UNIVERSITY of York

This is a repository copy of *Reinforcement learning based MAC protocol (UW-ALOHA-Q) for underwater acoustic sensor networks*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/153604/

Version: Accepted Version

Article:
Park, Sung Hyun, Mitchell, Paul Daniel orcid.org/0000-0003-0714-2581 and Grace, David orcid.org/0000-0003-4493-7498 (2019) Reinforcement learning based MAC protocol (UW-ALOHA-Q) for underwater acoustic sensor networks. IEEE Access. ISSN 2169-3536

https://doi.org/10.1109/ACCESS.2019.2953801

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Reinforcement Learning based MAC Protocol (UW-ALOHA-Q) for Underwater Acoustic Sensor Networks

SUNG HYUN PARK[1], PAUL DANIEL MITCHELL[1], (Senior Member, IEEE), DAVID GRACE[1], (Senior Member, IEEE)

[1]Department of Electronic Engineering, University of York, York YO10 5DD, U.K.

Corresponding author: Sung H. Park (e-mail: sp1356@york.ac.uk).

This work was supported in part by the U.K. Engineering and Physical Science Research Council through the USMART project under Grant EP/P017975/1.

ABSTRACT The demand for regular monitoring of the marine environment and ocean exploration is rapidly increasing, yet the limited bandwidth and slow propagation speed of acoustic signals leads to low data throughput for underwater networks used for such purposes. This study describes a novel approach to medium access control that engenders efficient use of an acoustic channel. ALOHA-Q is a medium access protocol designed for terrestrial radio sensor networks and reinforcement learning is incorporated into the protocol to provide efficient channel access. In principle, it potentially offers opportunities for underwater network design, due to its adaptive capability and its responsiveness to environmental changes. However, preliminary work has shown that the achievable channel utilisation is much lower in underwater environments compared with the terrestrial environment. Three improvements are proposed in this paper to address key limitations and establish a new protocol (UW-ALOHA-Q). The new protocol includes asynchronous operation to eliminate the challenges associated with time synchronisation under water, offer an increase in channel utilisation through a reduction in the number of slots per frame, and achieve collision free scheduling by incorporating a new random back-off scheme. Simulations demonstrate that UW-ALOHA-Q provides considerable benefits in terms of achievable channel utilisation, particularly when used in large scale distributed networks.

INDEX TERMS MAC Protocol, Medium Access Control, Reinforcement Learning, Underwater Acoustic Networks

I. INTRODUCTION

The Earth's surface comprises 71% water [1] and the market value of coastal resources is estimated to be 3 trillion USD per year [2], with our oceans contributing 1.5 trillion USD annually in value-added to the global economy [3]. It is therefore unsurprising that the marine environment is central to a vast diversity of industries and areas of scientific importance. Examples of underwater applications include disaster detection far off coast, underwater security surveillance, as well as environmental and ecosystem data gathering. However, most of the ocean has not been explored since ocean exploration is significantly hampered by the inherently hostile and harsh environment for both people and equipment. To deal with the challenges of the underwater environment, wire free communication is necessary in order to monitor the oceans more effectively, remotely, and potentially in real time.

Wireless Sensor Networks (WSNs) using radio technology are used for monitoring purposes in many applications in the terrestrial environment. However, this technology cannot be directly applied to the underwater environment since radio signals are heavily absorbed by water. Acoustic signals are the most viable means of communicating underwater, but technologies for underwater acoustic communications are complex and demand sophisticated signal processing, hence underwater devices tend to be bulky and expensive [4]. Moreover, the slower propagation speed ($\approx$ 1500 m/s) of acoustic signals in water compared to radio signals in the air ($\approx 3 \times 10^8$ m/s) leads to poor channel utilisation in underwater networks, and the limited and distance dependent

bandwidth brings about low fundamental capacity based on Shannon's channel capacity theory [5].

To address these problems limiting the efficient use of acoustic networks for underwater monitoring, we describe a novel reinforcement learning based Medium Access Control (MAC) protocol, UW-ALOHA-Q. The merits of UW-ALOHA-Q lie in providing a low complexity approach through reinforcement learning to achieve high channel utilisation in distributed networks where centralized scheduling is not feasible and distributed scheduling introduces significant signalling overheads and complexity.

ALOHA-Q was designed for WSNs in the terrestrial environment and uses reinforcement learning as a technique whereby nodes learn through trial-and-error interactions with the environment [6]. The underwater environment continuously changes and hence underwater networks need to be capable of adapting to such time varying changes. Reinforcement learning based protocols are able to inherently adapt to these environmental changes through the learning process. Therefore, the objective of this study is to transform the design of an established reinforcement learning based protocol (ALOHA-Q) into one suitable for the underwater environment (UW-ALOHA-Q).

Specific contributions of this paper include:

- Transformation of ALOHA-Q (developed for terrestrial networks) to a new protocol for underwater acoustic networks (UW-ALOHA-Q) through three improvements: asynchronous operation; optimisation of the number of slots in a frame; incorporation of a new back-off scheme.
- Design of the new protocol for asynchronous and self-organised distributed underwater networks, achieving collision free scheduling and high channel utilisation alongside low overheads.
- Investigation of the baseline channel utilisation of the new protocol for different network sizes and topologies through a simulation.

A preliminary paper was presented at the IEEE International Conference on Computing, Electronics and Communication Engineering (iCCECE' 2018) which received a best paper award [7].

Section II of this paper provides a summary of the related literature. Section III details the ALOHA-Q protocol and provides a summary of the preliminary paper [7]. Section IV describes the transformation processes underpinning the development of UW-ALOHA-Q from ALOHA-Q. Section V presents simulation results showing key performance characteristics of UW-ALOHA-Q under various network configurations.

II. PREVIOUS WORK

The MAC layer is responsible for organising the access of each node to their shared transmission medium. The general objective of the MAC layer is to minimise collisions and overheads in the channel through a suitable protocol. The operation of the MAC layer also has an impact on achievable Quality of Service (QoS) including latency, energy efficiency, network scalability, and adaptability. Therefore, the MAC layer can play a key role in underwater acoustic networks in maximising channel utilisation, both in the presence of a limited bandwidth and slow propagation speed.

MAC protocols can be generally categorised as centralised or distributed. Centralised MAC protocols can achieve good channel utilisation through collision-free scheduling, but require infrastructure to provide a coordinating node and time synchronisation. Typical examples of centralised MAC protocols include Time Division Multiple Access (TDMA) and polling based protocols. Distributed MAC protocols do not require such infrastructure, however, significant additional overheads are incurred for distributed scheduling, or to otherwise incorporate techniques such as handshaking or carrier sensing whenever a sender initiates transmission in order to help reduce the probability of collision. Examples of these include Carrier Sense Multiple Access (CSMA) [8] and Multiple Access with Collision Avoidance (MACA) [9].

Recently, reinforcement learning schemes have been applied to MAC protocols in WSNs for terrestrial networks and the results are promising [10-16]. ALOHA-Q [13] is a reinforcement learning based protocol designed to be used in Low Rate - Personal Area Networks (LR-PANs). The protocol is based on framed slotted ALOHA [17] which is a distributed protocol employing time synchronisation to reduce data packet collisions. Due to its low complexity and lack of infrastructure requirements, framed slotted ALOHA is used as a fundamental system for many different types of network. For example, it is a primary protocol in Radio Frequency Identification (RFID) tag systems [18] and has also been considered for use in Machine to Machine (M2M) networks [19].

In framed slotted ALOHA, all nodes are synchronised into time frames and slots across the network. Each node must deliver a data packet within a defined slot period. Since there is no means of coordinating the times in which data packets are transmitted by nodes, collisions occur regularly leading to an unreliable service. ALOHA-Q takes the advantages of framed slotted ALOHA which are simplicity and low overheads. However, ALOHA-Q avoids collisions through a reinforcement learning process as nodes in the network can determine which slots to transmit in to avoid collisions. As a consequence, the ALOHA-Q protocol approaches centralised style scheduling without the need for any form of central controller and achieves a nearly identical level of channel utilisation [13] as that of a centralised scheme in steady-state conditions. ALOHA-Q is discussed further in section III.

While reinforcement learning based MAC protocols have been researched extensively for terrestrial networks, there has, however, been very little research into underwater reinforcement learning based protocols. Most of these are for routing [20-24] and only one protocol has been found for the MAC layer [25] which uses a reinforcement learning approach

to extend the lifetime of underwater acoustic wireless sensor networks. The study was proposed in 2013 and the aim of the proposed protocol is to extend the lifetime of a network. It is a distributed protocol based on slotted CSMA with time synchronisation. Nodes learn optimal decisions for three aspects of the next data packet transmission: the next relay node, the sub-channel to sense, and the level of transmission power to use. The protocol requires periodic control message exchange for neighbour discovery which can lead to high overheads and thereby a decrease in channel utilisation due to the slow propagation speed. In addition, multi-channel communication is used in the design, which is not optimal for underwater acoustic networks since the channel bandwidth is so limited, especially over longer distances. Moreover, the protocol uses carrier sensing and exponential random back-off which can deteriorate channel utilisation. Carrier sensing, in particular, potentially requires long guard bands due to the long propagation delay, otherwise it is ineffective underwater.

III. ALOHA-Q

ALOHA-Q is a reinforcement based MAC protocol designed for WSNs in the terrestrial environment. All nodes in an ALOHA-Q network are time synchronized. Table I gives typical parameters related to the slot and frame structures of ALOHA-Q as used in the terrestrial environment [13] and Fig. 1 illustrates an example of a packet flow between a generating node and a sink.

TABLE I
TYPICAL ALOHA-Q PARAMETERS FOR TERRESTRIAL USE

Parameter	Value
Duration of a data packet of 1044 bits (T_{dp})	4.176 ms
Duration of an acknowledgement packet of 20 bits (T_{ap})	0.08 ms
Duration of a guard time of 36 bits (T_g)	0.144 ms
Duration of a slot (T_s)	4.4 ms
Distance between a generating node and a sink node	12.9 m
Tx/Rx data rate (r_{tr})	250,000 bps
The number of generating nodes (N)	50 nodes
Propagation speed (v_{tr})	$3 \cdot 10^8$ m/s
Propagation delay (τ_p)	negligible

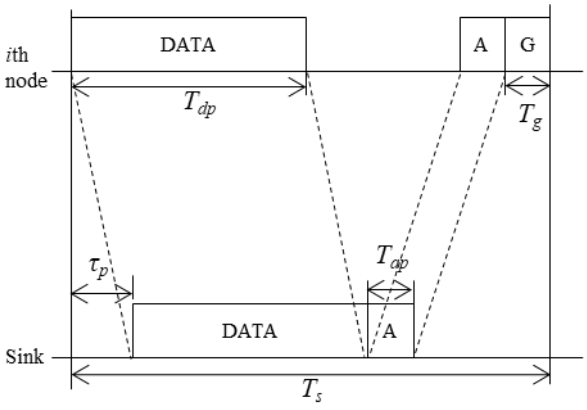

FIGURE 1. Packet flow between a generating node and a sink node

For the terrestrial environment, the propagation speed of 3×10^8 m/s is used for the radio signals and a 250,000 bps data rate is used reflecting IEEE 802.15.4 LR-WPANs [26]. One slot is sufficient to accommodate a data packet, an acknowledgement packet, and a guard time. After sending the data packet, if the generating node does not receive an acknowledgement from the sink node before the guard time ends (i.e. a stop and wait acknowledgement policy), the transmission is assumed to have failed and a retransmission must be initiated.

A. REINFORCEMENT LEARNING

Reinforcement learning enables agents to learn an optimal action through trial-and-error interactions in a dynamic environment, with future actions determined by prior experience [9]. This established artificial intelligence strategy has recently been applied to MAC layer protocols for terrestrial networks and shows promising results [10-16]. Stateless Q-learning [27] is used in the ALOHA-Q protocol, in which each node uses the Q-learning scheme to select one slot in a frame to send one data packet at the start of each frame. All nodes have their own Q-table which contains individual Q-values for each slot in a frame. Equation (1) is used to determine how Q-values are updated:

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

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, a is the learning rate, and r is the reward. A standard implementation of ALOHA-Q uses $a = 0.1$ and $r = 1$ if the transmission is successful, otherwise, $r = -1$.

Fig. 2 illustrates a simple example of how the Q-values of each frame in the Q-table might become updated. Since all Q-values in the Q-table are initially zero in this example, a node randomly selects a slot in the next frame for data packet transmission. If the node receives a positive acknowledgement before the guard time ends, meaning the transmission was successful, the Q-value for the first slot in the Q-table becomes updated to 0.1 as shown through the application of (1). Thus, after one frame, the Q-table has Q-values of 0.1/ 0/ 0/ 0 and the first slot has the highest Q-value in the node's Q-table.

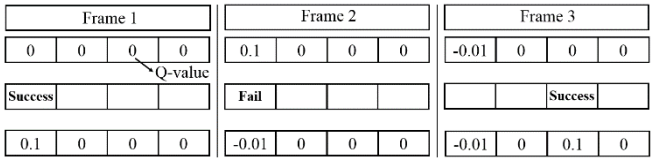

FIGURE 2. An example of Q-table in a single node when one frame comprises four slots

At the start of the second frame, the node transmits a data packet in the first 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 acknowledgement packet before the guard time ends, the node assumes that the transmission has failed and the Q-value for the first 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 selects a slot number randomly since the 2nd, 3rd, and 4th slots 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 the same network.

Importantly, each node operates independently of each other as each node only refers to its own Q-table to determine the transmission order in a frame. ALOHA-Q does not need any periodic message exchange for neighbor discovery nor any control message exchange for scheduling. These characteristics of low overheads and high simplicity are highly significant and unique to ALOHA-Q because existing distributed protocols require each node to have information about its neighbors or to reserve a channel before every transmission to avoid collisions.

B. LIMITATIONS OF ALOHA-Q FOR UNDERWATER ACOUSTIC NETWORKS

It is expected that a reinforcement learning based protocol can offer underwater networks the capability of adapting through constantly interacting with the time-varying underwater conditions. Therefore, it is of interest to explore the possibility that ALOHA-Q can be used for underwater networks. An initial simulation based study has been undertaken in [7], comprising 50 generating nodes in a single-hop ring topology with one sink node located centrally. All nodes are considered to be within interfering range. The packet inter-arrival time is exponentially distributed and a collision-based error model is used for reception in the simulation. The purpose of the initial simulation is to compare the performance of ALOHA-Q in both terrestrial and underwater environments. Table II shows the simulation parameters used for ALOHA-Q in the underwater environment. The same simulation parameters for the previous study [13] are used in this section and only two notable parameters for the underwater network have been changed for fair comparison: the propagation speed of 1500 m/s is used for acoustic signals under water and the use of a state of the art underwater modem which is currently on the market with a data rate of 62,500 bps [28] is considered.

TABLE II
TYPICAL ALOHA-Q PARAMETERS FOR UNDERWATER USE

Parameter	Value
Duration of a data packet of 1044 bits (T_{dp})	16.704 ms
Duration of an acknowledgement packet of 20 bits (T_{ap})	0.32 ms
Duration of a guard time of 36 bits (T_g)	0.576 ms
Duration of a slot (T_s)	34.8 ms
Distance between a generating node and a sink node	12.9 m
Tx/Rx data rate (r_{uw})	62,500 bps
The number of nodes (N)	50 nodes
Propagation speed (v_{uw})	1500 m/s
Propagation delay (τ_p)	8.6 ms

Not all parameters are realistic for a practical underwater deployment, but it is important to keep the network topology parameters unchanged for the comparison to be useful. Beyond this initial comparison, realistic parameters are used for underwater network simulations in section V.

The result of this simulation shows that ALOHA-Q can be operated in the underwater environment but that the protocol only achieves a channel utilisation of 0.48 Erlangs, much lower than the 0.95 Erlangs achieved by the same protocol within a terrestrial environment [7]. The unit of Erlang corresponds to the fractional proportion of time during which data traffic is usefully received. 1 Erlang therefore corresponds to the fundamental capacity of the channel. The slow propagation speed of acoustic signals is the primary cause for low channel utilization. Equation (2) shows the calculation for the duration of a slot (T_s) which is proportional to the propagation delay (τ_p). During the propagation of the data and acknowledgement packets, the channel remains in an idle state which consequently causes a decrease in achievable channel utilisation.

$$T_s = (T_{dp} + T_{ap} + T_g) + 2 \times \tau_p \quad (2)$$

Therefore, conclusions from the initial simulations [7] show that although ALOHA-Q can be operated in an underwater environment, it is constrained by low channel utilisation due to the slow propagation speed of acoustic signal underwater.

IV. UW-ALOHA-Q

To transform ALOHA-Q for the underwater environment, we consider three improvements to the protocol: asynchronous operation, optimisation of the number of slots per frame, and a new random back-off scheme. Each improvement is discussed in this section.

A. ASYNCHRONOUS OPERATION

Generally, terrestrial networks can be time synchronised based on use of a global time reference, thereby reducing the probability of collision in contention based schemes by shortening the vulnerable period. For example, ALOHA-Q also uses time synchronisation in the terrestrial environment and achieves collision free scheduling through reinforcement learning, but for the same topology and parameters, it shows a decrease in channel utilisation without time synchronisation from 0.95 Erlangs to 0.64 Erlangs [13]. However, the reliance on time synchronisation in the underwater environment is costly and complex since GPS is not available [29]. Consequently, as a first step we consider asynchronous implementation of ALOHA-Q for underwater networks. It would be expected that collisions will occur in the absence of time synchronisation, since transmissions from nodes will arrive at a receiver at random times. However, utilising the idle time caused by the propagation delay (τ_p), reinforcement learning can still achieve collision free reception in the same way as described in section III in the underwater environment. Fig. 3 compares the difference in reception patterns of data packets at a sink node with ALOHA-Q in the two different

environments. In the terrestrial environment, packet receptions are time synchronised and the propagation delay is negligible, so that the data packets from different generating nodes arrive close to each other at the sink and only small guard bands are required. Channel utilisation is high under this condition; however, if asynchronous operation is applied, a significant number of collisions occur because data packets will then overlap with each other at the receiver due to the short slot duration. In the underwater environment, however, the length of a slot needs to be much greater for stop and wait ALOHA-Q, to accommodate the long propagation delays. The long propagation delay results in a long idle time at the sink node such that the channel utilisation becomes lower, but the idle time tends to be sufficient to avoid overlapping reception, so the protocol is less prone to experiencing collisions.

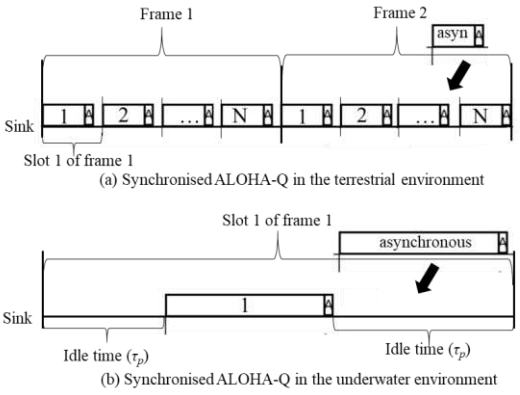

FIGURE 3. Reception of data packets at a sink node in the terrestrial and underwater environments.

Even if packets overlap at the receiver, reinforcement learning can achieve collision free operation using the idle time without relying on synchronisation in the underwater environment as shown in Fig. 4.

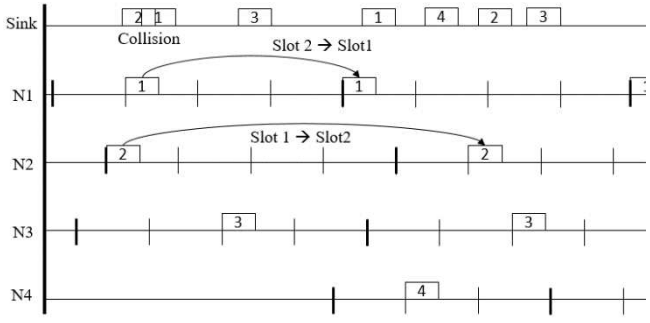

FIGURE 4. How reinforcement learning removes collisions in the underwater environment in the absence of time synchronisation (The acknowledgment processes are omitted in Figs. 4, 5 and 6 for the purpose of simplicity.)

The four nodes (N1~N4) have to choose a slot number from slot1, slot2, slot3 or slot4 for their data packet transmissions in each frame. The nodes are not synchronised, so the frame start time for each node is different. In the first frame, N1 randomly chooses slot2 and transmits a data packet in the slot, N2 in slot1, N3 in slot3, and N4 in slot2. At the sink node, packets from N1 and N2 overlap with each other and collide in the first frame transmission process. Therefore, the two nodes do not receive acknowledgements from the sink node. As a result, the Q-values of the slots in the Q-table are negatively reinforced so the two nodes change slot numbers for the next transmission: N1 chooses slot1 and N2 chooses slot2. The new order no longer results in overlapping data packets at the receiver from N1 and N2. Whereas N3 and N4 continue to use the same slot numbers they used for their first transmissions since they successfully received acknowledgements.

By repeating the learning cycle, the four nodes can learn which slot number they need to use and finally all four packets can arrive at the sink node without interfering with reception from other nodes in the network: this status is called convergence. Convergence only applies in a relatively static environment. In practical underwater scenarios, what is required is effective adaptation of transmission timing in response to changing conditions to retain higher utilisation than can be achieved without reinforcement learning. Though, the scope of this paper is understanding the baseline capability and fundamental behavior of UW-ALOHA-Q, convergence and conditions where network convergence are considered as discussed in section V.

The slots allow collisions to be avoided despite the absence of time synchronisation through reinforcement learning due to the long propagation delay (τ_p) and consequently long slot duration (T_s). However, despite the reduction in collisions, the achievable channel utilisation remains low.

B. OPTIMISATION OF THE NUMBER OF SLOTS

Building on the benefits of asynchronous operation, it then becomes feasible to explore the possibility of increasing channel utilisation by reducing the number of slots per frame. This concept is depicted in Fig. 5 which shows an example of how collision free reception can be obtained when only two slots are used to support four generating nodes in a frame.

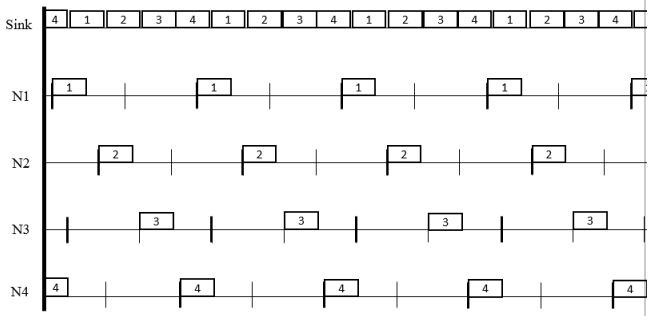

FIGURE 5. Reduced number of slots per frame and increased channel utilisation

By comparing Figs. 4 and 5, it is clear that channel utilisation can be improved simply by reducing the number of slots in a frame. In time synchronised networks such as ALOHA-Q, if a smaller number of slots is used than the number of interfering nodes, collisions occur since all transmitting nodes cannot

obtain a dedicated slot to send their packets reliably. However, in the absence of time synchronisation, reducing the number of slots is feasible since there is space to accommodate multiple packets within a single frame in the underwater environment, due to the long propagation delay and given different frame start times.

C. UNIFORM RANDOM BACK-OFF SCHEME

Incorporation of the first two improvements provides the potential for high channel utilisation to be achieved underwater. However, using a reduced number of slots per frame, a possibility arises that the network cannot converge due to the randomly inherited frame start time which cannot be changed. A new time-based random back-off scheme is proposed to address this problem and allow convergence to be achieved.

Traditionally, in wireless communication networks, when a transmission fails, a node does not send the retransmission immediately, but delays it in order to avoid a potential collision. This delay is called back-off and the delayed time is often calculated as a number of slots. As an example, the back-off algorithm in the IEEE 802.11 Wireless Local Area Networks (WLANs) standard [30] delays retransmissions based on the number of slots in a contention window with an exponential increase in the window size in response to successive failures.

However, if the same slot based strategy is applied to ALOHA-Q with the two proposed improvements in the underwater environment, the possibility of non-convergence continues to exist since some nodes cannot find a distinct slot from the reduced number of slots per frame having the fixed frame start time. Therefore, we propose a new back-off scheme called uniform random back-off. This scheme operates independently from the slot learning process (described in section A) and provides a chance for nodes to adapt their frame start times. Using this scheme, for every collision, nodes randomly delay the next frame start time according to a uniform distribution. By repeated trial-and-error learning, all nodes can discover an appropriate frame start time and slot to use in successive frames. Operation of the proposed uniform random back-off scheme is illustrated in Fig. 6 in which one slot is used in a frame for two generating nodes in the network.

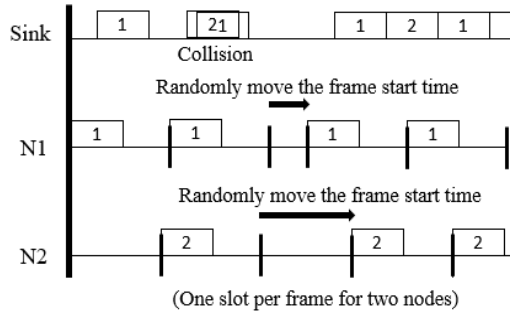

FIGURE 6. Uniform random back-off scheme for UW-ALOHA-Q when one slot per frame is used for two generating nodes

Inclusion of this scheme leads to collision free scheduling and permits convergence in UW-ALOHA-Q underwater acoustic networks under the assumption that any environmental changes are covered by the guard duration (T_g). Therefore, an appropriate guard duration needs to be chosen for a particular environment to accommodate for changes in propagation delay arising from node movement in the water.

In summary, the proposed UW-ALOHA-Q scheme can achieve high channel utilisation with low costs and overheads without the need of time synchronisation and any centralised controller in the underwater environment. The following simulations demonstrate the behaviour of UW-ALOHA-Q with different network configurations and serve to validate the envisaged channel utilisation capacity of the protocol.

V. SIMULATIONS

Simulations have been carried out to understand the baseline channel utilisation of UW-ALOHA-Q. Identical configurations and parameters to those described earlier in section III. B are used and simulations are carried out for different network topologies, comprising 25 and 50 nodes, as well as with propagation distances varying from 100 m to 1000 m.

A. PARAMETERS AND PERFORMANCE MEASURE

Channel utilisation (U) is evaluated as fractional amount of time in which data traffic is successfully received at the sink node and is calculated by (3),

$$U = \frac{R \times T_{dp}}{Mesaurement\ duration} \tag{3}$$

where, R is the number of data packets successfully received at the sink node over the period of interest which is the measurement duration from network convergence frame to the end of a simulation.

We define two parameters for simulation analysis:

- *Scvg*: the number of slots per frame which can permit convergence to be achieved for a certain size of a network
- Index B: the ratio between 'the duration of a single frame excluding acknowledgement packets and guard times' and 'the total duration of data packets in a frame generated from all nodes in a network'

As described in (4), this ratio represents the overheads of a system to total capacity of frame related to the data carrying capacity of the frame:

$$B = \frac{S \times (2 \times \tau_p + T_{dp})}{N \times T_{dp}} \tag{4}$$

where, S is the number of slots per frame. The potential range of S considered in this paper is $0 < S \leq N$.

B. THE TRADEOFF BETWEEN CHANNEL UTILISATION AND CONVERGENCE AS A FUNCTION OF THE NUMBER OF SLOTS PER FRAME

The number of slots per frame is a key parameter of UW-ALOHA-Q since the selection of the number of slots significantly impacts upon the achievable channel utilisation and the end to end delay performance of UW-ALOHA-Q networks. An excessive number of slots in the frame will lead to poor channel utilisation, whereas insufficient slots in a frame will not provide a sufficient duration for the transmitting nodes to find collision free space. Therefore, this section provides simulation results of channel utilisation according to the number of slots per frame and highlights a tradeoff between channel utilisation and the probability of convergence.

Table III shows the simulated channel utilisation of UW-ALOHA-Q when the number of slots per frame varies, for a network comprising 25 generating nodes equally spaced around a 100 m radius ring topology with a central receiver. The simulations include the first two improvements and exclude the uniform random back-off scheme in order to particularly understand the impact of changing the number of slots per frame. For each value of the number of slots per frame, 100 simulations are carried out and one simulation comprises 5000 frames to ensure sufficient time to converge. Convergence is considered to have occurred when all generating nodes send packets using the same distinct slot for 800 consecutive frames.

TABLE III
TRADE-OFF BETWEEN CHANNEL UTILISATION AND THE CHANCE OF CONVERGENCE ACCORDING TO THE NUMBER OF SLOTS PER FRAME

Number of slots in a frame (S)	Index ratio B	The number of simulations where the network converges	Average channel utilisation (Erlangs)
4	1.44	1	0.44
5	1.80	28	0.46
6	2.16	63	0.42
7	2.51	80	0.36
8	2.87	97	0.34

During the simulations, each of the 25 nodes uses reinforcement learning to find a distinct slot in a frame which does not interfere with the transmission of any of its neighbors. Increasing the number of slots up to 8 per frame increases the flexibility in the selection of any particular slot and it is, therefore easier for the network to converge through the learning process of each node, despite a relatively low channel utilisation of 0.34 Erlangs. However, as shown in the results, a trade-off is observed when the number of slots is lowered from 8 to 5, with the highest average of channel utilisation is achieved at 0.46 Erlangs but with convergence occurring less frequently: the UW-ALOHA-Q network converges 28 times out of 100 simulation trials. Therefore, it is observed that UW-ALOHA-Q shows a trade-off between average channel utilisation and the chance of convergence as the number of slots varies.

As stated earlier, these simulations do not include the new back off scheme. As shown in Table III, the network fails to converge on 3 occasions out of 100 trials when 8 slots per frame is used. This low probability of convergence failure can be overcome by the uniform random back-off scheme by finding the appropriate frame start time, and thereby allowing the UW-ALOHA-Q protocol to converge every time.

Table IV compares simulation results with and without the uniform random back-off scheme. Applying the scheme, nodes which cannot find a distinct slot are able to adjust their frame start time. Consequently, all nodes can find an appropriate frame start time and a distinct slot so that simulation results shows that the network converges 100 times out of 100 trials. However, during this process, the scheme disturbs nodes which already find their own distinct slot and thus triggers additional learning processes. Therefore, overall network convergence takes more frames (i.e. more trial-and-error learning processes) than UW-ALOHA-Q without the back-off scheme.

TABLE IV
SIMULATION RESULTS WHEN UNIFORM RANDOM BACK-OFF IS USED

Number of slots in a frame (S)	Uniform random back-off scheme	The number of simulations where the network converges	Average channel utilisation (Erlangs)	Average number of frames used for network converge (frames)
8	Not used	97	0.34	20.04
8	Used	100	0.35	158.51

Simulations have also been carried out for different sizes of networks, using 25 and 50 nodes, as well as with different propagation distances varying from 100 m to 1000 m. An identical tradeoff is observed for all variables under a condition that the index ratio (B) is greater than 1.5. This also implies that the highest average channel utilisation of UW-ALOHA-Q is achievable under a condition of the index ratio equal to 1.5. However, this paper focuses on validating the baseline channel utilisation of UW-ALOHAQ, therefore, simulation results of this paper demonstrate UW-ALOHA-Q in case when the network reliably converges, rather than when the highest average channel utilisation is achieved.

C. CHANNEL UTILISATION AS A FUNCTION OF NETWORK SIZE

In terms of network deployment, the size of a network and the number of nodes in the network are determined by the requirements of individual applications. Therefore, it is necessary to predict the channel utilisation of UW-ALOHA-Q across a range of different size networks in order to define the baseline performance which UW-ALOHA-Q can provide for a range of different applications. Fig. 7 illustrates the simulated channel utilisation of UW-ALOHA-Q following convergence in a ring topology where the network size varies

from 100 m to 1000 m radius with 25 nodes. Identical configurations to those in the earlier section B are used for the simulations, but the uniform random back-off scheme is applied for network convergence.

These results present the detailed UW-ALOHA-Q behavior based on the index ratio (*B*). The main observation is that network convergence is achievable when the index ratio (*B*) is greater than 2.6 as Fig. 7 specifies. The number of slots per frame for network convergence (*Scvg*) varies from 1 to 8 as the network size decreases. In the larger networks, such as those with a 900 m and 1000 m radius, the propagation delay primarily accounts for one slot as referred to by (2). During the propagation delay, the channel is idle and the amount of idle time in one slot is sufficient for 25 nodes to find a distinct time period for transmission. Therefore, the network can converge and achieve collision free scheduling when the number of slots per frame is 1. However, the amount of available time in one slot for 25 nodes in an 800 m network is insufficient, therefore, adding one more slot in a frame is necessary so that the network achieves convergence when the number of slots per frame equals 2. Adding one more slot in a frame, however, causes a decrease in channel utilisation due to redundant idle time. We term this change in channel utilisation as 'the effect of a slot'.

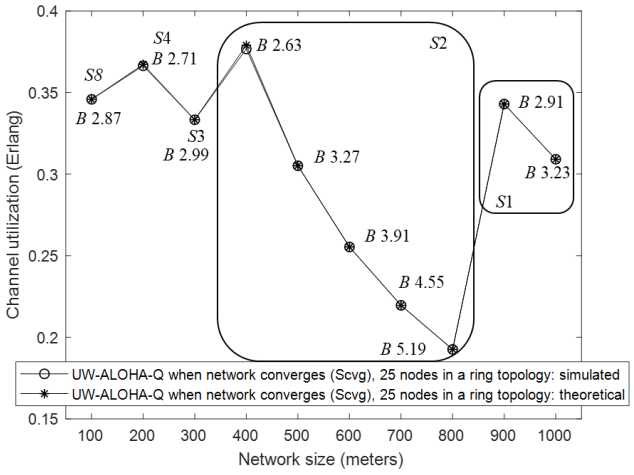

FIGURE 7. Channel utilisation of UW-ALOHA-Q networks for a 25 node ring topology at a variable network sizes when network converges (Scvg)

Once a network has converged, all nodes use the same number of slots and timing in a frame. Therefore, a centralised data transmission pattern is formed and this pattern is repeated as long as convergence is maintained. Based on this, the theoretical channel utilization under network convergence can be determined by considering the proportion of time available for data transmission in just a single frame, as given by (5):

$$Ucvg = \frac{N \times T_{dp}}{Scvg \times T_s} \qquad (5)$$

Fig. 7 shows a comparison of the theoretical channel utilization (based up on the frame parameters and calculation using (5)) with the simulation results for the purpose of validation. It can be seen that a very close match is obtained.

Fig. 8 illustrates simulation results of channel utilisation of UW-ALOHA-Q using 50 nodes and shows a similar trend to the channel utilisation results obtained when 25 nodes are used. The number of slots for network convergence (*Scvg*) varies from 2 to 17 as the network size decreases and is achieved when the index ratio is larger than 3.0. 'The effect of a slot' is moderated in the network with 50 nodes compared to the network with 25 nodes, because a greater number of data packets compensates for the inefficient use of time in a frame.

For a comparative analysis, simulation results of framed slotted ALOHA and ALOHA-Q are also shown in Fig. 8 when 50 slots per frame and the number of slots for network convergence (*Scvg*) are used. UW-ALOHA-Q achieves a much higher channel utilisation compared to ALOHA-Q when the number of slots per frame is equal to the number of nodes (i.e. 50). This improvement is greater in larger networks, for example, a 2.8 fold increase in a 100 m size network and a 24.6 fold increase in 900 m and 1000 m size networks. This result demonstrates the great benefits of UW-ALOHA-Q particularly in large networks where most underwater acoustic networks struggle due to the increasing propagation delay in the acoustic channel. Compared with framed slotted ALOHA, UW-ALOHA-Q shows lower channel utilisation. However, framed slotted ALOHA cannot guarantee collision free communication and requires time synchronisation. When framed slotted ALOHA is simulated using the number of slots for network convergence (*Scvg*), most cases show almost zero channel utilisation.

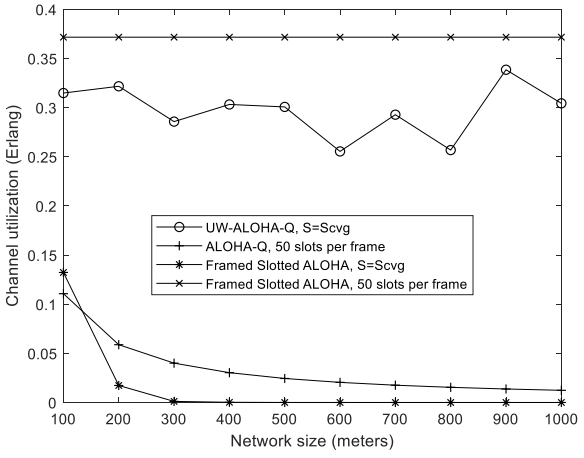

FIGURE 8. Channel utilisation with a 50 node ring topology and variable network size

C. END TO END DELAY

Most importantly, one of the outstanding benefits of UW-ALOHA-Q is that the network achieves maximum channel utilisation when the number of slots for network convergence (*Scvg*) is used whereas ALOHA-Q and framed slotted

ALOHA achieves maximum channel utilisation when the number of slots per frame is equal to the number of nodes as Fig. 9 shows.

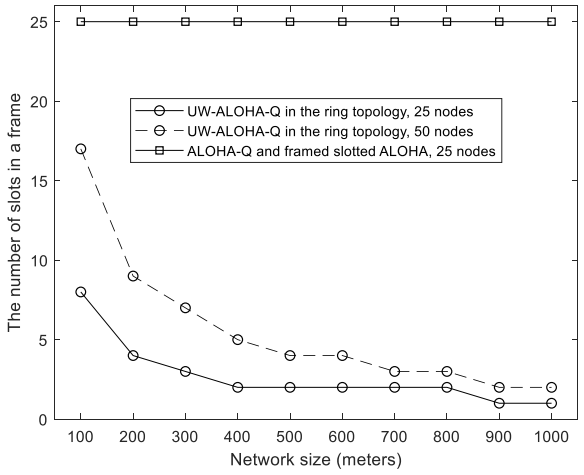

FIGURE 9. The number of slot per frame used for UW-ALOHA-Q, ALOHA-Q and framed slotted ALOHA in different sizes of network

In any size of networks, one node of ALOHA-Q (and framed slotted ALOHA) needs to wait for a much longer time for the next transmission than UW-ALOHA-Q and this becomes more serious in the underwater environment. In a 1000 m network, a slot duration is 1.35 seconds calculated by (2). UW-ALOHA-Q uses only one slot to accommodate 25 nodes in a frame to achieve network convergence, so the frame duration is 1.35 seconds. However, ALOHA-Q needs 25 slots in a frame, hence the frame duration becomes 33.75 seconds. Using the reduced number of slots per frame, UW-ALOHA-Q can provide the significantly lower end to end delay than ALOHA-Q as shown in Table V. The table shows the average end to end delay and channel utilisation of 100 simulation trials for each result.

TABLE V
END TO END DELAY OF UW-ALOHA-Q AND ALOHA-Q IN A 100 M AND 1000 M NETWORK WHEN 25 NODES ARE DEPLOYED

Protocol	Number of slots in a frame (*S*)	Network size (m)	The average end to end delay of successfully delivered data packets (seconds)	The average channel utilisation (Erlangs)
UW-ALOHA-Q	8	100	243	0.35
ALOHA-Q	25	100	758	0.11
UW-ALOHA-Q	1	1000	271	0.31
ALOHA-Q	25	1000	6787	0.01

When 50 nodes are deployed, this benefit of UW-ALOHA-Q is magnified as shown in Table IV. UW-ALOHA-Q uses 2 slots in a frame for a 1000 m size network, so the frame duration becomes 2.7 seconds, whilst ALOHA-Q needs 50 slots in a frame which has 67.55 seconds duration.

TABLE VI
END TO END DELAY OF UW-ALOHA-Q AND ALOHA-Q IN A 100 M AND 1000 M NETWORK WHEN 50 NODES ARE DEPLOYED

Protocol	Number of slots in a frame (*S*)	Network size (m)	The average end to end delay of successfully delivered data packets (seconds)	The average channel utilisation (Erlangs)
UW-ALOHA-Q	17	100	540	0.31
ALOHA-Q	50	100	1516	0.11
UW-ALOHA-Q	2	1000	555	0.30
ALOHA-Q	50	1000	13576	0.01

Through reducing the number of slots per frame, UW-ALOHA-Q improves channel utilisation and decreases the end to end delay. Notably, greater benefits can be obtained in larger networks using a greater number of nodes in a network. This results demonstrate that UW-ALOHA-Q becomes more efficient in large scale networks where high propagation delay and high collision probability exist.

D. NETWORK CONVERGENCE

As tables V and VI show, standard ALOHA-Q using 25 and 50 slots per frame (i.e. $S = N$) exhibits low channel utilisation due to the propagation delay. However, the protocol achieves network convergence in a short time since the slots allow the network to converge easier.

It is useful to see a clearer picture of how the channel utilisation varies over time, to better understand the impact of the network being able to converge. Fig. 10 shows the channel utilisation as a function of time of UW-ALOHA-Q with/without the uniform back-off scheme and compared with ALOHA-Q in a 200 m network where 25 nodes are deployed. Three asterisk marks in Fig. 10 indicate the times at which the network converges. Channel utilisation is measured using (3) from the first frame at the end of every frame.

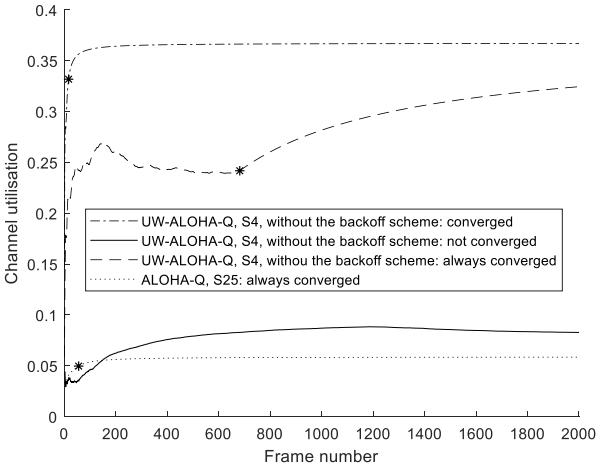

FIGURE 10. Channel utilisation as a function of time of ALOHA-Q using 25 slots per frame and UW-ALOHA-Q using 4 slots per frame (*Scvg*) in a 200 m network when 25 nodes are deployed.

Applying two improvements on top of ALOHA-Q (i.e. UW-ALOHA-Q without the uniform random back-off scheme), most simulation result shows fast (77% converge within 34 frames) convergence so that the network reaches the maximum channel utilisation rapidly. However, there is a small possibility that the network cannot converge due to the randomly inherited frame start time which cannot be changed. In that case, the network never converges hence the channel utilisation remains low. It is because there is a high instance of collisions in the channel and these collisions are not avoidable using the fixed frame start time. The back-off scheme solves this problem.

UW-ALOHA-Q using 4 slots per frame (*Scvg*) needs more frames to converge since the uniform back-off scheme disturbs nodes which achieves convergence and consequently triggers multiple additional learning processes. However, applying the scheme, the protocol can provide network convergence and collision free scheduling. The channel utilisation of UW-ALOHA-Q using 4 slots per frame in Fig. 10 fluctuates when the simulation starts which shows that nodes are learning the optimised frame start time and a distinct slot number through trial-and-error learning processes. Once the network converges, the result shows an increasing channel utilisation due to collision free scheduling.

It is important to note that UW-ALOHA-Q achieves much higher channel utilisation than standard ALOHA-Q when it converges, and its channel utilisation performance remains superior to ALOHA-Q even in a situation where it does not converge. This implies that UW-ALOHA-Q can obtain higher channel utilisation in the time-varying environment: if environmental changes occurs the channel utilisation and the end to end delay performance fluctuate temporarily but the scheme is capable of adapting and maintaining a good level of performance overall.

Please note that each graph in Fig. 10 shows typical examples of four individual results rather than the average of multiple simulation trials. The time at which convergence occurs varies and Table VII shows the results of 100 simulation trials.

TABLE VII
THE RANGE OF NUMBER OF FRAMES USED FOR NETWORK CONVERGENCE

Protocol	Number of slots in a frame (*S*)	Min number of frames used for network convergence (frames)	Max number of frames used for network convergence (frames)	The average number of frames used for network convergence (frames)
UW-ALOHA-Q	4	25	1811	400
ALOHA-Q	25	21	60	43

This paper focuses on the network performance following convergence where collision free scheduling is achieved. Collisions occur during the initial learning process, but this period of time is very small with respect to the period over which such a network would be operational. The achievable channel utilization following convergence is therefore an important metric and we do not consider performance metrics during the learning process, such as collision ratio.

E. RANDOM TOPOLOGY

Let's now look at a more practical underwater topology for environmental monitoring where the position of each sensor node is dictated by the location at which data must be gathered. Nodes tend to be deployed in a random topology rather than in a well aligned ring topology and this feature of underwater applications necessitates UW-ALOHA-Q simulations in a random topology to determine whether the protocol can function in the topology.

For simulations of a random topology, generating nodes are located randomly within a circle of each network size. Simulation results show that UW-ALOHA-Q achieves convergence using the identical number of slots per frame described in section C. This is the most interesting benefit of UW-ALOHA-Q since the protocol can provide the identical baseline performance in the random topology. Fig. 11 shows channel utilisation of UW-ALOHA-Q when 25 nodes are deployed in different sizes of networks.

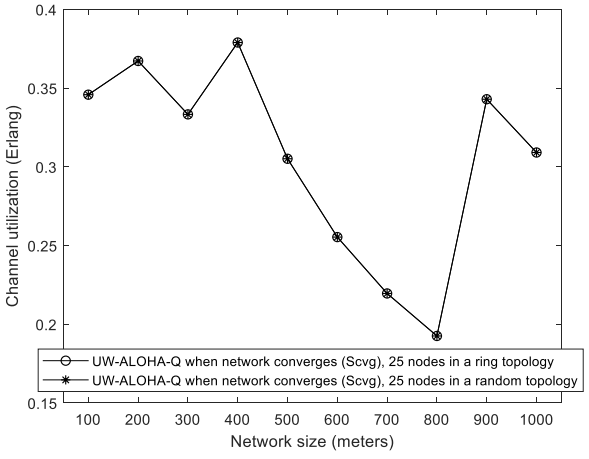

FIGURE 11. Channel utilisation of UW-ALOHA-Q network converges (*Scvg*) in the two different topologies using 25 nodes

A successful data packet transmission is determined by an acknowledgement packet if it is delivered before the guard time ends. Therefore, UW-ALOHA-Q operates identically irrespective of whether the nodes are equally spaced or not. Nodes conduct ordinary trial-and-error learning and can find an appropriate frame start time and a slot number for data transmission in a random topology. A random topology in a circle is simulated, but in principle the random topology in a spherical area also can achieve the identical performance.

ALOHA-Q also achieves convergence and the same channel utilisation in a random topology as it does in a ring topology as Table VIII shows.

TABLE VIII
END TO END DELAY OF UW-ALOHA-Q AND ALOHA-Q IN A 100 M AND 1000M RANDOM TOPOLOGY WHEN 25 NODES ARE DEPLOYED

Protocol	Number of slots in a frame (*S*)	Network size (m)	The average end to end delay of successfully delivered data packets (seconds)	The average channel utilisation (Erlangs)
UW-ALOHA-Q	8	100	243	0.35
ALOHA-Q	25	100	758	0.11
UW-ALOHA-Q	1	1000	271	0.31
ALOHA-Q	25	1000	6788	0.01

Fig. 12 shows the real time channel utilisation of ALOHA-Q and UW-ALOHA-Q in the random topology. This shows four individual results rather than the average value and the similar trend is shown as same as the UW-ALOHA-Q in a ring topology.

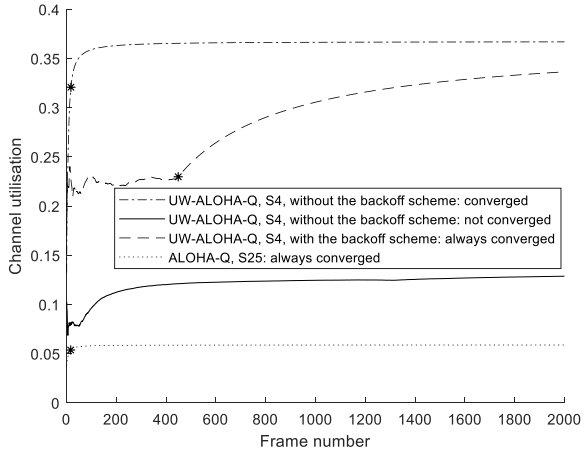

FIGURE 12. Real time channel utilisation of ALOHA-Q using 25 slots per frame and UW-ALOHA-Q using 4 slots per frame (*Scvg*) in a 200 m network when 25 nodes are deployed.

The results demonstrate that UW-ALOHA-Q is robust and tolerant to randomness in a network implying that UW-ALOHA-Q can potentially incorporate random-moving nodes in the operation of underwater acoustic networks.

VI. CONCLUSION

In this paper, we have proposed a reinforcement learning based MAC protocol for underwater acoustic sensor networks, namely UW-ALOHA-Q. ALOHA-Q is designed for the terrestrial environment and this paper has transformed the protocol to UW-ALOHA-Q for use in underwater acoustic networks. Three improvements are proposed for UW-ALOHA-Q: asynchronous operation, reduction in the number of slots per frame, and a uniform random back-off scheme. End to End learning is achieved by the interaction using acknowledgement packet reception between a sink node and generating node. UW-ALOHA-Q takes the benefits of ALOHA-Q which are low complexity and low overheads to achieve collision free high channel utilisation for distributed networks where centralised scheduling is not feasible and distributed scheduling introduces significant signaling overheads and complexity. Practically, H/W computation for UW-ALOHA-Q requires minimum integer values of Q-learning and little storage for Q-values of one frame. Moreover, UW-ALOHA-Q significantly improves performance for use in underwater networks without the need for time synchronisation. A comprehensive simulation study shows that UW-ALOHA-Q has considerable potential for use in practical random and large scale underwater applications. For the example scenario considered, UW-ALOHA-Q achieves up to a 24.6 times improvement in channel utilization with much lower end to end delay than ALOHA-Q in a 1000m radius underwater network.

ACKNOWLEDGMENT

S. H. P. Author thanks D.B. for help with academic writing in English.

REFERENCES

[1] *Unites States Geological Survey (USGS)*, May. 2019. [Online]. Available: https://water.usgs.gov/edu/earthhowmuch.html

[2] *United Nations (UN)*, May. 2019. [Online] Available: https://www.un.org/sustainabledevelopment/oceans

[3] *Orgnisation for Economic Cooporation and Development (OECD)*, May. 2019. [Online] Available: http://www.oecd.org/sti/inno/ocean-economy/

[4] J. Partan, J. Kurose, and B. N. Levine, "A survey of practical issues in underwater networks," *ACM SIGMOBILE Mobile Computing and Communications Review,* vol. 11, no. 4, pp. 23-33, 2007.

[5] C. E. Shannon, *The mathmetical Theory of Communication,* Urbana, IL, USA: University of Illinois Press, 1949.

[6] L.P. Kaelbling, M. L. Littman and A. W. Moore, "Reinforcement learning: A survey", *Journal of Artificial Intelligence Research*, vol. 4, pp. 237-285, 1996.

[7] S. H. Park, P. D. Mitchell, and D. Grace, "Performance of the ALOHA-Q MAC Protocol for Underwater Acoustic Networks," in *IEEE International Conference on Computing, Electronics & Communications Engineering*, 2018, pp. 189-194.

[8] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part I - carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE transactions on Communications,* vol. 23, no. 12, pp. 1400-1416, 1975.

[9] P. Karn, "MACA - a new channel access method for packet radio," in *ARRL/CRRL Amateur radio computer networking conference*, 1990, pp. 134-140.

[10] Z. Liu and I. Elhanany, "RL-MAC: A QoS-aware reinforcement learning based MAC protocol for wireless sensor networks," in *IEEE International Conference on Networking, Sensing and Control,* 2006, pp. 768-773.

[11] S. Galzarano, A. Liotta, and G. Fortino, "QL-MAC: a Q-learning based MAC for wireless sensor networks," in *International Conference on Algorithms and Architectures for Parallel Processing*, 2013, pp. 267-275.

[12] J. Niu and Z. Deng, "Distributed self-learning scheduling approach for wireless sensor network," *Ad Hoc Networks*, vol. 11, no. 4, pp. 1276-1286, 2013.

[13] Y. Chu, S. Kosunalp, P. D. Mitchell, D. Grace, and T. Clarke, "Application of reinforcement learning to medium access control for wireless sensor networks," *Englieering Applications of Artificial Intelligence*, vol. 46, pp. 23-32, 2015.

[14] H. Bayat-Yeganeh, V. Shah-Mansouri, and H. Kebriaei, "A multi-state Q-learning based CSMA MAC protocol for wireless networks," vol. 24, no. 4, pp. 1251-1264, 2018.

[15] G. Chen, Y. Zhan, G. Sheng, L. Xiao, and Y. Wang, "Reinforcement Learning-Based Sensor Access Control for WBANs," vol. 7, pp. 8483-8494, 2018.

[16] Y. Yu, T. Wang, and S. C. Liew, "Deep-reinforcement learning multiple access for heterogeneous wireless networks," vol. 37, no. 6, pp. 1277-1290, 2019.

[17] H. Okada, Y. Igarashi, and Y. Nakanishi, "Analysis and application of framed ALOHA channel in satellite packet switching networks-FADRA method," *Electronics Communications of Japan*, vol. 60, pp. 72-80, 1977.

[18] H. Wu and Y. Zeng, "Efficient framed slotted Aloha protocol for RFID tag anticollision," *IEEE Transaction on Automation Science and Engineering*, vol. 8, no. 3, pp. 581-588, 2011.

[19] A. George and T. G. Venkatesh, "Performance Analysis of M2M Data Collection Networks Using Dynamic Frame-Slotted ALOHA," *IEEE Transactions on Green Communication and Networking,* vol. 2, no. 2, pp. 493-505, 2018.

[20] S. Wang and Y. Shin, "Efficient Routing Protocol Based on Reinforcement Learning for Magnetic Induction Underwater Sensor Networks," vol. 7, pp. 82027-82037, 2019.

[21] X. Li, X. Hu, W. Li, and H. Hu, "A Multi-Agent Reinforcement Learning Routing Protocol for Underwater Optical Sensor Networks," in ICC 2019-2019 IEEE International Conference on Communications (ICC), 2019, pp. 1-7: IEEE.

[22] Z. Jin, Q. Zhao, and Y. Su, "RCAR: A Reinforcement-Learning-Based Routing Protocol for Congestion-Avoided Underwater Acoustic Sensor Networks," 2019.

[23] 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," 2019.

[24] 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 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), 2018, pp. 702-706: IEEE.

[25] L. Jin and D. D. Huang, "A slotted CSMA based reinforcement learning approach for extending the lifetime of underwater acoustic wireless sensor networks," *Computer Communications,* vol. 36, no. 9, pp. 1094-1099, 2013.

[26] *IEEE 802.15 WPAN Task Group 4*, May. 2019 [Online]. Available: http://www.ieee802.org/15/pub/TG4.html

[27] R.S. Sutton, and A.G. Barto, *Reinforcement learning: An introduction*, Cambridge, MA, USA: MIT Press, 1998.

[28] *Evo Logics*, May. 2019 [Online]. Available: https://evologics.de/acoustic-modem/hs

[29] L. Paull, S. Saeedi, M. Seto, and H. Li, "AUV navigation and localization: A review," *IEEE Jounal of Oceanic Engineering*, vol. 39, no. 1, pp. 131-149, 2014.

[30] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function", *IEEE Journal on Selected Areas in Communications*, vol 18, no. 3, pp. 535-547, 2000.