that of ground sensors nodes (GSNs) because there is less bandwidth available for UTM. USNs therefore use more energy than GSNs when transferring packets of a same length [9]. A strong, reliable, and energy-efficient routing protocol is needed to deal with the dynamically varying UTM circumstances and boost the efficiency of energy and data transmission reliability in USNs.

Machine learning has recently become a potential method for proactively enhancing system reliability in a complex and demanding environment. To adjust the reinforcement learning (RL) driven weight of nodes and select the best participating nodes (PN) from a list of prospective PNs, the FN rely on adjacent nodes (ANs). In RL-based technique, the model picks up information and behavior from the surrounding area to improve the rewards function. The action gives the reward that shows a good or bad performance of RL. The agent receives limited information from the external environment (UTM). Therefore, RL needs on exposure to obtain data for feedback during the action evaluation phase. The procedure improves the ability to adapt in the UCC's dynamic environment [10].

In addition, deep neural network (DNN) is combined with RL (DRL), to optimize the RL system with a convergence to give an optimal solution. As shown in Figure. 1, the agents in DRL first interact with the surroundings to modify the reward. The agent's reward may increase as a result of this contact, or it may decrease. Surroundings and agents frequently interact during the entire process to generate a vast amount of data. The agent eventually learns the best course of action to take in order to complete the task after many iterations. Simultaneously, the data is inputted to DNN to obtain the convergence point while considering collision, energy and latency.

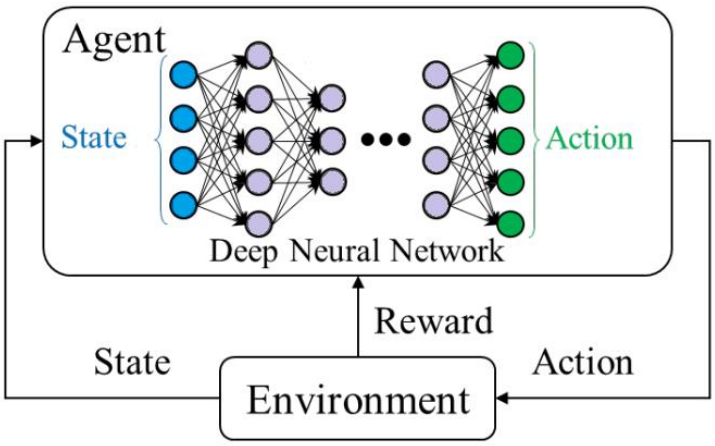


Figure 1: A DRL Approach.

1.1 Related Works

Different routing algorithms for USNs had been investigated in previous research. USNs do not have access to the Global Positioning System to gather real-time position data, in contrast to GSNs. It is difficult to adapt routing methods for USNs because UTM caused an increased in energy

consumption and have an additional latency. Similar to this, the centralized routing techniques suggested in [11] fall short of meeting the requirements of USNs. The origin based routing algorithm was recently proposed by authors in [12]–[14]. The pipeline idea is put into practice. Each SN in the pipeline has a corresponding FN and information about its position. An implicit pipe is used to convey the data from the source node (SRN), which is now storing the data packet, to the FN. In an environment with saturated nodes, controls system flooding to minimize energy loss brought on by a high density of data relay nodes. In contrast, the authors of [12] and [13] used the concept of an adaptive pipeline to advance data packets, ensuring the reliability of the data packets' conveyance in the limited system.

The anycast routing approach that relies on hydraulic pressure was introduced by the authors in [15]. Based on various pressure levels, it shows the prioritization for sending network packets in each group. The authors in [16] presented the collaborative anycast routing mechanism relying on hydraulic pressure as a way to enhance the protocol introduced in [15]. By allowing SN cooperation, it enhances the forwarding mechanism and thus boosts data rate. Despite this, that strategy might limit the variety of options. In [17], the authors suggested a cooperative, secure, and energy-efficient routing strategy for UWSNs. The energy efficient SNs and the confidentiality of data packets were two noteworthy aspects in this case. The impact of control and data packet collision was avoided by the authors. The authors of [18] developed a method that could be easily adapted to different underwater conditions in order to achieve improved effectiveness with the least amount of computational complexity. The authors employed opportunistic routing (OR), which resulted in a data packet collision at FN. When IN wanted to convey data packets, [19], and [20] devised a collaborative routing scheme. Separate PNs were chosen using the signal-to-noise ratio (SNR) and distances from the FN.

Multi-agent RL was recently proposed by authors in [21] for subsequent decision. The authors of [22] suggested an RL-based low energy, adaptable, and life span based routing system. The Q-value is created by the authors, who also optimize the network lifetime and energy usage. Considering this, the computer simulation chosen was insufficient for a UTM that underwent constant change. A multistage routing protocol based on Q-learning was suggested by the authors in [23]. All of the SNs delivered the collected packets to the cluster's leader, who then used a cooperative transmission to send the network packets to the DN. Although the performance is improved, the system has more redundant data packets, which results in collisions and energy losses.

1.2 Contributions and Organization

The severe underwater environment in the earlier research left open issues with data packet collision, excessive energy consumption, and delay. We employed the DRL technique coupled with the shortest latency (DRL-SL) for the optimal PN selection in order to increase system reliability, efficiency in real-time, and data delivery efficiency. First, in this study, we compile a list of possible PNs by taking propagation delay, the quantity of ANs, and remaining energy into account. Then, by routinely changing the reward function of SNs, we select the best PN. The main contributions of this paper is summarized as follows:

- A list of potential PNs is compiled in DRL-SL utilizing the shortest latency. A DRL-SL approach is used to choose the best PN, with each PN in the potential list of candidate continuously calculating the reward function.
- The collected data of reward, states, and actions are fed to DNN to converge to an optimal action value. The optimal value of each PN in the potential list of candidate is sent to SRN.
- The best PN with the highest optimal point is chosen by SRN. The data packet is forwarded to the chosen PN by SRN, who serves as a potential helper.

The remainder of the study is structured as follows: in Section II, we provide a summary of the system model. We define the problem statement in Section III. In Section IV, we outline the proposed DRL-SL-based relay node selection method. In addition, Section V uses simulations to assess how well the proposed work performs. Finally, Section VI conclude this article.

2. Application of Artificial Intelligence in SDN & NFV Based Networks

The SRN, FN, and number of SNs make up the USNs. The total ANs for the kth SN is presented as Mk. The set W, where { k = 1,...,M^k ∈W }, donates the list of ANs. During the system's initialization, each SN obtains the propagation latency from the kth SN to the FN, which is denoted as Lk. Additionally, LMF is donated as the propagation delay from SRN to FN. The average SNR for the kth SN, which is now storing a data packet, is provided as

$$\gamma_k = \frac{A_e / \alpha}{n} \tag{1}$$

Where, n shows the noise spectral density; we consider it to be a Gaussian noise, α is the pathloss component and A_e represents the average energy required to send one bit of data. Now, the bit error rate of kth SN is presented as

$$\rho_k = \frac{1}{2}(1 - \sqrt{\frac{\gamma_k}{1+\gamma_k}}) \tag{2}$$

Additionally, using QPSK modulation [25], we calculate the kth SN's successful transmission probability for n data packets of p bits each, and this is denoted as

$$\mathrm{Pr}_k^n = \sum_{n=1}^{N}([1-\rho_k^n]^p) \tag{3}$$

Where, N represents the total packets. The kth SN's average throughput is described by

$$\Gamma_k^n = \mathrm{Pr}_k^n \times R_t \tag{4}$$

Where, R_t is expressed as the transmission rate.

The amount of energy required by k to accept n number of data packets is expressed as

$$\zeta_k^r = n \times \varepsilon_k^r \tag{5}$$

Where, ε_k^r is a device-specific reception coefficient of k. The total amount of energy used while sending packets n at distance d is shown as

$$\zeta_k = \zeta_k^r + \zeta_k^t(d,n) \tag{6}$$

Where, the energy $\zeta_k^t(d,n)$ used by the kth SN to transmit data at a distance d. Its overall energy usage ζ_k must be less than the remaining energy ζ_k^{RE}. The main aim to minimize total energy consumption ζ_k and latency.

3. DRL-SL Algorithm

We present a DRL-base relay selection with shortest latency (DRL-SL) to enhance the learning process in the selection of relay policy with the optimization objective function. This method uses DNN to compress the USN state space and calculate Q-values for relay nodes with shortest latency to support deep learning. As illustrated in Algorithm 1, the DNN parameters are initialized by

utilizing the prior relay experiences in comparable USNs. The relay acquires the BER of the prior sensing data by receiving the performance feedback from the SRN. A markov decision process is given by the tuple (S, A, P, R), Where P is the likelihood that state S will transition to state S' after an action is conducted, and R is the function that decides how much is rewarded immediately following an action. Where A, and S are a bounded collection of actions and states, respectively.

State: The system states of USNs at the start of every time slot, are propagation delay, and residual energy.

Action: The system actions are selecting the PN as a relay to transfer the data packets and computing successful transmission probability.

Reward: Decreasing the overall energy consumption, BER, latency and maximizing the probability of success transmission. The system computes the reward resulting action performed and then updates the state space.

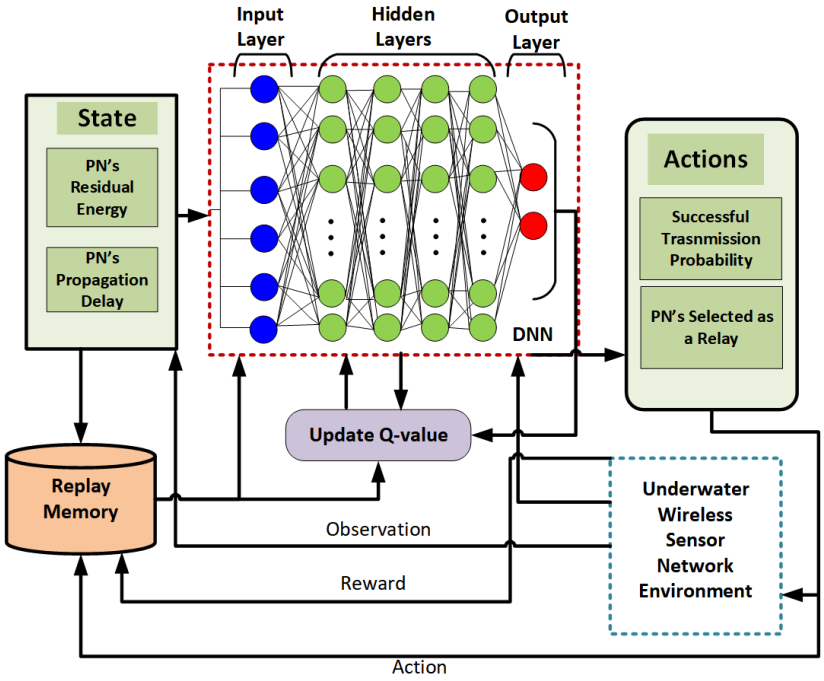


Figure 2: A proposed DRL-SL Approach.

Online reinforcement learning and offline DNN training are the two phases of DRL. The DNN is learned offline to estimate the correlation between the state-action pair and the Q-function, which reflects the expected reward with the penalty of choosing to remain in-state S and performing action A. The Q-function is defined as

$$Q(S,A) = \mathrm{E}[R + \upsilon \max Q(S',A') \mid S,A] \quad (7)$$

Where, υ is defined as the discount factor. For the offline DNN learning phase must have sufficient experience samples at each iteration. The state, action, instantaneous reward, and change in states experiences are updated into the DNN by selecting them at random from the replay memory. This offline method can smooth learning, avoid oscillations, and divergent parameter estimates by using experience storage. The online probability control based on the offline DNN is then improved using deep Q-learning. In each decision iteration, the DRL agent retrieves the predicted Q from the offline DNN using the information about the current state and each possible

action. The execution action is then selected using the greedy policy. Choosing the course of action with the maximum Q value is our policy of greed. The DRL process then chooses the set R associated with the highest output value of the neural network, that is,

$$R_f = \arg\max_R Q(S_t \mid R_t; \theta) \tag{8}$$

A discounted average reward, often known as $Q(S_t \mid R_t; \theta)$, is an approximate of $Q(S, A)$ at time slot t, and θ denotes the weight of DNN. The proposed DRL-SL approach is presented in Figure. 2.

Algorithm 1: Pseudocode of the proposed DRL-SL

1: ***Tasks Performed At PN***
2: Get the propagation delay and Residual Energy of each SN
3: ***For all*** *SN retrieve the latency*
4: ***IF*** *current PN latency < AN latency*
5: PN act as a potential node to participates in data transmission
6: Get the list of potential candidates for data transmission
7: Train the ***DNN*** with buffer memory
8: Observe ***Underwater Wireless Sensor Network Environment***
9: Define states ***S***
10: Perform Action ***A***
11: Get Reward ***R***
12: Update ***Q-value***
13: ***END***
14: ***END***
15: ***Tasks Performed at SRN***
16: Retrieve the best PN
17: Transmit Data packet to the PN
18: ***Tasks Performed at FN***
19: ***IF*** *Collision*
20: Ask for ***Retransmission***
21: ***END***

4. Simulation Results

We dispersed 100 static SNs at random throughout a 500m x 500m x 500m region. Each SN receives a distinct node SRN and a set amount of energy. Each node has a finite range of communication and is aware of the propagation delay to the FN. All SNs are considered to be static (attached to the seafloor) and to have equivalent functional and physical characteristics.

For the purpose of determining the impact of acoustic communication in real time, we adopt the empirical underwater channel standard [8]. SNs are unable to acquire precise time synchronization because to underwater channel features. We did not go into greater detail because this isn't a significant topic in this paper. As no error correcting mechanism is used, the data packet retrieved is incorrect even if there is a single bit of error. We assume that because the control packets are small and error-free, they are. By taking UTM into account, SNs' communication success or failure is determined.

Figure. 3(a) shows that the latency of the OR increases with retransmissions, whereas the latency of the DRL-SL is constant. The quantity of retransmissions significantly affects delay in

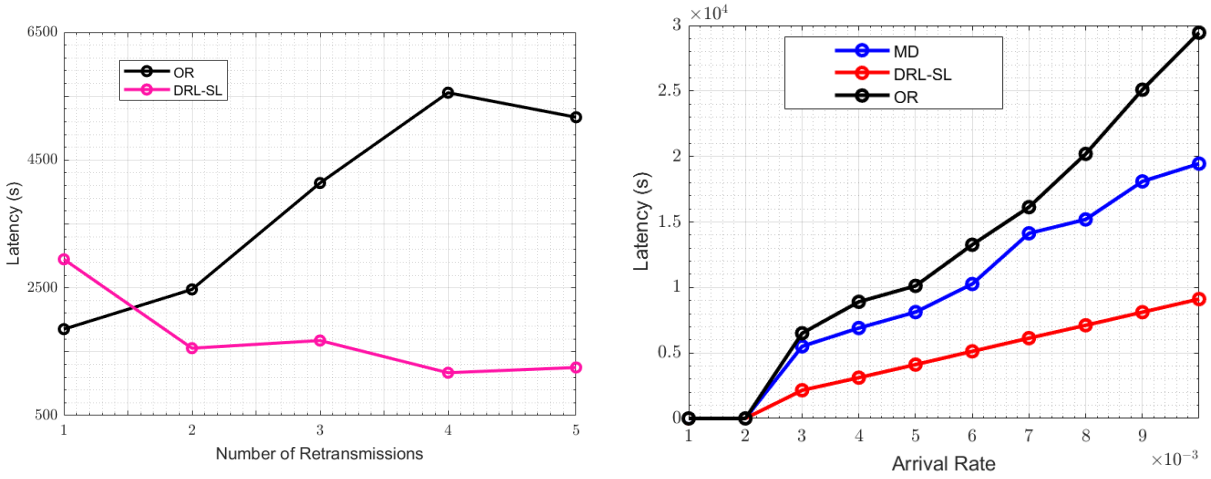
the underwater environment. Generally, it is assumed that the PN closest to the FN is preferable for the retransmission process because of the long propagation delay in the underwater channel. The PN that is farthest from the FN is the one closest to SRN. OR works as all the nodes opportunistically sends the stored data packets and causes numerous collision at the FN. Moreover, improper selection of PNs results in the transmission being a time and energy wastage. The PN with DRL-SL has an intelligence in selecting the PN with minimum latency with highest transmission probability. In this manner, the propagation delays from SRN to PN and PN to FN takes less time and PN will have a greater chance of receiving error-free data packets.

The proposed scheme's network performance is depicted in Figure. 3(b) along with comparisons to the OR and MD protocols. The DRL-SL has a lower latency than the OR and MD. While the OR and MD's maximum latency are 3x104 second and 1.92x104 seconds, respectively, the DRL-SL's network latency is around 0.8x104 seconds. As a result, DRL-SL has a 50 percent lower network latency than other protocols. The most reliable PN is chosen in the DRL-SL. Unexpectedly, OR's latency is lower than DRL-SL and MD. The reason is every opportunistic PN sends the controlling packet almost simultaneously, the drop is caused by collisions at SRN. Control packet collision rate is quite higher in OR and occurring more frequently as a result of different propagation delays.

Figure. 3(c) illustrates the energy efficiency of DRL-SL, OR and MD. he energy consumption measured when a data packet is created and sent at SNs. Energy consumption will increase if there exits more erroneous data packets. DRL-SL has a better energy efficiency that is around 42, while MD and OR energy efficiency is less then DRL-SL, which is 39 and 37, respectively. The one-time transmission at a specific time reduces lost data packets in DRL-SL. As seen in Figure 3(d), it also improves the ratio of packet delivery (PDR) accordingly. The maximum PDR for MD and OR is around 82% and 84%, respectively. Whereas, PDR is around 98% in DRL-SL. In every set of results, DRL-SL performs better than MD and OR.

5. Conclusion

USNs are susceptible to packet collisions because of their constrained frequency range and quickly fading communication channels. In this study, we present a DRL-SL for USNs that allows users to select a relay based on the BER of the previous transmission and the jamming power as detected by the relay node. The computational complexity of the deep-RL-based relay method is examined, and it provides a bound on how well the system works. The suggested DRL-SL based relay selection strategy can improve relay performance in comparison to the benchmark underwater relay techniques, according to the numerical results.



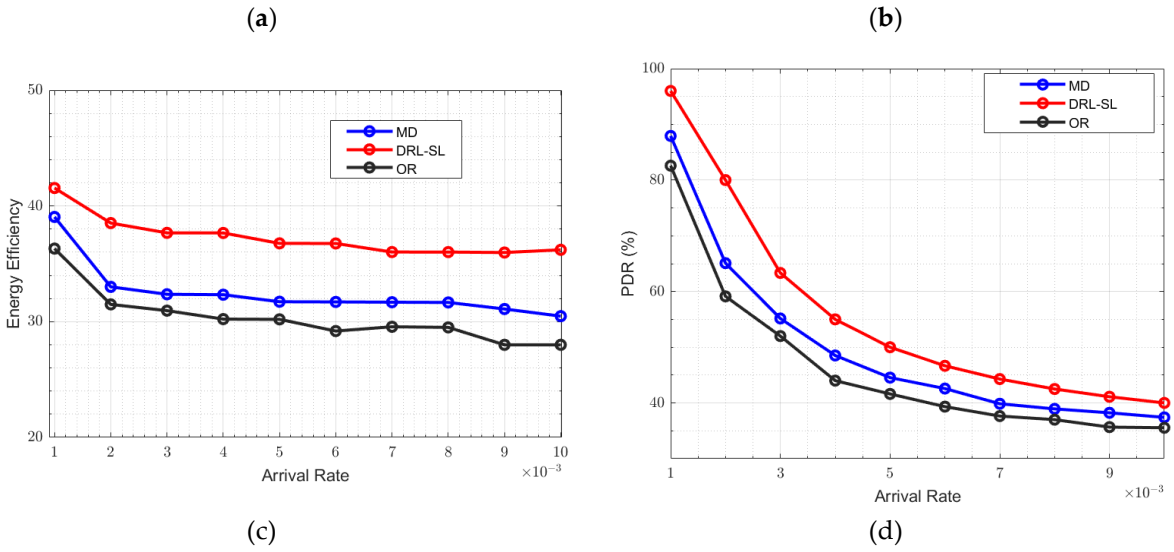


Figure 3: Performance matrix to compare proposed DRL-SL with MD and OR: (a) Latency Vs Number of Retransmissions; (b) Latency Vs Arrival Rate; (c) Energy Efficiency Vs Arrival Rate; (d) PDR Vs Arrival Rate.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. Latif, N. Javaid, I. Ullah, Z. Kaleem, Z. Abbas, and L. D. Nguyen, "DIEER: Delay-intolerant energy-efficient routing with sink mobility in underwater wireless sensor networks," Sens., vol. 20, no. 12, p. 3467, 2020.

[2] G. Khan, R. Dwivedi, and K. K. Gola, "Comparative analysis of routing algorithms for underwater sensor network," in Pervasive Computing A Networking Perspective and Future Directions. Springer, 2019, pp. 31–46.

[3] A. Abuarqoub, M. Hammoudeh, B. Adebisi, S. Jabbar, A. Bounceur, and H. Al-Bashar, "Dynamic clustering and management of mobile wireless sensor networks," Comput. Netw., vol. 117, pp. 62–75, 2017.

[4] Y. Su, R. Fan, and Z. Jin, "ORIT: A transport layer protocol design for underwater DTN sensor networks," IEEE Access, vol. 7, pp. 69 592–69 603, 2019.

[5] I. Ahmad and K. Chang, "Downlink power allocation strategy for next-generation underwater acoustic communications networks," Electr., vol. 8, no. 11, p. 1297, 2019.

[6] T. Melodia, H. Kulhandjian, L.-C. Kuo, E. Demirors, S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic, Advances in Underwater Acoustic Networking. Wiley Online Library, 2013.

[7] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," IEEE Trans. Signal Process., vol. 58, no. 3, pp. 1708–1721, 2009.

[8] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," IEEE Commun. Mag., vol. 47, no. 1, pp. 84–89, 2009.

[9] S. Jabbar, M. Ahmad, K. R. Malik, S. Khalid, J. Chaudhry, and O. Aldabbas, "Designing an energy-aware mechanism for lifetime improvement of wireless sensor networks: A comprehensive study," Mobile Netw. Appl., vol. 23, no. 3, pp. 432–445, 2018.

[10] S. H. PARK and D. GRACE, "Reinforcement learning based mac protocol (uw-aloha-qm) for mobile underwater acoustic sensor networks," IEEE Access, pp. 5906–5919, 2020.

[11] R. I. B. S. D. Muruganathan, D. C. F. Ma and A. O. Fapojuwo, "A centralized energy-efficient routing

protocol for wireless sensor networks," IEEE Commun. Mag., vol. 43, no. 3, pp. S8–13, 2005.
[12] B. S. F. Boavida and E. Monteiro, "Vbf: Vector-based forwarding protocol for underwater sensor networks," Networking Technologies, Services and Protocols, pp. 1216–1221, 2006.
[13] N. Y. H. Yu and J. Liu, "An adaptive routing protocol in underwater sparse acoustic sensor networks," Ad Hoc Netw, pp. 121–143, 2015.
[14] A. S. N. Nicolaou and D. Maggiorini, "Improving the robustness of location-based routing for underwater sensor networks," IEEE OCEANS, pp. 1–6, 2007.
[15] R. Coutinho and A. Boukerche, "Hydrocast: Pressure routing for underwater sensor networks," IEEE Trans. Veh. Technol., vol. 65, no. 1, pp. 333–347, 2016.
[16] N. J. et al, "Cooperative opportunistic pressure based routing for underwater wireless sensor networks," Sensors, vol. 17, no. 3, p. 629, 2017.
[17] K. Saeed, W. Khalil, S. Ahmed, I. Ahmad, and M. N. K. Khattak, "SEECR: secure energy efficient and cooperative routing protocol for underwater wireless sensor networks," IEEE Access, vol. 8, pp. 107 419–107 433, 2020.
[18] A. A. A. El-Banna, K. Wu, and B. M. ElHalawany, "Opportunistic cooperative transmission for underwater communication based on the water's key physical variables," IEEE Sensors J., vol. 20, no. 5, pp. 2792–2802, 2019.
[19] S. Khadim, F. Riaz, S. Jabbar, S. Khalid, and M. Aloqaily, "A non-cooperative rear-end collision avoidance scheme for non-connected and heterogeneous environment," Comput. Commun., vol. 150, pp. 828–840, 2020.
[20] H. Tran-Dang and D.-S. Kim, "Channel-aware cooperative routing in underwater acoustic sensor networks," J. commun. Netw-s, vol. 21, no. 1, pp. 33–44, 2019.
[21] X. L. et al, "Marl-based distributed cache placement for wireless networks," IEEE Access, vol. 7, pp. 62 606–62 615, 2019.
[22] T. Hu, "Qelar: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks," IEEE Trans. Mobile Computing, vol. 6, no. 6, pp. 796–809, 2010.
[23] T. Hu and Y. Fei, "Murao: A multi-level routing protocol for acoustic optical hybrid underwater wireless sensor networks," in 9th Annu. IEEE Commun. Soc. Conf. Sensor, 2012, pp. 218–116.
[24] I. Ahmad and K. Chang, "Effective SNR mapping and link adaptation strategy for next-generation underwater acoustic communications networks: a cross-layer approach," IEEE Access, vol. 7, pp. 44 150–44 164, 2019.
[25] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," IEEE J. Sel. Areas Commun., vol. 24, no. 3, pp. 659–672, 2006.
[26] G. W. H. Wu and X. Lin, "On the effectiveness of opportunistic routing over wireless mesh networks," in 6th Int. Conf. Wireless Commun. Netw. Mobile Comput, 2010, pp. 1–4.
[27] H. X. K. Tang, C. Li and J. Zou, "Reinforcement learning-based opportunistic routing for live video streaming over multihop wireless networks," in IEEE 19th Int . Workshop Multimedia Signal Process, 2017, pp. 1–6.
[28] H. J. X. He and H. Xiao, "Routing selection with reinforcement learning for energy harvesting multi-hop crn," IEEE Access, vol. 7, pp. 54 435–54 448, 2019.
[29] Z. Z. Ying Zhang and X. Wang, "Reinforcement learning-based opportunistic routing protocol for underwater acoustic sensor networks," IEEE Transaction on Vehicular Technology, vol. 70, no. 3, pp. 2756–2770, 2021.
[30] R. Narmeen and S. Muhaidat, "Shortest propagation delay-based relay selection for underwater acoustic sensor networks," IEEE Access, pp. 37 923–37 935, 2021.