
Article:
Park, Sunghyun, Mitchell, Paul Daniel orcid.org/0000-0003-0714-2581 and Grace, David orcid.org/0000-0003-4493-7498 (2020) Reinforcement Learning Based MAC Protocol (UW-ALOHA-QM) for Mobile Underwater Acoustic Sensor Networks. IEEE Access. pp. 5906-5919. ISSN: 2169-3536

https://doi.org/10.1109/ACCESS.2020.3048293

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 13, 2020, accepted December 28, 2020, date of publication December 30, 2020,
date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048293

Reinforcement Learning Based MAC Protocol (UW-ALOHA-QM) for Mobile Underwater Acoustic Sensor Networks

SUNG HYUN PARK, (Graduate Student Member, IEEE), PAUL DANIEL MITCHELL, (Senior Member, IEEE), AND DAVID GRACE, (Senior Member, IEEE)
Department of Electronic Engineering, University of York, York YO10 5DD, U.K.

Corresponding author: Sung Hyun Park (sp1356@york.ac.uk)

The work of Paul Daniel Mitchell was supported in part by the U.K. Engineering and Physical Sciences Research Council under Grant EP/P017975/1.

ABSTRACT The demand for ocean exploration and exploitation is rapidly increasing and this has led to rapid growth in the market of mobile vehicles. Given the mobility, the key challenge is to design a highly adaptive solution with minimal signalling (and the associated delays) which current techniques have not fully addressed. Therefore, the mobility and associated challenges in the underwater channel necessitates the design of a new approach to Medium Access Control (MAC) which provides the capability to adapt to rapidly changing conditions with no reliance on signalling which causes delays. This paper proposes the UW-ALOHA-QM protocol, which uses reinforcement learning to allow nodes to adapt to the time varying environment through trial-and-error interaction and thereby improve network resilience and adaptability. Simulations are carried out in four distinct scenarios in which node mobility patterns are significantly different. Simulation results demonstrate that UW-ALOHA-QM provides up to 300% improvement in channel utilisation with respect to existing protocols designed for mobile networks.

INDEX TERMS Medium access control, mobile sensor networks, reinforcement learning, Q-learning, underwater acoustic networks.

I. INTRODUCTION

The marine environment is becoming increasingly important to a vast diversity of industries and is subject to rising scientific interest. There are a huge range of applications for underwater networks, such as: pollution monitoring (e.g. ocean plastics [1]), seismic detection (Tsunami warning system [2]), military (notably underwater surveillance [3]), and managing food stocks (e.g. fish farms [4]). However, most of the ocean has not yet been explored, since ocean exploration is hampered by the hostile and harsh underwater environment. The traditional approach to underwater monitoring is highly inefficient. Sensor nodes need to be carried to the sea by ship and deployed to collect data for the mission period. At the end of the mission, sensors typically need to be retrieved and taken back to the shore, so the collected data can then be analysed. Underwater communication offers the potential for continuous long-term monitoring from land, at much lower cost, but amongst many other things, it is reliant on efficient and adaptable networking.

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

The use of Wireless Sensor Networks (WSNs) to monitor and collect data in the terrestrial environment has been intensively researched. Unfortunately, well established radio technologies cannot be directly applied to the underwater environment since radio signals are heavily absorbed by water. Acoustic signals are the most feasible means of communicating underwater due to their ability to propagate over long distances compared to alternative signals. However, acoustic signals have a slow propagation speed ($\approx$ 1,500 m/s) in water compared to radio signals in the air ($\approx 3 \times 10^8$ m/s). The long propagation delays that result make it difficult to achieve high channel utilisation in underwater networks. In addition, the limited and distance dependent bandwidth results in low fundamental channel capacity [5]. Furthermore, any reliance on time synchronisation for data communication is costly and complex in the underwater environment since GPS signals are not available.

The objective of the Medium Access Control (MAC) layer is to make effective use of shared a channel by coordinating the multiple access of nodes to provide high channel utilisation and good Quality of Service (QoS). Therefore, the MAC layer plays a vital role in underwater acoustic networks in

an effort to maximise channel utilisation in the presence of the slow propagation speed. MAC protocols can generally be categorised as centralised or distributed. Centralised protocols can often achieve good channel utilisation through collision-free scheduling but require infrastructure such as a coordinating central node. Distributed MAC protocols do not require such infrastructure, but distributed scheduling or reservation schemes incur significant overheads for processes such as neighbour discovery, schedule exchange, and associated handshaking techniques. Techniques based on carrier sensing are less effective for underwater acoustic communication due to the long and variable propagation delays, often requiring long guard bands. Centralised protocols are more appropriate for static networks where a coordinating node knows (or is able to gather) information from all nodes such as their locations, transmission priorities, or individual traffic loads. Therefore, transmission scheduling can be relatively static and potentially pre-defined by a central node. Distributed protocols are necessary for networks where centralised scheduling is not feasible. The problem is that the signalling overheads of distributed protocols can impair channel utilisation, especially when the propagation delay is significant as it is in underwater acoustic networks.

Reinforcement learning is a form of machine learning which enables agents (nodes) to learn an optimal action at each unit of time (called an epoch) through trial-and-error interactions in a dynamic environment [6]. The underwater environment continually changes and thus underwater nodes need to be capable of adapting to such time-varying changes. Therefore, it is expected that a reinforcement learning based protocol can offer underwater networks this capability by interacting with the underwater channel. Previously, the authors designed a reinforcement learning based protocol for quasi-stationary networks and the initial studies [7], [8] have shown that the distributed protocol can achieve collision-free access without the need for time synchronisation in underwater networks comprising fixed nodes. Therefore, it is of interest to explore the use of reinforcement learning for mobile underwater networks. Although such learning algorithms will be unable to converge in such dynamic environments, they have the potential to track changes in the environment at a rate which can significantly reduce the probability of collision with respect to alternatives such as random access. This capability, combined with minimal overheads and low complexity, is favourable with respect to more complicated distributed protocols.

Specific contributions of this paper include:

- This is the first paper to the authors' knowledge to explore the use of reinforcement learning for mobile underwater networks.
- This paper introduces a reinforcement learning based protocol (UW-ALOHA-QM) to provide resilience and adaptability in response to environmental changes in mobile underwater networks.
- This paper demonstrates that reinforcement learning techniques are a powerful means of providing a flexible topology agnostic solution for medium access control in underwater networks without the need for time synchronisation.
- It is shown that UW-ALOHA-QM can provide significantly enhanced channel utilisation in a range of distinct scenarios compared to alternative existing protocols designed for mobile underwater networks.

TABLE 1. Comparision between UW-ALOHA-QM and existing protocols for mobile underwater networks.

Features	UW-ALOHA-QM	Current protocols
Time synchronisation	No	Commonly used [10][12][13][14][15]
Handshaking	No	Commonly used [12][13][14][15]
Carrier sensing	No	Some protocols use [10][14]
Localisation information	No	Typical approach [10][12][13][14][15][16]

Compared to other non-reinforcement learning methods, the key strength of UW-ALOHA-QM is that it is: 1) a fully distributed algorithm which allows nodes (and a network) to self-organise and 2) It provides a level of resilience and adaptability to cater for node mobility and the constantly changing environment. A potential weakness is that UW-ALOHA-QM requires sufficient iterations to learn the environment otherwise, it cannot achieve the desirable channel utilisation. The protocols performance and ability to support mobility are evaluated and demonstrated for a wide range of scenarios in section IV. Table 1 compares key features of UW-ALOHA-QM and current MAC protocols designed for mobile networks.

Section II of this paper provides a literature review on protocols designed for mobile underwater networks and reinforcement learning based protocols for fixed underwater networks. In section II, subsections A to E describe the design process and related parameter settings and subsection B in particular presents the reinforcement learning algorithm used in UW-ALOHA-QM. Section III introduces UW-ALOHA-QM for mobile networks. This section discusses node mobility in underwater networks and analyses the impact of node mobility on the reinforcement learning process. Section IV presents a comparative performance evaluation through simulation, showing the key performance characteristics of UW-ALOHA-QM under various network configurations. Finally, section V concludes the paper.

II. PREVIOUS WORK

This section reviews existing protocols in two parts. First, a brief review of state of art protocols for mobile underwater networks is provided. Node mobility is a critical issue to consider in protocol design since some new considerations and tasks arise due to time varying node locations and the need to provide resilience [9] to rapidly changing environmental conditions. Secondly, existing reinforcement based MAC protocols are reviewed and it is observed that the

majority of them consider only pseudo static networks due to the emphasis on reaching a converged state.

Location based TDMA MAC (LTM-MAC) [10] is an extended version of Location Based MAC (LT-MAC) [11]. LT-MAC is designed for fixed networks and LTM-MAC is designed to support the use of Autonomous Underwater Vehicles (AUVs) in conjunction with fixed nodes. LTM-MAC assumes time synchronisation and adds carrier sensing to support data packet transmission from the AUVs. First, the reliance on time synchronisation in the underwater environment is potentially costly and complex since GPS signals are not available. Although it may be feasible in some instances to synchronise nodes prior to development, clock drift is likely to be a problem for the envisaged long-term monitoring applications. Moreover, the carrier sensing mechanism added to cope for AUV mobility requires long guard bands due to the long propagation delays, otherwise it cannot operate effectively. This represents a significant overhead with respect to channel utilisation.

Delay-aware Opportunistic Transmission Scheduling (DOTS) [12] is a distributed protocol which is designed primarily for fixed node deployments, but the study in [12] investigates the protocol in mobile networks as well. Nodes overhear one-hop neighbour transmissions for neighbour discovery and build a propagation delay map. Using the map, the protocol is able to appropriately schedule concurrent transmissions. The map quickly becomes out of date if a node moves continuously, hence DOTS uses guard bands in the scheduling to accommodate some changes after the map is updated. It uses RTS-CTS handshaking for channel reservation and requires time synchronisation across all nodes in a network. Adaptive MAC [13] uses RTS-CTS handshaking but one CTS packet can correspond to multiple RTS messages received during a RTS waiting period in order to reduce the number of control messages exchanged. Load-adaptive CSMA/CA MAC [14] is designed for single hop networks and uses RTS-CTS handshaking. It has two operational modes based on traffic load. In the high-load mode, one node can send two data packets after one handshaking process to decrease the number of control message exchanges. As the protocol name suggests, this protocol uses carrier sensing. If the channel is sensed busy, a Binary Exponential Back-off (BEB) algorithm is used which reduces achievable channel utilisation. Juggling like Stop and Wait (JSW) based MAC [15] also uses RTS-CTS handshaking and assumes multi-channel use.

Asymmetric Propagation Delay aware TDMA (APD-TDMA) [16] is designed for AUV networks and is an extension of TDMA without clock synchronisation MAC (TDA-MAC) [17] for static networks. This protocol estimates the future locations of AUVs using the data packet arrival times in the previous cycle. Therefore, this protocol is appropriate for networks comprising nodes moving at a constant velocity, rather than those comprising nodes with dynamically varying speed or direction changes.

Most protocols [10]–[15] use handshaking processes to reserve the channel, but the duration of such procedures means that this process can struggle to keep up with the topology changes in networks comprising mobile nodes. Also, frequent control message exchanges for neighbour discovery or channel reservation can lead to long idle times in the channel, high overheads, and low channel utilisation, especially in underwater acoustic networks due to the slow propagation speed. Moreover, in the case of JSW [15], the required multi-channel operation is not easily realisable for underwater acoustic networks since the usable channel bandwidth is so limited, especially over longer distances. APD-TDMA [16] is a distinct protocol because it estimates the future locations of nodes, however it is not an efficient scheme when nodes move at variable speeds or in different directions because it estimates future locations of AUVs based on the latest data packet arrival time at the central node.

There have been a number of studies applying reinforcement learning to the medium access control problem in terrestrial WSNs and the results are promising [18]–[24]. However, there have been few publications on such approaches for underwater networks. A few papers propose routing algorithms [25]–[29] and we have only found four studies recently published [30]–[32], [43] for fixed node underwater networks excluding our own prior work [7], [8].

A conference paper [30] discusses the use of slotted ALOHA in distributed networks comprising fixed nodes. It assumes time synchronisation across the network. A large enough slot is divided into a data transmission phase and an acknowledgement (ACK) phase. Using reinforcement learning, each node learns a suitable transmission order, which means an individual node finds a distinct slot to send a data packet in a frame in a slotted time system. Once the order is determined, the protocol omits the ACK phase and can increase channel utilisation. However, the protocol is highly vulnerable to any future changes because the protocol cannot be aware of such changes without feedback (ACKs) about the data transmissions.

Two conference papers [31], [32] were published in 2019. They are inspired by a journal paper [24] which discusses Deep Reinforcement Learning (DRL) for heterogeneous wireless networks. Underwater and terrestrial environments are totally different and it is not therefore efficient if the wireless techniques are directly used in underwater networks. First, the journal paper [24] considers a heterogeneous wireless network including LTE user equipment and WiFi devices. Time synchronisation and high propagation speeds are considered in the wireless network which are not appropriate to underwater networks. Lastly, DRL increases complexity and wastes computing resources in determining only a transmission order for fixed nodes in the underwater networks.

One of the conference papers [31] suggests a time synchronised mode and two non-synchronised modes to setup different simulation scenarios and the synchronised network

shows the best performance in terms of the average channel throughput. This paper ignores the propagation delay of ACK packets from the sink nodes, which is not a practical assumption. The paper does not describe how the required propagation distance information (estimates) can be obtained, yet the information is a prerequisite for data transmission.

Another conference paper [32] assumes time synchronisation and acquires estimates of the distances between a sink node and sensor nodes before data transmission subsequently takes place through beacon message exchanges. Due to the long propagation delay, the scope of the learning history for the current action does not include recent feedback. This reward approach can work in fixed networks, but it cannot work properly in mobile networks since the environmental conditions associated with the previous experience has significantly changed.

The most recent study [43] has been published in September 2020. It proposes to use deep learning for channel selection in a multi-channel system. The protocol uses a slotted ALOHA structure and assumes time synchronisation. Simulation compares results obtained with the learning scheme, random selection, and optimised traditional selection which requires the network information in advance. Random selection shows the worst channel utilisation and the optimised selection shows the best performance. The learning approach does not achieve the best throughput at the beginning of the simulation, however it approaches the optimised throughput after sufficient learning iterations. The study also wastes computing resources in that it merely selects one channel in a slot. Moreover, the acoustic signalling channel is very limited so that the multi-channel system is not ideal for the underwater communication.

In summary, most of the existing protocols for mobile underwater networks are extended versions of MAC protocols designed for networks comprising fixed nodes where time synchronisation is assumed. Most of them add extra functions such as frequent control message exchange or carrier sensing with long guard bands to handle node mobility. However, these solutions incur high propagation delay or low channel utilisation hence they are not efficient in mobile underwater networks. Rather than these supplementary measures to deal with node mobility, the learning approach provides network adaptability which can achieve good channel utilisation, low overheads, and low complexity in the face of changes in the network. Moreover, all existing reinforcement learning based protocols designed for underwater networks consider networks comprising fixed nodes. Reinforcement learning is potentially effective in a time-varying environment since it provides inherent adaptability based on continued interaction with an environment. With regard to underwater networks comprising mobile nodes, we cannot seek convergence. The effectiveness of such an approach boils down to whether the learning algorithm is able to adapt at a sufficient speed with respect to the key environmental changes. We propose what we believe to be the first reinforcement learning based MAC protocol (UW-ALOHA-QM) for mobile underwater networks and investigate its potential to provide a topology agnostic approach to medium access control.

III. UW-ALOHA-QM

Mobility always causes complexity in a network since it brings a lot of variability to the network including more significant time-varying channel conditions, changes in connectivity and propagation delays. Therefore, node mobility represents a specific challenge which needs to be addressed in the design of MAC protocols [9].

For static topologies, it has been shown that it is possible to achieve a scheduled outcome from initial random access, through the learning process, to achieve a high channel utilisation. The merit of employing such an approach lies in the inherently distributed nature of typical algorithms such that there is no reliance on infrastructure, making it a useful approach for a wide range of network topologies and potentially those with changing connectivity over time. Typical algorithms are also characterised by low signalling overheads and low complexity. In a mobile network, convergence is unlikely to be achieved, and it would otherwise be very short lived. Therefore, network resilience needs to be considered in the mobile network. We consider network resilience to be the ability to provide and maintain a good level of service in the face of changes to normal operation [33]. Reinforcement learning provides a means of adapting to a time-varying environment, with nodes learning from their experience. If the learning process can be sufficiently rapid with respect to the changing environment, then reinforcement learning based MAC protocols can provide useful adaptation in dynamic environments and achieve superior performance with respect to the alternative approaches that are known in the literature.

The desired capability of a reinforcement learning based MAC protocol for mobile networks is to provide more effective adaptation to the time-varying environmental conditions such that an improved level of performance (e.g. in terms of channel utilisation) can be achieved with respect to baseline protocols that do not incorporate learning. Superior channel utilisation performance can be potentially achieved with respect to alternative state of the art protocols owing to the minimal signalling overheads and absence of inefficient handshaking procedures.

In Fig. 1, it is expected that a standard distributed protocol which is designed with the appropriate guard time is able to withstand any envisaged changes in environments.

For example, if the propagation delay changes through mobility, it is expected that the protocol has sufficient guard bands to deal with that mobility. On the other hand, with the learning scheme, it is expected that the learning process iterates for nodes in a static or pseudo-static environment and the learning approach is able to converge on a stable solution. However, if there are any changes in the environment convergence cannot be maintained. Fig. 1 shows an example where there are quite significant changes in the environment at discrete times. This will cause the learning process to be disturbed and the performance would be expected to drop.

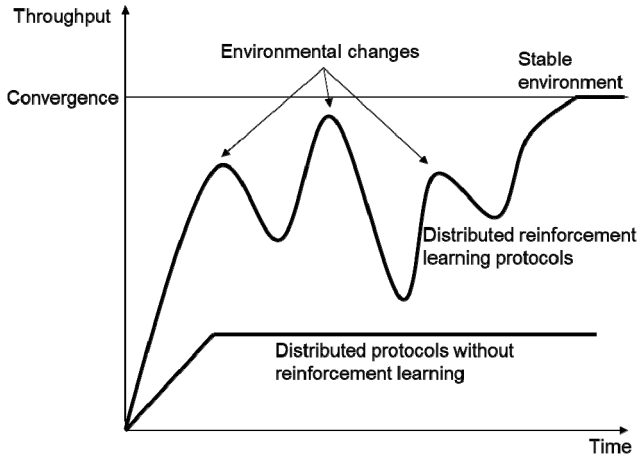


FIGURE 1. Adaptability and resilience.

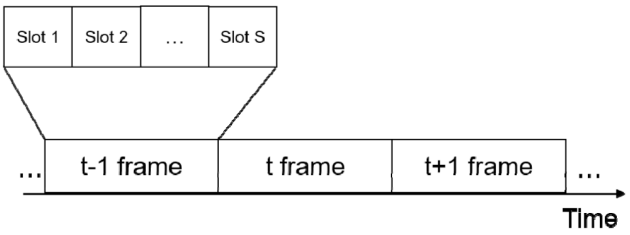


FIGURE 2. The fundamental frame structure used for UW-ALOHA-QM.

However the learning approach can then start to improve the situation again until there is another significant change in the network.

A. UW-ALOHA-QM FRAME STRUCTURE

Time is divided into repeating frames where each frame comprises a number of slots as depicted in Fig. 2. The main idea of UW-ALOHA-QM is to apply reinforcement learning to enable each node (agent) to independently learn a preferred slot in the repeating frame structure based on past experience and to send data packets in the preferred slots of successive frames. For UW-ALOHA-QM, nodes are not assumed to be synchronised and the frame start times are therefore considered to be randomly distributed.

An example of an UW-ALOHA-QM network is a simple topology where four different sensor nodes (N1, N2, N3, and N4) are deployed at different distances from a sink node. The four nodes collect data and transmit the information to the single sink node. Each node in UW-ALOHA-QM is permitted to transmit in one slot per frame. For this example network, each frame comprise four slots (the frame size, S = 4) for the four nodes in the network (N = 4) shown in Figs. 3 and 4.

Therefore, each node has an opportunity to transmit collected data once per frame and needs to select one of four slots in each successive frame to transmit a data packet. In the example shown, N1 uses slot2 and N2 uses slot2. (Note that data transmission flows for N3 and N4 are omitted in Fig. 3 for the purpose of simplicity).

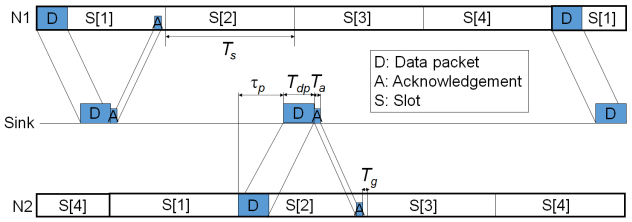


FIGURE 3. An example of UW-ALOHA-QM timing for a four-node network.

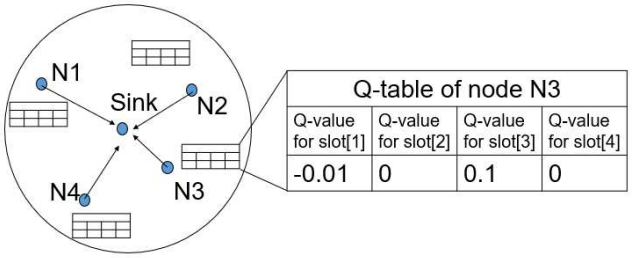


FIGURE 4. Example of Q-table of UW-ALOHA-QM.

Each individual slot is designed to support the transmission of a data packet to the sink node and reception of an acknowledgement (ACK) packet back. To achieve this, the slot duration needs to account for the maximum propagation delay from a mobile node to a receiver in both directions and incorporate a guard band. This very small guard band is merely for the case that the maximum delay is underestimated. One slot duration (T_s) is sufficient to accommodate a data packet (T_{dp}), twice the propagation delay (τ_p), an ACK packet (T_a), and guard time (T_g). The slot duration (T_s) can be calculated by Eq. (1).

$$T_s = (T_{dp} + T_a + T_g) + 2\times\tau_p \tag{1}$$

UW-ALOHA-QM uses ACKs to determine if data packets are delivered. This serves to provide reliable communication but is additionally a key requirement for operation of the learning algorithm. After sending a data packet, if the generating node does not receive an ACK from the sink node before the guard time ends, the transmission is assumed to have failed and a retransmission must be initiated.

The most significant element which impacts upon slot duration (T_s) is the propagation delay (τ_p) as Eq. (1) shows. Therefore, the duration becomes large in typical underwater networks because of the slow propagation speed. During the propagation of a data packet and ACK, the channel remains in an idle state as shown in Fig. 3. This overhead is a potential problem and a key difference in applying this frame structure to underwater acoustic networks with respect to terrestrial radio networks. It can be overcome, however, as described in subsequent sections of this paper.

B. Q-LEARNING

UW-ALOHA-QM is based on stateless Q-learning [6], which is used where an environment does not have to be represented by state. In the UW-ALOHA-QM protocol, nodes use the learning scheme to choose a distinct slot in each frame to

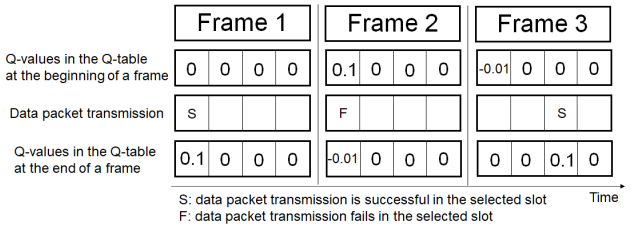


FIGURE 5. Example of Q-table update process of N3 for the first 3 frames.

transmit one data packet. All nodes maintain a Q-table which contains individual Q-values (one for each slot in the frame) as Fig. 4 shows.

Eq. (2) is used to determine how Q-values are updated:

$$Q_{t+1}(i,k) = Q_t(i,k) + a(r - Q_t(i,k)) \tag{2}$$

where, the ith node has sent a data packet in the kth slot in a frame. Q_t is the Q-value at time t, t is a time epoch (i.e. a frame), a is the learning rate, and r is the reward value. A standard implementation of ALOHA-Q [21] uses $a = 0.1$ and $r = 1$ if the transmission is successful (i.e. a generating node successfully receives an ACK) otherwise, $r = -1$.

Consider the example shown in Fig. 5, depicting the Q-table at a single node. Since all Q-values in the Q-table are initially zero, the node randomly selects a slot in the next frame for data packet transmission. If the node receives a positive ACK before the guard time ends, meaning the transmission was successful, the Q-value for the 1st slot in the Q-table becomes updated to 0.1 as shown through the application of Eq. (2). Thus, after one frame, the Q-table has Q-values of 0.1/ 0/ 0/ 0 and the 1st slot has the highest Q-value in the node's Q-table.

At the start of the second frame, the node will transmit a data packet in the 1st slot since the Q-value of the slot has the highest value (i.e. 0.1) in the node's Q-table. If the node does not receive an ACK packet before the guard time ends, the node assumes that the transmission has failed and the Q-value for the 1st slot in the Q-table is updated to -0.01. Therefore, after the second frame, the Q-values of the Q-table are -0.01/ 0/ 0/ 0.

At the beginning of the third frame, the node chooses between the 2nd, 3rd, and 4th slots at random, with equal probability since they all have the same highest Q-value of zero. By repeating this trial-and-error learning, and as long as there are sufficient slots in a frame, it can be shown that individual nodes are able to find distinct slots to transmit in, and thereby avoid collisions with other nodes in networks where the environment is sufficiently stationary [8].

C. ASYNCHRONOUS OPERATION

Asynchronous operation is proposed for UW-ALOHA-QM since GPS signals are not available underwater. Reliance on time synchronisation in the underwater environment is potentially costly and complex and clock drift is likely to be a problem for the envisaged long-term monitoring applications.

TABLE 2. Channel utilisation according to the frame size (s) in a network of 100m radius (r) comprising 25 nodes.

Frame size (S)	Index ratio (B)	Average channel utilisation (Erlangs)	Theoretical max channel utilisation (Erlangs)
4	1.44	0.44	0.69
5	1.80	0.46	0.55
6	2.16	0.42	0.46
7	2.51	0.36	0.40
8	2.87	0.34	0.36
…	…	…	…
25	8.98	0.11	0.11

Fig. 3 provides an example of the asynchronous timing of UW-ALOHA-QM. The two generating nodes N1 and node N2 start their frames at different moments and the sink node does not need to work to a frame structure. Therefore, all nodes in the network (including the sink node) do not need to synchronise to other nodes and each node operates completely independently.

It is expected that collisions occur at the sink node in the absence of time synchronisation, since data packets sent from sensor nodes will arrive at a sink node at random times, but UW-ALOHA-QM can achieve an identical channel utilisation with and without time synchronisation in the underwater environment where there are the large number of nodes with relatively large propagation delays. This is observed from the intensive simulation for networks comprising more than 25 nodes and the full details are shown in [8]. This is achievable as the significant idle time at the sink node is large enough to accommodate data packets sent from all sensor nodes in a frame. In addition, reinforcement learning allows nodes to learn the distinct slot in which a data packet can be successfully received in the idle time at the sink node when there are sufficient slots in a frame.

D. REFINEMENT OF FRAME SIZE

The standard implementation of ALOHA-Q schemes where the frame size (S) is set equal to the number of nodes (N) is not efficient for this underwater network due to the long slot duration and the different and time varying propagation delays from mobile nodes. Therefore, it is useful to explore how we can refine the frame size (S) to improve the theoretical channel utilisation. Table 2 shows the difference in achievable channel utilisation according to the frame size (S) from [8]. For each result shown in the Table 2, 100 simulations were carried out and each simulation run had different and random frame start times for all nodes. The theoretical maximum channel utilisation values in the table refer to the channel utilisation which UW-ALOHA-QM achieves in a network comprising fixed nodes.

The index ratio (B) is introduced in the previous study [8] and represents the ratio between 'the total available time which can be used for the sink node to receive in a frame' and 'the sum of data packet durations generated by all nodes in a frame'. The index (B) can be expressed by Eq. (3). This index represents the theoretical available space at the sink

node to be used for reception of data packets related to the duration within a frame that is used for data bits. For example, in Table 2, when N = S = 25, the sink node theoretically has 8.98 times more space than the duration required to receive one packet from each of the 25 nodes in a frame.

$$B = \frac{S \times (2 \times \tau_{\mathrm{p}} + T_{dp})}{N \times T_{dp}} \tag{3}$$

Table 2 shows a trade-off between frame size (S) and the average channel utilisation. In the table, as the frame size (S) is reduced, the average channel utilisation is improved. However, at a certain threshold, the channel utilisation decreases because the level of contention increases. It refers to the fact that more slots in a frame wastes channel capacity because there will be a number of unused slots in each frame whereas an insufficient frame size (S) will not provide a sufficient duration for the contending nodes to find collision free spaces. Therefore, it is necessary to find an optimum frame size to maximise the average channel utilisation.

Referring to the simulation results of [8], it is found that channel utilisation of UW-ALOHA-QM can be improved by reducing the frame size (S) up to an index ratio (B) value of 1.5 when 25 nodes are used in a network sized from 100 m to 1000 m [8], otherwise (i.e. B<1.5) the channel utilisation decreases due to collisions. For example, in Table 2, UW-ALOHA-QM shows the highest channel utilisation using 5 slots in a frame (B=1.8>1.5) and the channel utilisation decreases when 4 slots per frame (B=1.44<1.5) is used since the index (B) value is under 1.5. Therefore, this paper proposes to use the optimum frame size called S_m which is the smallest frame size for maximum channel utilisation under a condition that the index (B) is equal to or greater than 1.5 with given network size (R) and the number nodes in a network (N). This is a generally applicable approach, and in order to aid understanding, some very specific parameters are described in this paper to serve a specific example.

E. UNIFORM RANDOM BACK-OFF

Although a small number of slots per frame (S) is desirable with respect to higher channel utilisation, an additional mechanism is required to achieve this. A small number of slots corresponds to a limited action space and distinct transmission timings for a relatively large number of nodes. For example, in Table 2, the optimum frame size (S_m) is 5 for 25 nodes. In this case, all nodes have only 5 slots in a frame and this implies that there is high possibility of residual contention at the sink node with only five transmission time options. It is necessary to allow nodes to additionally adjust their frame start times, such that they can offset themselves with respect to other nodes and fill the idle time at the sink node. In other words, each node also needs to learn an appropriate frame start time and not only the distinct slot to avoid collisions.

A Uniform Random Back-off (URB) scheme is proposed to provide the opportunity for nodes to adjust their frame start time. URB can improve the flexibility of UW-ALOHA-QM in mobile networks by adjusting frame timing according to the constantly changing network environment caused by node movement.

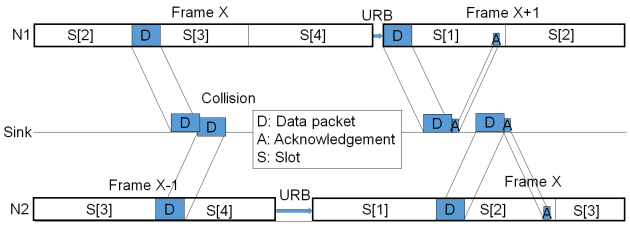


FIGURE 6. Concept of Uniform Random Back-off (URB).

Fig. 6 shows the concept of URB. Note that data transmission flows for N3 and N4 are omitted in Fig. 6 for the purpose of simplicity. URB provides chances nodes to adjust their frame start time, but significantly decreases the channel utilisation because of two reasons. First, the action currently taken is based on learning conducted in different network circumstances in the past, which means neighbour nodes have moved so that their locations have changed. Therefore, the largest Q-value (Q_t) in the Q-table is not always the best action for a node, so transmitting a data packet in the selected slot can generate collisions in the mobile network. Moreover, mobility makes learning of UW-ALOHA-QM 'myopic' [34]. Every collision initiates URB and moving the frame start time brings about a new network configuration for a node. This frequent URB wastes historical experience since the optimal action is based on heuristic awards and punishment. Consequently, URB (the new frame start time) causes a situation that all nodes must learn the new environment from the scratch at every frame resulting in inefficient and unnecessary learning processes.

Therefore, a new URB design is required to achieve the more efficient learning. The URB must be initiated only when a node can determine that the current highest Q-value is not the optimum action. Using Eq. (2), we can calculate when a node needs to trigger a new learning process. Assuming a node experiences collisions at every transmission because of mobility and setting the initial Q-value to 1 (i.e. $Q_0 = 1$), the Q-value is changed from $1 \rightarrow 0.8 \rightarrow 0.62 \rightarrow 0.458 \rightarrow 0.3122 \rightarrow 0.18098 \rightarrow 0.062882$ and to -0.043406 at a 7th consecutive collision. Seven consecutive failures cause the Q-value to return to $\approx$ zero at a learning rate (a) of 0.1. The previous study [35] carried out the analysis of Q-value in a radio wireless network and shows the same result.

Therefore, this paper proposes the 7-Uniform Random Back-off (7-URB) scheme which raises the URB scheme after a seven consecutive collision for the mobile network. 7-URB utilises the experienced Q-value and removes unnecessary learning processes. As a result, UW-ALOHA-QM can improve network resilience and adaptability and avoid collisions by adjusting its frame timing.

IV. SIMULATIONS

Simulations have been carried out in order to evaluate the capability of this reinforcement learning based MAC

protocol, UW-ALOHA-QM for mobile underwater acoustic networks. Four distinct scenarios have been modelled and the performance is evaluated in these scenarios. The intention of simulations in four different scenarios is to demonstrate the potential of UW-ALOHA-QM when applied to distinct scenarios, to better illustrate the scope of its use and provide an insight into how UW-ALOHA-QM performs with respect to different protocols designed for each type of network.

These scenarios broadly are

- Moored or anchored sensor network
- Free floating sensor network [12]
- AUV assisted network [10]
- AUV sensor network [16].

The first scenario is a reference scenario serving to illustrate the fundamental operation of UW-ALOHA-QM with typical parameters. The other three scenarios and corresponding parameters are defined in other MAC protocol studies, [12], [10], and [16]. These four scenarios have been considered to provide comprehensive evaluation of UW-ALOHA-QM and have been chosen for two primary reasons: 1) to provide very distinct mobility setups and cases and to provide a wide evaluation of the capability of UW-ALOHA-QM. 2) The latter three scenarios and parameters are taken from the literature [12], [10], and [16] of other MAC protocol research which have been developed for the specific scenarios and therefore this allows a direct comparison of UW-ALOHA-QM with the results presented for the state of the art schemes presented in those papers. For each scenario, results from the respective paper from which a scenario is taken have been extracted. This typically includes the scheme which was proposed by the authors and also some other comparative schemes. In addition, UW-ALOHA-QM is simulated in their scenario.

In these scenarios, the channel utilisation of UW-ALOHA-QM is measured at the sink node. Channel utilisation is evaluated as the fractional amount of time in which data traffic is successfully received at the sink node. Eq. (4) shows how the channel utilisation (U) of UW-ALOHA-QM is measured:

$$U = \frac{D}{r_{uw} \times F \times S \times T_s} \tag{4}$$

where, D is the total number of data bits successfully received at the sink node, r_{uw} is the data rate in bps, F is the total number of framed measured, S is the number of slots in a frame, and T_s is the duration of a slot.

A. MOORED OR ANCHORED SENSOR NETWORKS

This scenario represents underwater networks which consists of moored or anchored nodes. To show the network resilience of UW-ALOHA-QM, this discontinuous movement scenario is considered where anchored or moored nodes move according to currents at random speeds with the assumption that nodes are spatially correlated. Spatial correlation is generally used as a fundamental assumption for studies of underwater node localisation [38]–[40].

TABLE 3. Typical UW-ALOHA-QM parameters.

Parameter	Value
Duration of a data packet of 1044 bits (T_{dp})	75.108 ms
Duration of an acknowledgement packet of 20 bits (T_a)	1.439 ms
Duration of a guard time of 36 bits (T_g)	2.59 ms
Duration of a slot (T_s)	212.47 ms
Distance between a generating node and a sink node (R)	100 m
Tx/Rx data rate (r_{tr})	13,900 bps
The number of generating nodes (N)	25 nodes
Propagation speed (v_{tr})	1500 m/s
Maximum propagation delay (τ_p)	66.6667 ms
Frame size (S_m)	14
Maximum theoretical channel utilisation	0.631 Erlangs

There are 25 sensor nodes in a single-hop random topology where generating nodes are located randomly within a circular coverage area with one sink node located centrally. All nodes are within interfering range of each other. All lost packets are due to packet collisions. To provide a worst case model, any overlap in packet reception is considered to result in the complete packet being lost.

Typical parameters for UW-ALOHA-QM are listed in Table 3. Data packet size, ACK size, and guard time size in bits are based on previous studies [21]. For practical underwater environment settings, the data rate (r_{uw}) of 13,900 bps is chosen by referring to an underwater modem currently on the market [41]. In terms of node speed, this paper refers to a velocity profile at tidal-stream energy sites [42] in the sea between Ireland and Britain. It shows that the tidal stream speed is less than 4 m/s between 0 to 40 m above the seabed. Therefore, this scenario uses random speeds for node mobility between 2 to 4 m/s. Nodes will start at a uniformly distributed random position in 2 dimensions, within a 100 m radius circle (R). Each node starts the first movement at 30 min, the second movement at 60 min, and the last one at 90 min. For each movement, nodes move in a random direction for a period of 30 seconds at a random speed which is in the range between 2 to 4 m/s and the actual value is uniformly distributed. The movement direction is randomly chosen in a 0 to 2π radius. In order to demonstrate the impact of mobility with respect to stationary we consider some events which are short periods of motion intermittent with peers that are assumed to be stationary.

Fig. 7 shows the changes in channel utilisation of UW-ALOHA-QM over time and demonstrates the network resilience of the protocol. As soon as the network is deployed, all nodes initiate a learning process and can achieve the theoretical channel utilisation. After 30 minutes, all nodes simultaneously start to move (for example, by waves) and this leads to changes in node locations and hence the topology and propagation delays are changed as well in the network. Therefore, nodes need to learn the new environment and can achieve the maximum channel utilisation again. This demonstrates that UW-ALOHA-QM is able to learn and adapt to changes in the network without a coordinating node or additional control message exchanges.

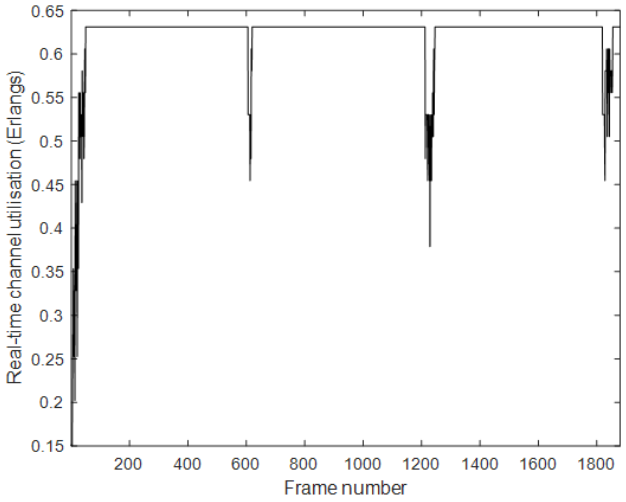


FIGURE 7. Real time channel utilisation of UW-ALOHA-QM in the moored or anchored sensor network scenario.

The theoretical maximum channel utilisation represents the stable channel utilisation which UW-ALOHA-QM achieves in the network comprising fixed or quasi-fixed nodes. In that situation, all nodes use the same slot numbers and maintain timing of a frame thus a centralised data transmission pattern is formed and this pattern is repeated. Therefore, the theoretical channel utilisation can be determined by considering the proportion of time available for data transmission in a single frame [8] as shown in Eq. (5). This moored / anchored mobile scenario, for example, achieves 0.631 Erlangs using parameters from Table 3: (25 nodes × 75.108 ms) / (14 slots × 212.47 ms).

$$Theoretical\ maximum U = (N \times T_{dp})/(S_m \times T_s) \quad (5)$$

B. FREE FLOATING SENSOR NETWORKS

This type of mobile network is characterised by free floating nodes distributed by currents. UW-ALOHA-QM is evaluated and compared to DOTS [12] which is designed for free floating sensor networks. DOTS uses the Meandering Current Mobility (MCM) model [36] for node movement.

DOTS is originally designed for networks comprising fixed nodes, however, it was evaluated for networks comprising mobile nodes. DOTS uses RTS-CTS-DATA-ACK processes but allows concurrent transmissions exploiting temporal and spatial reuse. Nodes overhear one-hop neighbour transmissions and obtain neighbour node propagation delay information from the MAC headers. The MAC headers include a time stamp indicating when the data packet is sent from a sender in order to estimate the propagation delay between a sender and a receiver. This information is stored in a map in each node and each node expects the future data transmission based on the overheard information in the map. Parameters defined by DOTS are described in Table 4.

The maximum node speed is restricted to 0.3 m/s [36]. The study shows that DOTS achieves 0.2 Erlangs of channel utilisation when the offered load is above 1 Erlang. Although

TABLE 4. Parameters used for free floating scenario evaluation.

Scenario	Free Floating
Defined by	DOTS [12]
The number of nodes (N)	10 mobile nodes
Network size (R)	430 m
Data rate	50,000 bps
Packet size	512 bytes
Maximum node speed	0.3 m/s
Simulation time	50 runs × 1 hour

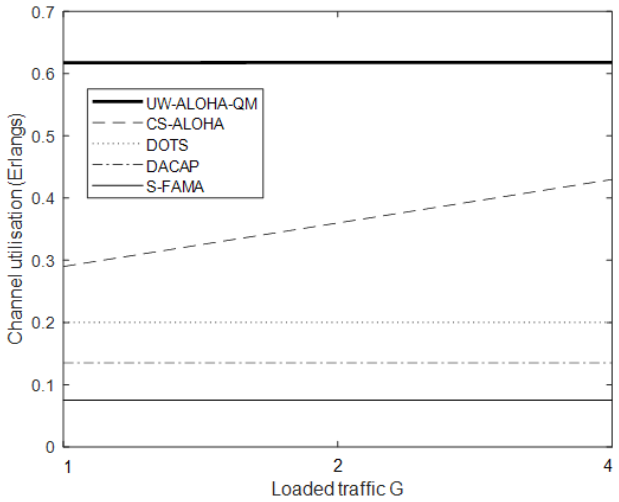


FIGURE 8. Channel utilisation according to different traffic loads.

beyond the practical operating capacity, it has been considered to evaluate DOTS under saturated traffic conditions. For a fair comparison, UW-ALOHA-QM is simulated using the parameters suggested by DOTS [12].

Fig. 8 compares the simulated channel utilisation of UW-ALOHA-QM to the other protocols as reported in [12]. Channel utilisation is measured in a consistent manner as the average value of 50 simulation runs with each simulation run lasting 1 hour. Nodes in the UW-ALOHA-QM evaluation start to move as soon as the simulation commences until the end of a simulation with the constant speed of 0.3 m/s. The only difference is that DOTS uses time synchronisation whilst UW-ALOHA-QM does not need to. The theoretical maximum channel utilisation of UW-ALOHA-QM in this network configuration is 0.624 Erlangs [8] but the protocol achieves 0.617 Erlangs due to the node mobility. The small difference in channel utilisation stands out given that the network comprises mobile nodes moving at a very slow speed.

Table 5 provides parameters for UW-ALOHA-QM in the network configurations defined by [12]. Given a network of 430 m size with 10 nodes [12], the smallest frame size under a condition that B is greater than 1.5 is 2 (S_m). In this network configuration, B is 1.6, which means a sink node has 60% more capacity than the total of 10 data packet durations. In other words, the sink node would be able to receive 16 data packets if the network was time synchronised and scheduled.

TABLE 5. Parameters used for free floating scenario.

Duration of a data packet (T_{dp})	0.08192 seconds
Duration of a slot (T_s)	0.656373 seconds
Duration of a frame (T_f)	1.312747 seconds
Frame size (S_m)	2
Index ratio (B)	1.6

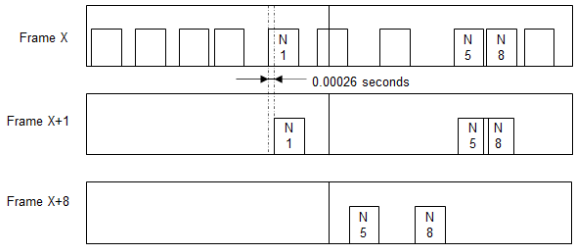


FIGURE 9. Data packet receptions at the sink node.

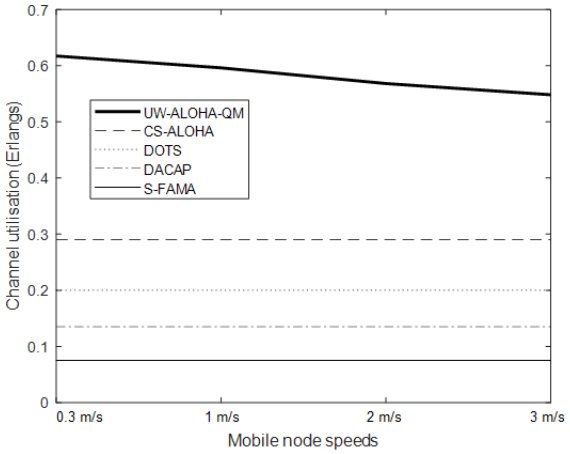


FIGURE 10. Data packet receptions at the sink node.

TABLE 6. The average number of times 7-URB is triggered.

	0.3 m/s	1 m/s	2 m/s	3 m/s
Average number triggered	32.4	134.1	267.1	362
7-URB is triggered every	85.6 frames	20.7 frames	10.4 frames	7.7 frames

Fig. 9 show the packet reception at the sink node at different frames. The sink node actually does not have the time slot and frame structure as shown in Fig. 3, but it is illustrated in Fig. 9 for an easy understanding of the theoretical concept of UW-ALOHA-QM. 10 sensor nodes send data packets to the sink node and they are not time synchronised, therefore 10 packets arrive at the sink node at random times. When the node speed is 0.3m/s, a sensor node moves 0.39 meters during a frame which results in a 0.00026 seconds change in the propagation delay per frame. This change accounts for merely 0.04% of a frame, which is very small compared to one data packet which accounts for 6.24% of a frame. Therefore, the idle time at the sink node functions as a guard band to deal with the small changes in propagation delay caused by the slow node mobility.

For example, in Fig. 9, N1 moves away from the sink node at a 0.3 m/s speed and the data packet sent from N1 of frame X+1 arrives slightly later than the previous frame X. However, the idle time at the sink node allows reception of the packet without a collision. N5 moves away from the sink node and N8 moves towards the sink node and their packets collide in frame X+1, and if the collision continues for 7 consecutive frames, the two nodes will trigger 7-URB and then they attempt a different frame start time to find an appropriate gap at the sink node. Therefore, with slow node movement (0.3 m/s), UW-ALOHA-QM can maintain good channel utilisation.

Fig. 10 compares the channel utilisation of the different protocols with various node speeds from 0.3 m/s to 3 m/s. All nodes continue to move during the simulation time for one hour. The DOTS protocol exhibits 0.2 Erlangs channel utilisation regardless of the node speed because DOTS incorporates guard bands of sufficient duration to accommodate changes in reception timing caused by node mobility and the impact this has on propagation delay. However, there is a 12% decrease in the average channel utilisation of UW-ALOHA-QM with respect to the theoretical maximum channel utilisation, with nodes moving at a 3 m/s speed. This is because the relative timing of packet reception from the different nodes at the sink changes more rapidly and the learning algorithm because less effective at adapting to the changes. The preferred slot is subject to reduction in its Q-value.

7-URB is triggered more often as the node speed increases. Table 6 provides the average frequency with which 7-URB is invoked across the 50 simulation runs for each speed. As the node speed increases, the 60% extra time at the sink node is not sufficient to deal with the high mobility. When the node speed is 0.3m/s, 7-URB is triggered on average every 86 frames whereas it is triggered more frequently (every 8 frames on average) when the node speed is 3m/s, in an attempt to find an appropriate frame start time for successful transmission.

In summary, the simulation results shows that UW-ALOHA-QM always provides a respectable channel utilisation and outperforms DOTS and other protocols in the free floating node scenario despite the asynchronous operation of UW-ALOHA-QM. DOTS uses a sufficient guard time to deal with the node movement and handshaking which significantly reduces the achievable channel utilisation. However, UW-ALOHA-QM uses the learning approach where all nodes independently learn and find a distinct slot and appropriate frame start time through their interaction with the time-varying environment, which brings about better adaptability and higher channel utilisation than other existing protocols. For example, UW-ALOHA-QM provides more than 3 times better channel utilisation (0.617 Erlangs) than DOTS (0.2 Erlangs) when the node speed is 0.3 m/s.

Figs. 8 and 10 also illustrate features of different MAC approaches. CS-ALOHA uses random access and the channel

utilisation is therefore heavily dependent on traffic load (G) but not dependent on the node speed. The DOTS, DCAP, and S-FAMA protocols conduct handshaking before data transmission and their performance is not affected by environmental changes (i.e. node speed in this scenario) since the handshaking scheme does not require prior information of the environment nor interaction with environmental changes. However, due to frequent control message exchange, the underlying performance of those protocols is very low and the handshaking process potentially fails if nodes move at a very high speed during the process in the mobile network.

On the contrary, the channel utilisation of the learning approach is related to the environmental changes since it interacts with the environment. High speed mobility implies that the network environment changes quickly. Consequently, the node speed has an impact on the performance of UW-ALOHA-QM. However, it can be seen that for this particular environment, the learning scheme is able to allow the network to adapt sufficiently rapidly to the environmental changes and achieve network resilience. Therefore, the channel utilisation of UW-ALOHA-QM can be significantly higher than other protocols.

C. AUV ASSISTED NETWORKS

These networks consist of fixed sensor nodes and one or more AUVs. The LTM-MAC [10] and Load adaptive CSMA/CA [14] protocols are designed for this type of mobile network. LT-MAC [11] was proposed for small-scale static underwater networks and LTM-MAC [10] is an extended version for the extra AUV in fixed underwater networks. LTM-MAC assumes the AUV has enough knowledge about the network topology to support the fixed sensor nodes. Basically, carrier sensing is added for the LTM-MAC protocol to handle the mobility of the AUV. However, the carrier sensing mechanism added to cope for AUV mobility requires long guard bands due to the long propagation delay, otherwise it cannot operate effectively in the underwater environment. LT-MAC and LTM-MAC are based on TDMA, therefore, time synchronisation is required and the transmission order of static nodes is decided before the data transmission. However, these protocols use dynamic time slot durations for each node based on the results obtained in the latency detection phase before the data transmission phase. Therefore, all nodes should broadcast a control message to indicate the slot duration before each data transmission.

In this AUV assisted network scenario, one AUV keeps moving throughout each simulation run whilst other nodes are static on the seabed. UW-ALOHA-QM uses the identical network configurations and parameters, but asynchronous operation is applied. With a frame size (S_m) of 6 based on the desired ratio B, the theoretical maximum channel utilisation of UW-ALOHA-QM is 0.58 Erlangs with a saturated traffic model in this scenario [8]. Table 7 summarise the parameters used in the AUV assisted scenario and they are defined in [10].

TABLE 7. Parameters used for AUV assisted scenario evaluation.

Scenario	AUV assisted
Defined by	LTM-MAC [10]
The number of nodes (N)	7 static nodes + 1 AUV
Network size (R)	1,500 m
Data rate	2,000 bps
Packet size	500 bytes
Node speed	1 to 3 knots (0.51 m/s to 1.54 m/s)
Simulation time	1,000 seconds

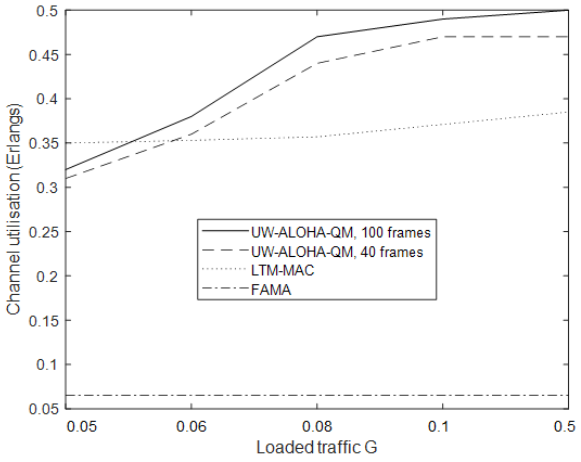


FIGURE 11. Channel utilisation according to different traffic loads.

Fig. 11 compares channel utilisation at the different traffic loads (G). LTM-MAC was evaluated for a 1,000 second period for one simulation trial, but the AUV only moves 1,540 m at a speed of 3 knots during the simulation time. Considering a network size of 1,500 m, it is not sufficient to visit every node located randomly in a circle, therefore for the UW-ALOHA-QM evaluation, additional simulations are executed with a longer simulation time of 100 frame durations for UW-ALOHA-QM as well as 1,000 seconds (40 frames).

When the traffic load (G) is very small, UW-ALOHA-QM exhibits a lower channel utilisation than LTM-MAC. If the frequency of data transmission is very small, there are insufficient trials for UW-ALOHA-QM to be able to find a suitable slot and frame start time in order to achieve collision free reception.

For the same traffic load levels, when the simulation time is longer (100 frames), UW-ALOHA-QM shows better performance since it has a longer period in which to find an appropriate transmission time. In a practical deployment, the duration of operation would of course be much longer than this and the results demonstrate that with the mobility levels in this scenario, UW-ALOHA-QM can provide higher channel utilisation than the alternatives for all but very low traffic load levels.

D. AUV NETWORKS

AUV networks consist of AUVs having sensing functionality. Path planning is generally used, for example, searching for wreckage in a zig-zag path in a crash area [37]. Therefore, the

TABLE 8. Parameters used for AUV assisted scenario evaluation.

Scenario	AUV network
Defined by	APD-TDMA [16]
The number of nodes (N)	5, 10, 15, and 20 AUVs
Network size (R)	1,500 m
Data rate	8,000 bps
Packet size	500 bytes
Node speed	5 m/s
Simulation time	10,000 cycles (frames) × 10

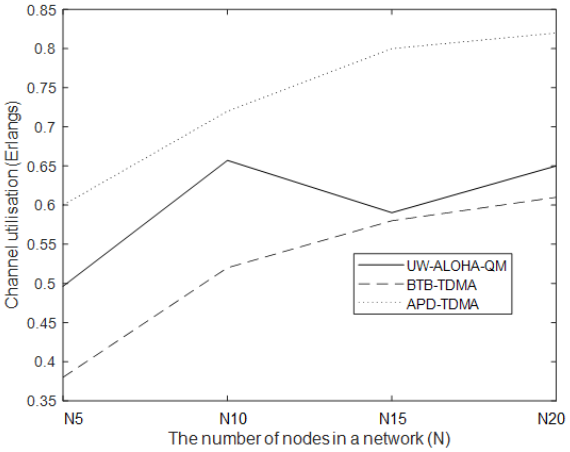


FIGURE 12. Channel utilisation in AUV network.

movement models are different depending on the application requirements. There are a couple of studies of AUV networks: APD-TDMA [16] and BTB-TDMA [43]. UW-ALOHA-QM is compared with APD-TDMA in the scenario described in [16] because APD-TDMA is the state-of-art and shows better performance than BTB-TDMA [43].

APD-TDMA [16] is designed for AUV sensor networks and it is an extension of the TDA-MAC protocol [17] designed for static networks. APD-TDMA consists of two phases: initialisation and transmission. APD-TDMA requires enough control message exchanges during the initialisation phase to get all AUV locations and then it can be ready to start the transmission phase for the data packet transmissions. A transmission phase consists of cycles which is a similar concept to frames of UW-ALOHA-QM but APD-TDMA does not use ACKs. During transmission phases, whenever the number of data packet losses at the sink node is greater than a certain value, APD-TDMA repeats the initialisation phases.

Table 8 provides AUV network configurations defined by APD-TDMA and Fig. 12 compares channel utilisation of the existing protocols with a different numbers of node (N) in a network.

APD-TDMA measures channel utilisation only during the transmission phases and does not reveal the certain level of packet loss for the re-initialisation, hence it is difficult to estimate how many times the re-initialisation occurs. Therefore, it is not fair to directly compare APD-TDMA and UW-ALOHA-QM since the channel utilisation of UW-ALOHA-QM is measured from the start of one simulation trial to the end. However, we compare those two protocols when the number of nodes (N) in a network is smaller, on the basis that fewer collisions are likely to occur using a smaller number of nodes. UW-ALOHA-QM shows lower channel utilisation, however it is predicted that, if the channel utilisation of APD-TDMA is measured also together with the multiple initialisation phases, UW-ALOHA-QM may provide better performance than APD-TDMA.

TABLE 9. Thoretical maximum channel utilisation of UW-ALOHA-QM with different N and SM.

Network size (R)	The number of AUVs (N)	Optimum frame size (S_m)	Maximum theoretical channel utilisation (Erlangs)
1500 m	5	2	0.5
1500 m	10	3	0.66
1500 m	15	5	0.60
1500 m	20	6	0.66

The theoretical maximum channel utilisation of UW-ALOHA-QM is calculated by Eq. (5). In the equation, the data packet duration (T_{dp}) and the slot duration (T_s) are constant during simulations of this scenario whilst the number of nodes (N) and the optimum frame size (S_m) are changing. Table 9 provides theoretical channel utilisation of UW-ALOHA-QM in different settings in the AUV network scenario. As the table shows, the number of nodes (N) increases linearly but the optimum frame size (S_m) does not change linearly due to the condition that B is greater than 1.5. Therefore, UW-ALOHA-QM shows the zig-zag style shape in Fig. 12 which is typical feature as explained in [8].

An initialisation phase is required for APD-TDMA and many other protocols to obtain the mobile nodes' location information in the underwater environment and then schedule the data transmissions. However, UW-ALOHA-QM does not need such a phase, because nodes do not need prior information for data transmissions and only the Q-value based on learning experience is important, which is independent from other nodes in the network. Although APD-TDMA knows the location information of AUVs, it becomes invalid quickly because AUVs continue to move. Therefore, the prediction approach of APD-TDMA based on the initialisation or the current data transmission receive timing is only reasonable for constant movements rather than random direction and speed movements. UW-ALOHA-QM, however, does not use prediction but learns and adapts to the changing environment, therefore UW-ALOHA-QM can be used in the network where nodes moves in an unpredictable manner. BTB-TDMA [44] shows the lowest channel utilisation in Fig. 12 because it fundamentally uses the long enough guard time to deal with AUV mobility though it requires time synchronisation.

V. CONCLUSION

In this paper, we have proposed a reinforcement learning based MAC protocol for underwater mobile sensor networks, namely UW-ALOHA-QM. Existing protocols designed for

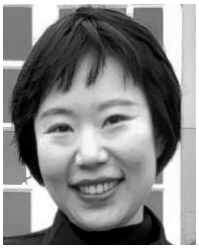
underwater mobile networks handle node mobility through additional supportive measures such as frequent control message exchange rather than our approach to improving network resilience.

Using the reinforcement learning approach, UW-ALOHA-QM provides good channel utilisation and adaptability for a range of mobile network scenarios. In the best scenario, the theoretical maximum channel utilisation of UW-ALOHA-QM reaches 0.66 Erlangs, which is comparable to centralised protocols for underwater networks. An approach has been proposed here for mobile underwater networks and the most appropriate application for this scheme are underwater networks in which node trajectories are unpredictable. Simulation results demonstrate that UW-ALOHA-QM generally outperforms existing protocols under various scenarios and configurations by improving network flexibility. It provides a useful topology agnostic solution to the medium access control problem.

REFERENCES

[1] A. S. Iminova and E. G. Ivanova, "New technologies for ocean cleaning," in *Proc. Conf. Achievements Prospects Innov. Technol.*, Sevastopol, Russia, Apr. 2018, pp. 335–340.
[2] D. Secrieru, G. Oaie, V. Radulescu, and C. Voicaru, "The black sea security system: A new early warning and environmental monitoring system," in *Sustainable Development of Sea-Corridors and Coastal Waters*. Cham, Switzerland: Springer, 2015, pp. 109–115.
[3] P. Braca, R. Goldhahn, G. Ferri, and K. D. LePage, "Distributed information fusion in multistatic sensor networks for underwater surveillance," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4003–4014, Jun. 2016.
[4] J. Lloret, S. Sendra, M. Garcia, and G. Lloret, "Group-based underwater wireless sensor network for marine fish farms," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps)*, Dec. 2011, pp. 115–119.
[5] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
[6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.
[7] S. H. Park, P. D. Mitchell, and D. Grace, "Performance of the ALOHA-Q MAC protocol for underwater acoustic networks," in *Proc. Int. Conf. Comput., Electron. Commun. Eng. (iCCECE)*, Aug. 2018, pp. 189–194.
[8] 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.
[9] J. H. Cui, J. Kong, M. Gerla, and S. Zhou, "The challenges of building mobile underwater wireless networks for aquatic applications," *IEEE Netw.*, vol. 20, no. 3, pp. 12–18, 2006. [Online]. Available: https://ieeexplore.ieee.org/document/1637927
[10] J. Mao, S. Chen, J. Yu, Y. Gu, R. Yu, and Y. Xu, "LTM-MAC: A location-based TDMA MAC protocol for mobile underwater networks," in *Proc. OCEANS*, Apr. 2016, pp. 1–5.
[11] J. Mao, S. Chen, Y. Liu, J. Yu, and Y. Xu, "LT-MAC: A location-based TDMA MAC protocol for small-scale underwater sensor networks," in *Proc. IEEE Int. Conf. Cyber Technol. Automat., Control, Intell. Syst. (CYBER)*, Jun. 2015, pp. 1275–1280.
[12] Y. Noh, U. Lee, S. Han, P. Wang, D. Torres, J. Kim, and M. Gerla, "DOTS: A propagation delay-aware opportunistic MAC protocol for mobile underwater networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 766–782, Apr. 2014.
[13] J. Lee, M. Riess, S. Moser, and F. Slomka, "An adaptive MAC protocol for underwater mobile ad-hoc networks," in *Proc. OCEANS*, May 2018, pp. 1–5.
[14] Y. Zhang, H. Chen, and W. Xu, "A load-adaptive CSMA/CA MAC protocol for mobile underwater acoustic sensor networks," in *Proc. 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2018, pp. 1–7.
[15] M. Gao, W. Li, and J. Li, "Performance analysis of a JSW-based MAC protocol for mobile underwater acoustic networks," in *Proc. 9th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Dec. 2015, pp. 1–6.
[16] A.-R. Cho, C. Yun, Y.-K. Lim, and Y. Choi, "Asymmetric propagation delay-aware TDMA MAC protocol for mobile underwater acoustic sensor networks," *Appl. Sci.*, vol. 8, no. 6, p. 962, Jun. 2018.
[17] N. Morozs, P. Mitchell, and Y. V. Zakharov, "TDA-MAC: TDMA without clock synchronization in underwater acoustic networks," *IEEE Access*, vol. 6, pp. 1091–1108, 2018.
[18] Z. Liu and I. Elhanany, "RL-MAC: A QoS-aware reinforcement learning based MAC protocol for wireless sensor networks," in *Proc. IEEE Int. Conf. Netw., Sens. Control*, Apr. 2016, pp. 768–773.
[19] S. Galzarano, A. Liotta, and G. Fortino, "QL-MAC: A Q-learning based MAC for wireless sensor networks," in *Proc. Conf. Algorithms Archit. Parallel Process.* Cham, Switzerland: Springer, 2013, pp. 267–275.
[20] J. Niu and Z. Deng, "Distributed self-learning scheduling approach for wireless sensor network," *Ad Hoc Netw.*, vol. 11, no. 4, pp. 1276–1286, Jun. 2013.
[21] Y. Chu, S. Kosunalp, P. D. Mitchell, D. Grace, and T. Clarke, "Application of reinforcement learning to medium access control for wireless sensor networks," *Eng. Appl. Artif. Intell.*, vol. 46, pp. 23–32, Nov. 2015.
[22] H. Bayat-Yeganeh, V. Shah-Mansouri, and H. Kebriaei, "A multi-state Q-learning based CSMA MAC protocol for wireless networks," *Wireless Netw.*, vol. 24, no. 4, pp. 1251–1264, May 2018.
[23] G. Chen, Y. Zhan, G. Sheng, L. Xiao, and Y. Wang, "Reinforcement learning-based sensor access control for WBANs," *IEEE Access*, vol. 7, pp. 8483–8494, 2019.
[24] Y. Yu, T. Wang, and S. C. Liew, "Deep-reinforcement learning multiple access for heterogeneous wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1277–1290, Jun. 2019.
[25] N. Javaid, O. A. Karim, A. Sher, M. Imran, A. U. H. Yasar, and M. Guizani, "Q-learning for energy balancing and avoiding the void hole routing protocol in underwater sensor networks," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 702–706.
[26] V. Di Valerio, F. L. Presti, C. Petrioli, L. Picari, D. Spaccini, and S. Basagni, "CARMA: Channel-aware reinforcement learning-based multi-path adaptive routing for underwater wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2634–2647, Nov. 2019.
[27] Z. Jin, Q. Zhao, and Y. Su, "RCAR: A reinforcement-learning-based routing protocol for congestion-avoided underwater acoustic sensor networks," *IEEE Sensors J.*, vol. 19, no. 22, pp. 10881–10891, Nov. 2019.
[28] X. Li, X. Hu, W. Li, and H. Hu, "A multi-agent reinforcement learning routing protocol for underwater optical sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–7.
[29] S. Wang and Y. Shin, "Efficient routing protocol based on reinforcement learning for magnetic induction underwater sensor networks," *IEEE Access*, vol. 7, pp. 82027–82037, 2019.
[30] L. Wang, C. Lin, K. Chen, and Y. Zhang, "A learning-based ALOHA protocol for underwater acoustic sensor networks," in *Proc. 13th ACM Int. Conf. Underwater Netw. Syst.*, Dec. 2018, pp. 1–2.
[31] X. Geng and Y. R. Zheng, "MAC protocol for underwater acoustic networks based on deep reinforcement learning," in *Proc. Int. Conf. Underwater Netw. Syst.*, Oct. 2019, pp. 1–5.
[32] X. Ye and L. Fu, "Deep reinforcement learning based MAC protocol for underwater acoustic networks," in *Proc. Int. Conf. Underwater Netw. Syst.*, Oct. 2019, pp. 1–5.
[33] P. Smith *et al.*, "Network resilience: A systematic approach," *IEEE Commun. Mag.*, vol. 49, no. 7, pp. 88–97, Jul. 2011.
[34] K.-L.-A. Yau, H. G. Goh, D. Chieng, and K. H. Kwong, "Application of reinforcement learning to wireless sensor networks: Models and algorithms," *Computing*, vol. 97, no. 11, pp. 1045–1075, Nov. 2015.
[35] S. Kosunalp, P. D. Mitchell, D. Grace, and T. Clarke, "Practical implementation and stability analysis of ALOHA-Q for wireless sensor networks," *ETRI J.*, vol. 38, no. 5, pp. 911–921, Oct. 2016.
[36] A. Caruso, F. Paparella, L. F. M. Vieira, M. Erol, and M. Gerla, "The meandering current mobility model and its impact on underwater mobile sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2018, pp. 221–225.
[37] M. Purcell, D. Gallo, A. Sherrell, M. Rothenbeck, and S. Pascaud, "Use of REMUS 6000 AUVs in the search for the air France flight 447," in *Proc. Oceans*, 2011, pp. 1–7.
[38] N. Chirdchoo, W. S. Soh, and K. C. Chua, "MU-Sync: A time synchronization protocol for underwater mobile networks," in *Proc. ACM Workshop Underwater Netw.*, 2018, pp. 35–42.
[39] F. Lu, D. Mirza, and C. Schurgers, "D-sync: Doppler-based time synchronization for mobile underwater sensor networks," in *Proc. ACM Workshop Underwater Netw.*, 2010, pp. 3–9.

[40] J. Liu, Z. Zhou, Z. Peng, J.-H. Cui, M. Zuba, and L. Fiondella, "Mobi-sync: Efficient time synchronization for mobile underwater sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 2, pp. 406–416, Feb. 2013.
[41] *EvoLogics*. Accessed: Apr. 13, 2020. [Online]. Available: https://evologics.de/acoustic-modem/18-34
[42] M. Lewis, S. P. Neill, P. Robins, M. R. Hashemi, and S. Ward, "Characteristics of the velocity profile at tidal-stream energy sites," *Renew. Energy*, vol. 114, pp. 258–272, Dec. 2017.
[43] H. Shin, Y. Kim, S. Baek, and Y. Song, "Distributed learning for dynamic channel access in underwater sensor networks," *Entropy*, vol. 22, no. 9, p. 992, Sep. 2020.
[44] A.-R. Cho, C. Yun, J.-W. Park, and Y.-K. Lim, "Design of a block time bounded TDMA (BTB-TDMA) MAC protocol for UANets," *Ocean Eng.*, vol. 38, nos. 17–18, pp. 2215–2226, Dec. 2011.



SUNG HYUN PARK (Graduate Student Member, IEEE) was born in Seoul, South Korea, in 1982. She received the B.S. and M.S. degrees in electronic engineering from Kwangwoon University, Seoul, in 2005 and 2007, respectively, and the M.B.A. degree (Hons.) from Aston University, Birmingham, U.K., in 2014. She is currently pursuing the Ph.D. degree with the Communication Technologies Research Group, Department of Electronic Engineering, University of York. She was invited to give an undergraduate lecture at Kwangwoon University. From 2008 to 2013, she became a Senior Researcher with Innowireless, a manufacturer of mobile technology test solutions, South Korea. Her expertise involved 4G technologies, such as WiMax, LTE, and LTE-A. She worked with the world's major mobile operators. From 2014 to 2016, she went on to work at Qucell, London, U.K., an LTE small cell provider. She was a Technical Support Manager and collaborated with many major British mobile service providers, including BT, EE, and Talk Talk. In 2016, she was awarded a research scholarship from the Department of Electronic Engineering, University of York, to undertake the Ph.D. degree. Her study is supervised by Prof. Paul Daniel Mitchell and Prof. David Grace. She is interested in the design and testing of novel intelligent medium access control (MAC) protocols, specifically in the use of reinforcement learning for underwater acoustic communication networks. In 2018, she received the Best Paper Award at the IEEE iCCECE annual International Conference.

PAUL DANIEL MITCHELL (Senior Member, IEEE) received the M.Eng. and Ph.D. degrees from the University of York, York, U.K., in 1999 and 2003, respectively. His Ph.D. research was on medium access control for satellite systems, which was supported by British Telecom. He has gained industrial experience with BT and QinetiQ. Since 2002, he has been a member of the Department of Electronic Engineering, University of York, where he is currently a Professor. He has authored more than 110 refereed journal and conference papers. His research interests include medium access control and routing, underwater acoustic networks, wireless sensor networks, cognitive radio, traffic modeling, queuing theory, and satellite and mobile communication systems. He is a member of the IET and a Fellow of the Higher Education Academy. He has served on numerous international conference program committees. He was the General Chair of the International Symposium on Wireless Communications Systems, which was held in York, in 2010, the Track Chair of the IEEE VTC, in 2014, and the TPC Co-Chair of the ISWCS 2019. He is an Associate Editor of the *IET Wireless Sensor Systems* journal and the *International Journal of Distributed Sensor Networks* (SAGE).

DAVID GRACE (Senior Member, IEEE) received the Ph.D. degree from the University of York, in 1999, with the subject of his dissertation on Distributed Dynamic Channel Assignment for the Wireless Environment. In 2000, he jointly founded SkyLARC Technologies Ltd. and was one of its directors. He was one of the lead investigators of FP7 ABSOLUTE and focused on extending LTE-A for emergency/temporary events through the application of cognitive techniques. He was the Technical Lead of the 14-Partner FP6 CAPANINA Project that dealt with broadband communications from high-altitude platforms. Since 1994, he has been a member of the Department of Electronic Engineering, University of York, where he is currently a Professor (Research) and the Head of the Communication Technologies Research Group. He is also the Co-Director of the York-Zhejiang Lab on Cognitive Radio and Green Communications and a Guest Professor with Zhejiang University. He is the Lead Investigator of H2020 MCSA 5G-AURA and H2020 MCSA SPOTLIGHT. He has authored over 220 articles and authored/edited two books. His current research interests include aerial platform-based communications; cognitive green radio, particularly applying distributed artificial intelligence to resource and topology management to improve overall energy efficiency; 5G system architectures; dynamic spectrum access; and interference management. He was the Chair of the IEEE Technical Committee on Cognitive Networks, from 2013 to 2014. He is a Founding Member of the IEEE Technical Committee on Green Communications and Computing.