Reinforcement Learning Based Medium Access Control for Underwater Acoustic Sensor Networks

Sung Hyun Park

Ph.D.

University of York

Electronic Engineering

September 2020

Abstract

This thesis studies the application of reinforcement learning to Medium Access Control (MAC) protocol design for underwater acoustic networks.

The underwater environment constantly changes due to many factors which have a significant impact on wire free communications. Therefore it is of interest to explore whether reinforcement learning can provide benefits in the underwater environment since reinforcement learning is capable of interacting and adapting to a changing environment.

Due to the limited bandwidth of acoustic signals, underwater networks have fundamentally low channel capacity and the slow propagation speed of the signals makes it very difficult to achieve high channel utilisation. MAC protocols play a key role in achieving efficient use of a shared channel since they govern the achievable channel utilisation and the corresponding quality of service required by the applications.

This thesis applies the reinforcement learning approach to the MAC protocol operating in the time-varying underwater channel. To utilise reinforcement learning effectively in underwater networks, three new schemes are proposed: asynchronous operation because time synchronisation is costly in the underwater environment, refinement of frame size for the desired channel utilisation, and finally a new random back-off scheme for better benefits from the reinforcement learning approach. Reinforcement learning based protocols can provide convergence for fixed networks and desirable channel utilisation and adaptability for mobile networks.

Intensive simulation results show that the proposed reinforcement learning based protocols in this thesis outperform existing protocols and can provide an agnostic solution for different underwater networks such as those comprising different types of nodes. This is achieved by applying reinforcement learning to remove the need for complex or inefficient operations that many existing protocols use to deal with the slow propagation delay of acoustic signals and environmental changes in the networks.

List of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my deepest gratitude to my supervisors Paul Mitchell and David Grace for providing me with their insightful guidance and enormous help throughout the past four years. Thanks too are due to underwater research group members in the Department of Electronic Engineering for creating intellectually stimulating discussion.

I would like to dedicate this thesis to my parents who sacrifice so much for me and my study. Their endless love and support have motivated me to complete this thesis through all the good times and difficult times.

I appreciate my friends – Dianna, Liz, Kevin, and most of all Bailey.

Declaration

All work presented in this thesis is original to the best knowledge of the author. References to other researchers have been given as appropriate. This work has not previously been presented for an award at this or any other institution. The research presented in this thesis features in a number of the author's publications listed below.

Conference paper

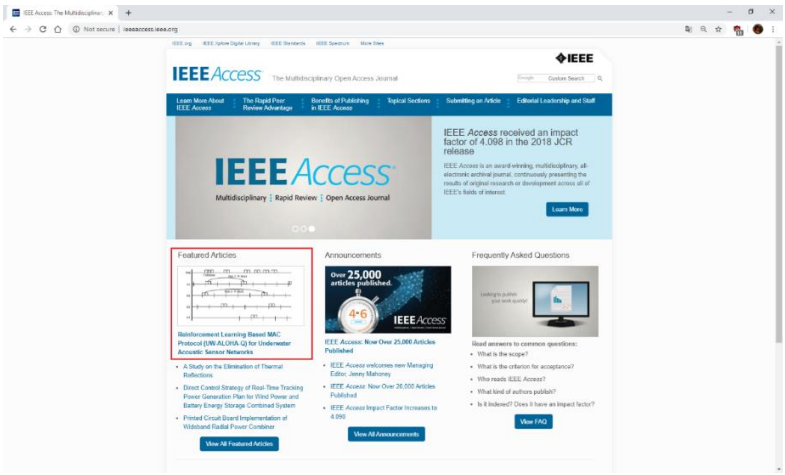
Sung Hyun Park, Paul Daniel Mitchell, and David Grace, "Performance of the ALOHA-Q MAC protocol for underwater acoustic networks," in International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, UK, Aug. 16-17, pp. 189-194. IEEE, 2018.

This paper received a best paper award at the conference.

Journal papers

Sung Hyun Park, Paul Daniel Mitchell, and David Grace, "Reinforcement Learning Based MAC Protocol (UW-ALOHA-Q) for Underwater Acoustic Sensor Networks," IEEE Access, vol. 7, pp. 165531-165542, Nov. 2019.

IEEE Access featured this published article as the IEEE Access "Article of the Week" and evaluated the paper as - *With the increased need for marine environment monitoring, a research team developed a novel approach to medium access control that engenders efficient use of an acoustic channel. Read more in this IEEE Access "Article of the Week"*



<Article of the Week of IEEE Access [106] >

1 Introduction

1.1 Background

The market value of coastal resources is estimated to be 3 trillion USD per year [1], contributing 1.5 trillion USD annually to the global economy [2]. Therefore, the marine environment has become central to a vast diversity of industries and areas of scientific importance. Examples of applications using underwater networks include disaster alarm systems from tsunami monitoring networks far off coast [3], natural resource exploration using seabed networks for gas or oil [4], underwater surveillance military networks [5], and ocean cleaning including ocean plastics [6]. However, most of the ocean is still unexplored: most of the underwater realm is unseen by human eyes because ocean exploration has been hampered by the 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 explore the oceans more effectively and to do so remotely, continuously and potentially in real time.

1.2 Current underwater networks

Current underwater networks have several limitations. To deploy sensor nodes, they need to be moved to the sea by ship and deployed on the sea bed or in the water column. Then the devices collect data for the mission period. After the mission, sensors needs to be taken back to the data centre, and finally the data sensed by nodes can be analysed. For example, in 2009, Air France flight 447 crashed in the equatorial Atlantic Ocean. The salvage was conducted through five phases for nearly two years. Multiple REMUS (Remote Environment Monitoring UnitS) 6000 Autonomous Underwater Vehicles (AUVs) [7] were used during phase 3 and phase 4 and the first plane wreckage was detected by the AUV side scan sonars 10 days into phase 4. Each AUV weights approximately 880 kg and a REMUS6000 mission includes the preparation, launch, descent, seafloor search, ascent, recovery, and data download. There are number of disadvantages of this existing approach which leads to the need for effective wireless communication in the underwater environment [8]:

- Real time monitoring is not possible.
- No interaction is possible between onshore control systems and monitoring instruments.
- If misconfiguration, failure, or loss occurs it may not be possible to detect such issues: they are not adaptable and not reconfigurable.
- Ocean data collection is limited by the duration of the mission.

1.3 Challenges

This section provides reasons why wire free communications are limited in their ability to achieve good performance in underwater networks.

1.3.1 Uncertainty

The underwater environment constantly changes due to many factors, notably variable wave motion which has a significant impact on wireless communications. This means that the underwater networks are situated in a time-varying environment. Therefore, it is required that underwater networks are capable of being adaptive to continuous environmental changes.

1.3.2 Energy sources

Energy sources are a challenge in the underwater environment as solar energy cannot be exploited because water absorbs much of the spectrum of sunlight, even though blue-green light is the last portion of the spectrum to be absorbed. Renewable underwater energy sources such as wave, tidal, and ocean thermal energy have been intensively studied nowadays but the underwater energy stations are not common and practical at the current time. Therefore, the underwater sensors have to be carried to the land or a ship for battery recharging.

1.3.3 Time synchronisation

Most wireless networks in the terrestrial environment use time synchronisation for data transmissions since it is easy and cheap to achieve. However, GPS signals are not available underwater because they are absorbed by water. Therefore, the reliance on time synchronisation for data communication in underwater networks becomes costly and increases the complexity of the system, especially the periodic applications. 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 lack of GPS also restricts options for navigation and tracking.

1.3.4 Inefficient channel use

Wireless Sensor Networks (WSNs) using radio technology have been widely used to collect data in many applications. Unfortunately, this technology cannot be directly applied under water since radio waves are absorbed by water. Acoustic signals are the most viable means of communicating underwater due to their longer propagation distance compared with alternatives such as radio or optical signals. However, the slower propagation speed of acoustic signals in water compared to radio signals in the air 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 [9].

1.3.5 Costs

The components for underwater networks tend to be bulky and expensive because of the housing requirements to deal with high pressures, the presence of salt, other minerals in the water, etc. For example, a simple underwater cable connector typically costs over one hundred USD [10]. Not only components but also deployment costs, for example, the cost of oceanographic research vessels supporting missions is significant even for a single days use [11].

1.4 Scope of the thesis

This thesis investigates whether reinforcement learning can be used in MAC protocols for mobile underwater networks as a means of improving performance and providing a flexible topology agnostic solution in particular, considering challenges of environment uncertainty, lack of GPS signal, limited channel, and the inefficient channel use discussed in sections 1.3.1, 1.3.3, and 1.3.4. MAC protocols play a key role in making efficient use of a multiple access channel since their operation governs the achievable channel utilisation and corresponding quality of service. Many state of the art MAC protocols designed for mobile underwater networks were originally designed for fixed sensor networks, and they were extended to mobile networks. The extension incorporates additional functions to deal with node movement, such as carrier sensing, transmission prediction schemes, or more frequent control message exchanges. Those existing MAC schemes can play a role of coordinating multiple accesses from mobile nodes, however their performance is significantly limited because those additional approaches are not optimal operations in the underwater environment and they are workarounds to deal with node movement rather than to improve network resilience and adaptability. Moreover, applying reinforcement learning to the medium access control problem in underwater networks is a new research area in so far that existing MAC protocols based on reinforcement learning only support networks comprising fixed nodes and cannot provide efficient learning. Therefore, this thesis proposes a set of reinforcement learning schemes to improve underwater resilience and adaptability. The work of this thesis falls into two main areas: firstly, the application of reinforcement learning to fixed networks and secondly, applying it to mobile networks. Using the fundamental nature of reinforcement learning, based upon trial-and-error interaction with a changing environment, the newly proposed protocol called UW-ALOHA-QM can achieve high channel utilisation without the need for time synchronisation and with a very low level of overheads. Results show that UW-ALOHA-QM

outperforms other existing protocols and achieves a significant improvement in channel utilisation in a number of distinct and representative scenarios.

1.5 Hypothesis

Reinforcement learning techniques are powerful means of providing an agnostic solution in different scenarios such as free floating networks, anchored or moored networks, AUV assisted networks, and AUV networks for medium access control in underwater networks.

1.6 Structure of the thesis

This thesis is defined into six chapters and the contents are outlined in this section. Chapter 2 provides a brief comparison of available signals which can be used for underwater communications and covers fundamental knowledge of multiple access techniques and MAC protocols. This chapter then provides a detailed literature review, describing state of the art protocols in detail, focusing on their features and relative merits in the underwater environment. Moreover, a Q-learning algorithm is introduced which is used in the UW-ALOHA-Q and UW-ALOHA-QM protocols which are proposed in chapters 4 and 5. Finally this chapter motivates the use of reinforcement learning for the medium access control problem and reviews the existing research literature on reinforcement learning based MAC protocols in both terrestrial and underwater environments.

Chapter 3 describes the ALOHA-Q protocol which was designed for WSNs and compares the initial simulation results of the protocol in terrestrial and underwater environments. The purpose of the initial simulation is to examine whether the reinforcement learning based protocol can be used in the underwater environment. Initial simulation results shows that there is a significant decrease in channel utilisation when ALOHA-Q is operated in the underwater environment.

Chapter 4 introduces UW-ALOHA-Q for underwater networks consisting of fixed sensor nodes. This protocol includes three novel schemes which are designed by considering the properties of underwater acoustic networks, notably the costs of time synchronisation and inefficient channel utilisation. This chapter provides a depth of understanding of the protocol and presents simulation results which demonstrate that the proposed protocol provides high channel utilisation as well as network convergence through the reinforcement learning approach.

Chapter 5 extends UW-ALOHA-Q to UW-ALOHA-QM in order to deal with significant and continuous changes in a network caused by mobile nodes. UW-ALOHA-QM is the first

reinforcement learning based MAC protocol for mobile nodes in underwater sensor networks. Simulation results show that UW-ALOHA-QM can increase network resilience and adaptability and hence can achieve a higher channel utilisation than existing protocols designed for underwater mobile networks.

Chapter 6 presents the conclusions of this thesis and provides suggestions for future work.

2 Literature review

2.1 Signals underwater

This section explains and compares features of radio, optical, and acoustic signals in the underwater environment.

2.1.1 Radio signals

Radio signals range between 3 kHz to 3 THz in the electromagnetic spectrum. High frequency radio bands facilitate networks requiring large bandwidths for high data rate applications and lower frequencies generally benefit from more favourable propagation for non-line of site and long distance communication in the terrestrial environment. However, radio waves are attenuated severely underwater and thus are not able to travel long distances.

2.1.1.1 Path loss

When a signal propagates in any channel, the loss increases with an increase in distance. The degradation over a particular distance from the transmitter to receiver is denoted by a term called path loss. The relationship between signal power and distance can be expressed as:

$$P_r \propto d^{-n} \tag{2-1}$$

where, P_r is received signal power, d is distance, and n is the path loss exponent. The path loss exponent is the loss of signal strength when it propagates in different environments and a higher values represent more lossy environments. Table 2-1 shows the practical values in different environments.

Environments	Path loss exponent (n)
Free space	2
Urban area	2.7 to 3.5
Suburban area	3 to 5
Indoor (Line of Sight)	1.6 to 1.8
Underwater (Line of Sight)	2 to 4

Table 2-1. Path loss exponent of radio signals for different environments [12]

2.1.1.2 Absorption loss

The primary reason for path loss in the underwater medium is the radio signal property of water (conductivity) which behaves differently over different frequency bands. Therefore, frequency has a significant influence on the radio wave propagation distances underwater. The electrical conductivity (δ) of the medium is measured in the unit of Siemens per meter (S/m).

As shown in Table 2-2, the conductivity of fresh water is typically around 0.001 S/m, sea water commonly 4 S/m (400 times higher), and the Red Sea is 8 S/m. An increase in conductivity results in an increase in attenuation, such that the higher the conductivity of the water, the shorter the propagation range.

Water	Conductivity values (S/m)
Fresh water	$0 \leq \delta < 1$
River water	$1 \leq \delta < 2$
Sea water	$2 \leq \delta$

Table 2-2. Water conductivity [12]

2.1.2 Optical signals

Optical signals are defined in the range from 400 THz to 700 THz (400nm to 700nm). The main characteristic of the signals underwater is the availability of a lot of bandwidth for high data rate communication, which can be achieved over distances of the order of a few hundred meters depending on the turbidity of the water. Different wavelengths of light are absorbed differently in water, for example, blues and greens penetrate deepest as shown in Figure 2-1 [13]. This leads to research subject projects being undertaken on the use of blue-green light for underwater communications [14-15]. Unlike the case of radio signals, conductivity does not play a major role in optical underwater communication [16].

2.1.2.1 Scattering

The propagation of optical signals through a medium is affected by absorption, emission, and scattering processes. Scattering is a dominant loss mechanism of optical signals and it is environmentally dependent and unknown, which means the available propagation distance of optical signals depends significantly on the environmental conditions. Moreover, transmission of optical signals requires high precision in directing the narrow laser beams. Therefore, optical signals are not appropriate to medium or large underwater networks.

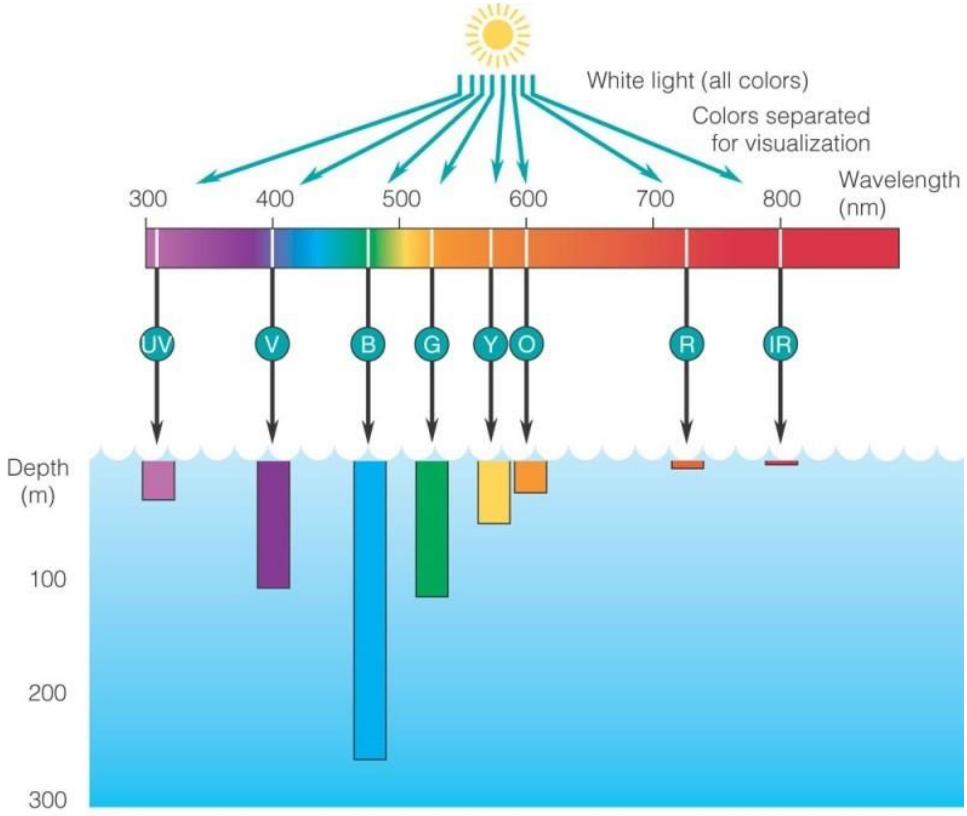


Figure 2-1. Characteristics of optical signals underwater: taken from [13]

2.1.3 Acoustic signals

The characteristics of radio and optical signals are significantly different in the terrestrial and underwater environments as discussed in sections 2.1.1 and 2.1.2. Acoustic signals which fall between 20 Hz and 20 kHz are considered to be more appropriate for underwater communication since they can propagate over longer distances than radio and optical signals.

2.1.3.1 Speed of acoustic signals in oceans

The speed of sound in sea water depends on its temperature, as well as on the salinity and hydrostatic pressure. For calculation of the speed of sound, Wilson's empirical formula offered in 1960 is in common use [17]. Wilson's formula is accepted by the National Oceanographic Data Centre (NODC) in the USA for computer processing of hydrological information. Equation (2-2) shows the simplified version of Wilson's formula:

$$c = 1449 + 4.6\,T - 0.055\,T^2 + 0.0003\,T^3 + 1.39\,(S-35) + 0.017\,D \qquad (2\text{-}2)$$

where, c is speed of sound (m/s), S is salinity (Practical Salinity Unit), T is temperature (°C), and D is depth (m). Using Equation (2-2), a sound speed profile underwater can be derived as shown

in Figure 2-2 [18]. Between surface and a depth of 100 m, the sound speed is approximately 1,500 m/s.

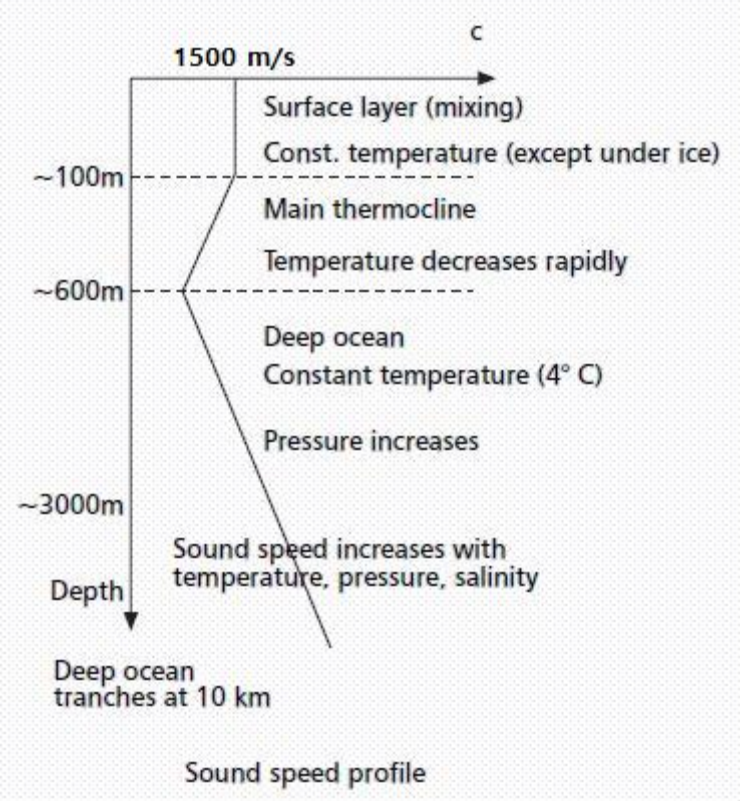


Figure 2-2. Depth profiles of sound speed: taken from [18]

This slow propagation speed impacts on MAC protocols for underwater networks. Existing protocols designed for wireless sensor networks in the terrestrial environment assume a high propagation speed ($\approx 3 \times 10^8$ m/s) and therefore the existing protocols cannot be directly applied to underwater networks. The slow propagation speed must be considered in the design of MAC protocols for underwater networks since the large propagation delay brings about low channel utilisation and high latency in the network.

2.1.3.2 Transmission loss

Transmission loss is the accumulated decrease in acoustic intensity between a transmitter and a receiver and the loss consists of spreading and attenuation. Attenuation can be divided into absorption and scattering. Scattering depends on frequency and a dominant factor below 100 Hz.

Spreading and absorption are primary causes of transmission loss of acoustic signals through underwater above 100 Hz. In shallow water, the transmission loss can be expressed by cylindrical spreading (spreading factor) plus absorption:

$$\text{Transmission loss} = \text{cylindrical spreading} + \text{absorption} = 10 \log r + \alpha r \times 10^{-3} \quad (2\text{-}3)$$

where, r is range in meters and α is the absorption coefficient in dB/km. The absorption coefficient depend on acoustic frequency, pressure, acidity, temperature, and salinity in the sea water.

In a similar way, the direct path model in deep water can be expressed as:

$$\text{Transmission loss} = \text{spherical spreading} + \text{absorption} = 20 \log r + \alpha r \times 10^{-3} \quad (2\text{-}4)$$

The propagation distances of sound waves depends to a great extent on frequency in the underwater environment. Therefore, acoustic signals at a higher frequency travel a shorter distance due to high transmission loss whilst at lower frequencies, they travel longer distances. Consequently, the available bandwidth of acoustic signals is limited and this leads to fundamental low channel capacity according to Shannon's theory [9].

2.1.4 Discussion

Section 2.1 compares features of radio, optical, and acoustic signals in the underwater environment. Each type of signal has its own pros and cons. However, acoustic signals are more feasible for use in underwater communications than other alternative signals due to their longer propagation distances. Table 2-3 summarises the characteristics of the three signals.

	Acoustic	Radio	Optical
Frequency	500 Hz (long range) to 50kHz (medium range)	100 Hz (up to 100m) to 100 kHz (few m)	Blue-green: 10^{14} Hz
Noise sources	Rain, marine lives, thrusters, electronic preamplifier noise	Motor, lightning, pump, solenoid electromagnetic noise, electromagnetic preamplifier nose	Sunlight, detector short, preamplifier noise
Latency	High 1450-1550 m/s	Frequency dependent 1500 m/s (1Hz) 1e6 m/s (1MHz)	Fixed low 2.25e8 m/s
Absorption loss	Low (frequency dependent) 0.05 dB/m (150 kHz) 0.0001 dB/m (1.5 kHz)	High (frequency dependent) 5.4 dB/m (25 kHz) 1.1 dB/m (1 kHz)	High (turbidity dependent) 0.1 dB/m (deep ocean) 10 dB/m (shallow coastal)

Data rate	10s bps to 10s kbps	10s bps to 100 kbps	100s kbps to several 10s Mbps
Antenna size	cm (medium range) few 10s cm (long range)	Few 10s cm to several meters	Up to 10 to 20 cm
Practical range	Meters to 10s km at kbps	Less than 10 m at kbps	10 m to 100 m
Antenna aspect	Omnidirectional and directional	Requires appropriate antenna orientation	Directional and narrow beam
Application examples	Long rang communication, long range sensor networks	Short range communication, short range cross media and NLOS communication (air-water, water-bottom), short range sensor networks	Short range high bandwidth data rate uplift, real time video command control

Table 2-3. Summary of signals

Based on the characteristics of the three types of signals underwater, the benefits and limitations are considered in Table 2-4.

	Acoustic	Radio	Optical
Benefits	Most widely used underwater wireless communication technology, long communication range over 100s of km	Relatively smooth transition to cross air and water boundaries, more tolerant to water turbulence and turbidity, loose pointing requirements	Ultra-high data transmission range (up to Gbps), immune to transmission latency, higher communication security
Limitations	Low data transmission rate (on the order of kbps), large communication latency (on the order of second), not proper to applications of real time large volume data exchange	Short link range (a few meters at extra low frequencies 30 to 300 Hz), huge transmission antenna	Cannot cross water and air boundary easily, suffer from severe absorption and scattering, moderate link range (up to ten of meters), transmission of optical signals requires high precision in pointing the narrow later beams

Table 2-4. Benefits and limitations

Figure 2-3 [19] summaries the available propagation distances of different signals and shows experimental results using them. While optical signals and radio signals provide higher capacities than acoustic signals, the available transmission range is much shorter than acoustic signals. Regarding the huge size of oceans and the cost of devices, acoustic communication is most

suitable for underwater networks but even so two critical limitations exist. First, the slower propagation speed (≈ 1,500 m/s) of acoustic signals in water compared to radio signals in the air (≈ 3×10^8 m/s) invariably leads to poor channel utilisation and inefficient channel use in underwater networks. Secondly, the limited and distance dependent bandwidth brings about low fundamental channel capacity.

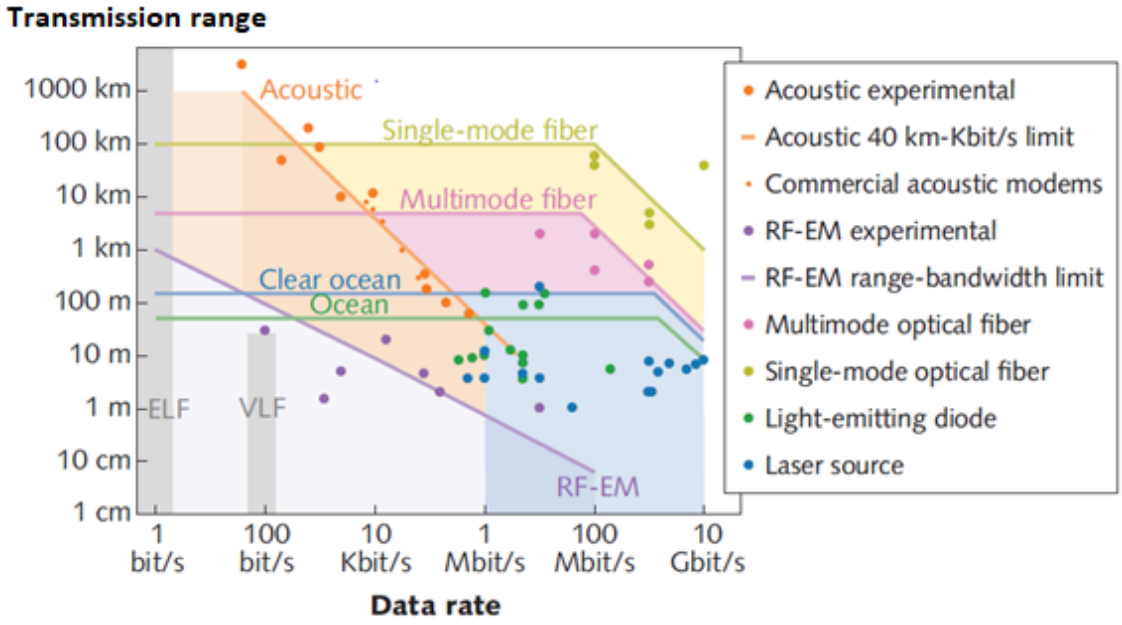


Figure 2-3. Comparison of the different underwater communication channels: taken from [19]

2.2 Medium access control

The International Standards Organisation (ISO) defines a reference model for packet switched networks called the Open Systems Interconnection (OSI) model [20]. The model consists of seven layers and the second layer is the called data link layer which is divided into Logical Link Control (LLC) and Medium Access Control (MAC). The MAC layer is positioned over the first physical layer and transforms the bit string into messages as shown in Figure 2-4. In the figure, H stands for Header of a message.

The objective of the MAC layer is to make efficient use of the available channel capacity in a network. At the same time, MAC protocol design has a notable impact on delivering the Quality of Service (QoS) requirements of applications, for example, packet error rate, end to end delay, energy consumption, throughput, etc. MAC protocols can achieve these objectives by assigning

channel capacity to multiple users and by coordinating and regulating their data transmissions on the shared channel.

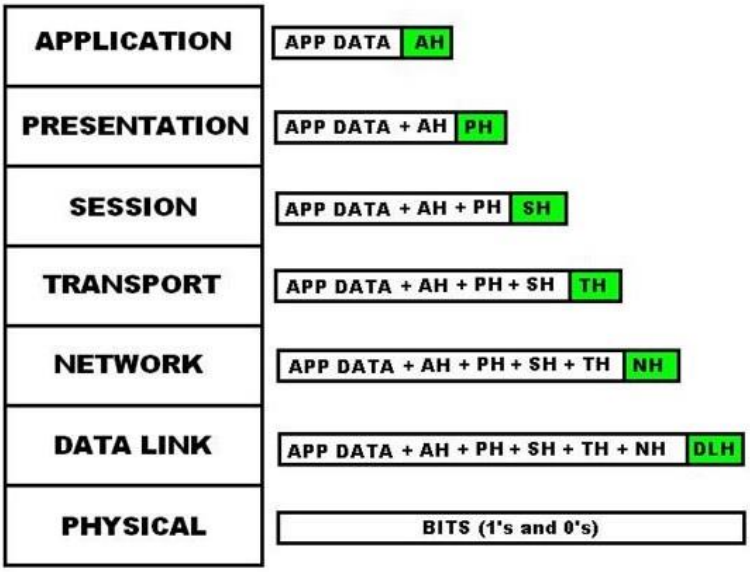


Figure 2-4. OSI seven layer model

Underwater acoustic channels are significantly limited in terms of bandwidth as described in section 2.1.3, hence the available capacity must be used effectively. The achievable utilisation efficiency is governed by the underlying MAC protocol. Therefore, the MAC layer can play a key role in underwater acoustic networks to handle the inefficient channel use which is due to the slow propagation speed.

2.2.1 Multiple access techniques

The required roles of the MAC layer vary depending on the needs of applications and how the applications are implemented. First, the MAC layer provides a multiple access technique(s) which enables multiple user access to a shared medium. According to the system implementation, either frequency, time, or codes are used to allow multiple users fundamentally to share the channel resources. Second, the MAC layer provides the process regulating and governing multiple access to the channel. This is the software control organising multiple access based on the multiple access technique(s) implemented. Section 2.2.1 explains multiple access techniques and section 2.2.3 reviews the associated medium access control protocols.

2.2.1.1 Frequency Division Multiple Access (FDMA)

FDMA divides a shared channel into a number of sub-frequency bands and these are assigned to individual nodes. All nodes can transmit packets simultaneously so that FDMA is appropriate for

constant bit rate traffic. However interference can be caused to other users operating on adjacent channels. This problem can be solved by inserting appropriate guard bands as shown in Figure 2-5.

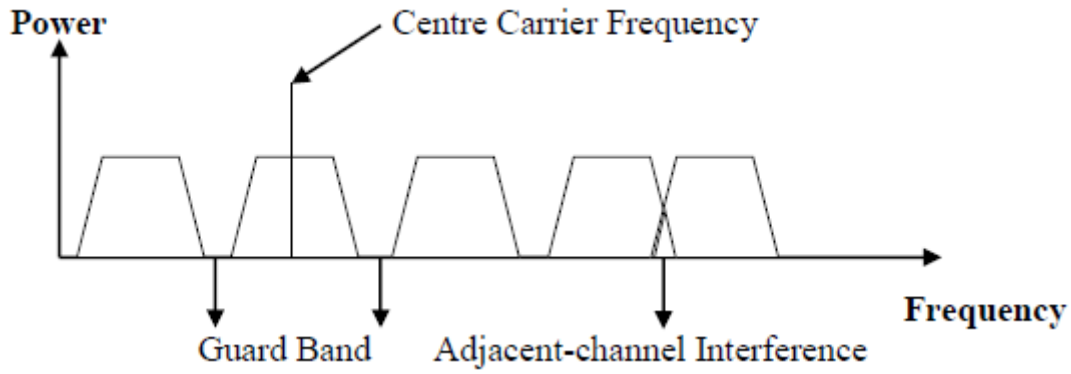


Figure 2-5. Frequency division multiple access

The FDMA technique was tried in some experiments as a key part of the Seaweb project early on in underwater communication system development in the 1990s [21]. Due to the lack of acoustic signal bandwidth as discussed in section 2.1.3, a FDMA system can only provide a very limited number of channels, so FDMA is not a dominant technology in underwater networks. The results of the experiments are available in the reports [21], and the maximum achievable data rate was 50 bits per second in 1999.

2.2.1.2 Code Division Multiple Access (CDMA)

A CDMA system uses (binary) codes to modulate the information stream in a spread spectrum fashion using different spreading sequences which have low cross-correlation. User information signals are multiplied by a unique wide bandwidth spreading code and the resulting signals from multiple users are modulated onto a common carrier frequency. All transmissions take place simultaneously on the shared channel, and CDMA is more robust than FDMA since the entire frequency band is used by all nodes. The received signal is multiplied by an identical spreading code to reproduce the original data, hence code synchronisation is required between a sender and a receiver. The spreading codes allocated to users must exhibit very low cross-correlation to effectively reject unwanted signals at the receiver. Other uses cause some interference due to residual correlation properties between spreading codes. Therefore, as the number of users in the system increases, the total level of interference increases, degrading the channel performance.

Studies of MAC protocols using the CDMA technique for underwater networks began to be published in the 2000s. Early studies [22-24] show initial experiments of CDMA in an underwater acoustic channel and later studies [25-27] propose MAC protocols based on CDMA. For example, Protocol for Long latency Access Networks – MAC (PLAN-MAC) [25] uses CDMA as a multiple access technique and a handshaking reservation scheme (refer to section 2.2.3.2.2) as a multiple access protocol.

2.2.1.3 Time Division Multiple Access (TDMA)

Time in TDMA systems is divided into time slots. Generally, slots have an identical and fixed time duration and one time slot is allocated to a single node for data transmission in each frame. Like CDMA, TDMA can be also more resistant to frequency selective fading than FDMA since one user uses the full bandwidth. Figure 2-6 shows an example of a basic TDMA system in an underwater network.

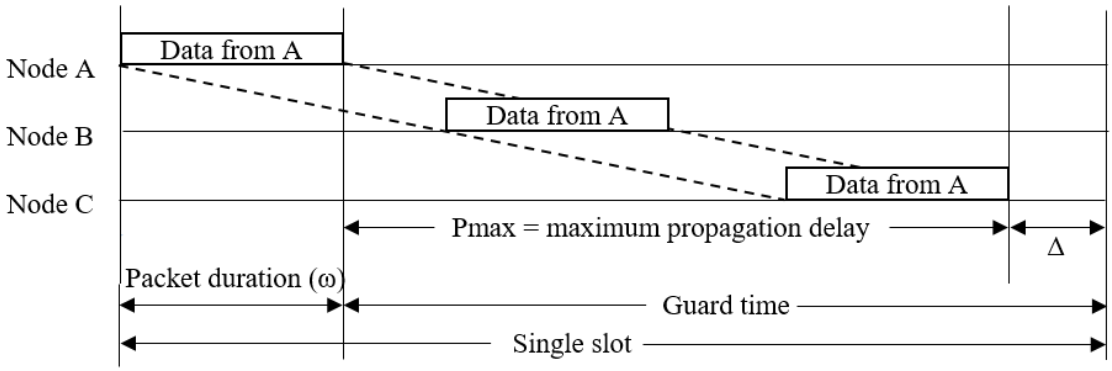


Figure 2-6. Basic TDMA time slot

The duration of the time slot includes the duration of a packet, and a guard time which consists of the maximum propagation time, with some additional allowance for synchronisation errors, and clock drift. Different from FDMA and CDMA, all nodes in a TDMA system need to be time synchronised so that the duration of guard time should be decided, considering the possible differences in time synchronisation of each node. Depending on the quality of the clocks, it may be feasible to synchronise devices prior to deployment and maintain adequate synchronisation during deployment. For less accurate clocks and/or longer term deployment, it is necessary to conduct time synchronisation after a deployment stage or as a separate initialisation stage leading to an additional cost for TDMA. There are a range of time synchronisation techniques for underwater networks [28-30].

The propagation distances need be considered to determine an appropriate duration of guard time in an underwater network to guarantee the data packet delivery to a node at the edge of the network. However, the increase in guard time of a slot results in low channel utilisation since the channel remains idle as for long periods of time shown in Figure 2-6. Therefore, TDMA can be inefficient for long range underwater networks due to the slow propagation speed of acoustic signals. However, TDMA provides good flexibility with more dynamic allocation of slots, in terms of being able to adapt the number of available time slots assigned to nodes based on the number of nodes and their changing requirements. Many studies [31-33] of MAC protocols are based on TDMA in underwater networks.

2.2.1.4 Space Division Multiple Access (SDMA)

SDMA is based on the use of multiple antennas or an antenna array where antenna elements are physically separated, either at the transmitter, the receiver, or both such as the SIMO, MISO, or MIMO systems. It is commonly used in terrestrial systems, as a means of providing diversity because multipath propagation links between different antenna elements will independently vary over time. In terrestrial communication systems, base stations usually have location information for the connected nodes. Therefore, using antenna techniques such as beam forming, the base station can focus the power of their signals in the directions of the associated users rather than radiating broadly to ensure wide area coverage which wastes energy and generates unnecessary interference. However, in the underwater network, the localisation information is costly as we discussed in section 1.3.3 and it is more difficult in mobile underwater networks. Therefore, SDMA is not the major multiple access technique in underwater networks.

2.2.2 Discussion

There are different opinions of how to define the role of the MAC layer in general and the roles vary based on the application requirements or implementation of systems in practice. There are no fundamental capacity differences between sharing a single channel in time, code, or frequency. However, this thesis focuses on TDMA because TDMA is practically more flexible in terms of network configuration. As we discussed in section 1.3.4, underwater acoustics have very limited bandwidth. Therefore, as the number of nodes increases, the limited frequency must be divided into very narrow bands. Moreover, CDMA requires precise power control and code synchronisation management, which could be less flexible in terms of varying time allocation in some scenarios and in providing topology agnostic solutions for underwater networks. Finally, for the practical assumption, this thesis focuses on the distributed networks where node location

information is unknown and GPS and time synchronisation are not supported for each node. Therefore, SDMA is not of concern of this thesis.

2.2.3 Medium access control protocols

MAC protocols can be categorised in various ways. This thesis categorises MAC protocols into centralised and distributed protocols. Centralised networks usually have a central node to coordinate channel accesses in a network. The central nodes are responsible for determining a transmission order of nodes in a network, therefore centralised protocols can achieve good channel utilisation through collision-free centralised scheduling. Centralised protocols are more appropriate for static networks in which a coordinating node knows (or can gather) relevant information form the network nodes, for example, locations, transmission priorities, or traffic loads. Therefore, transmission scheduling can be relatively static and potentially pre-defined by a central node.

However, such information about all nodes and network configurations is not usually available beforehand for most applications and node locations are not constant in a mobile networks. Therefore, distributed protocols are necessary for networks where centralised scheduling is not feasible. However, significant additional overheads are incurred in distributed scheduling, for example to conduct neighbour discovery, to reserve channels/slots using a handshaking mechanism which a sender initiates whenever it starts a new transmission, or to sense a channel in order to help reduce the probability of collision. These signalling overheads of distributed protocols can impair channel utilisation in particular, when the propagation delay is significant such as in underwater networks.

2.2.3.1 Random access based protocols

Random access protocols are uncoordinated or employ minimal coordination, therefore it is more appropriate to distributed networks where there is a lack of centralised infrastructure. Random access based protocols allow nodes to decide when to transmit a data packet on the shared channel. If more than one node tries to send a data packet simultaneously on the shared channel, it results in a collision on the channel. This section reviews the history of distributed MAC protocols which commenced with a random access approach and developed towards channel reservation approaches.

2.2.3.1.1 Pure ALOHA

The simplest and earliest MAC protocol in wireless networks is called ALOHA. It was developed in the 1970's at the University of Hawaii [34]. The university was located on several islands, so a wireless network to exchange data between colleges was desired. Therefore, Norman Abramson and his team started a project to develop the wireless packet switched network based upon ALOHA, which is the first random access scheme. Using this scheme, each node accesses a channel as soon as they have a data packet to transmit. It is a feasible natural approach for distributed networks because individual nodes determine when to access the radio channel. The study [35] of ALOHA induces Equation (2-5) specifying the theoretical performance of the ALOHA system:

$$\text{Channel utilisation (U) of pure ALOHA} = G \bullet e^{-2G} \qquad (2\text{-}5)$$

In Equation (2-5), the highest theoretical performance value of ALOHA system is 0.18 (1/2e) Erlangs when the traffic load (G) is 0.5 Erlangs (as shown in Table 2-5) under assumptions:

- There are a large number of transmission nodes.
- Packet are generated according to a random Poisson arrival process [107] with average time between packets.
- All packets have the same length and same transmission time.
- At any instant in time, each node has no more than one packet to transmit.
- All lost packets are due to packet collisions.
- Any overlap in packet transmission times causes the complete packet to be lost.

The unit of Erlang corresponds to the fractional proportion of time during which a channel (e.g. telephone wire, radio channel) is active. 1 Erlang therefore corresponds to the fundamental capacity of a single channel.

2.2.3.1.2 Slotted-ALOHA

Slotted ALOHA is an extended version of pure ALOHA in which time is divided into slots. For the time slots, time synchronisation across all nodes in a network is required. Users randomly access the channel at the beginning of a fixed time slot and if more than two nodes select the same time slot, a collision occurs at the receiver (in the case where the propagation delays are very small as they are for typical terrestrial radio systems). Whilst packets collide with pure ALOHA

if they overlap even partially, with slotted ALOHA, packets either overlap completely or not at all. In other words, the slotted access can limit the probability of a collision within a slot and as a result, slotted ALOHA has a lower probability of a collision than pure ALOHA. The vulnerable period during which no other terminal should transmit in order to avoid collision at the receiver is reduced from twice the packet duration to exactly one packet duration as illustrated in Figure 2-7. Therefore, the maximum theoretical channel utilisation of slotted ALOHA is double that of pure ALOHA. ALOHA reaches 0.18 Erlangs of channel utilisation when the offered load is 0.5 Erlangs whilst slotted ALOHA reaches 0.36 Erlangs at 1 Erlang of offered load. Table 2-5 summaries the performance of pure ALOHA and slotted ALOHA.

$$\text{Channel utilisation (U) of slotted ALOHA} = G \bullet e^{-G} \quad (2\text{-}6)$$

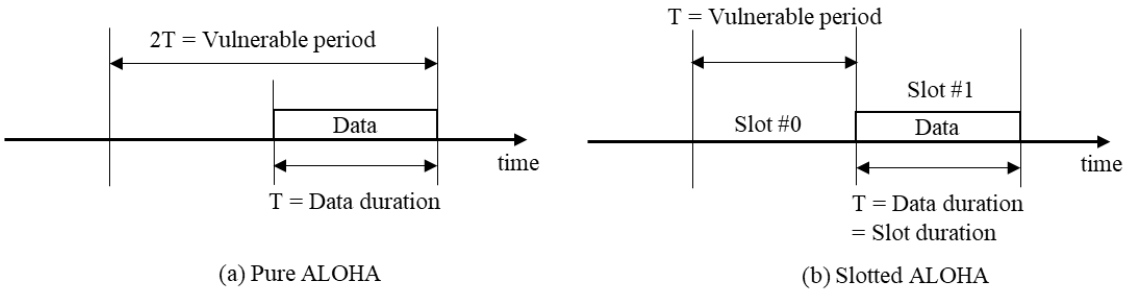


Figure 2-7. Vulnerable period

Protocol	Offered load (G)	Throughput rate to offered load	Max channel utilisation (U)	Offered load at the maximum U
Pure ALOHA	G	e^{-2G}	0.18	G = 0.5
Slotted ALOHA	G	e^{-G}	0.36	G = 1

Table 2-5. Comparison of channel utilisation

Figure 2-8 shows the pure ALOHA and slotted ALOHA performance with different propagation speeds of 3×10^8 m/s in the terrestrial environment and 1,500 m/s in the underwater environment. Note that Figure 2-8 is generated by the Riverbed modeler simulation tool in which pure ALOHA and slotted ALOHA in the terrestrial and underwater environments are implemented. MATLAB is used to plot the simulated and theoretical results. The Simulation tool, Riverbed modeler is discussed in the Appendices. The parameters of the simulation are presented in Table 2-6.

As Figure 2-8 (b) depicts, the benefit of slotted ALOHA (i.e. less time vulnerability) is lost by the effect of the slow propagation speed. The reason for this is that although a transmitter sends a

packet at the beginning of the time slot (synchronised), it will not arrive at the beginning of the time slot at a receiver (not synchronised) due to the long propagation delay when nodes are deployed with different propagation distances in a network. Therefore, the performance of slotted ALOHA shows the same performance level as pure ALOHA (0.18 Erlangs at G = 0.5) in the underwater environment.

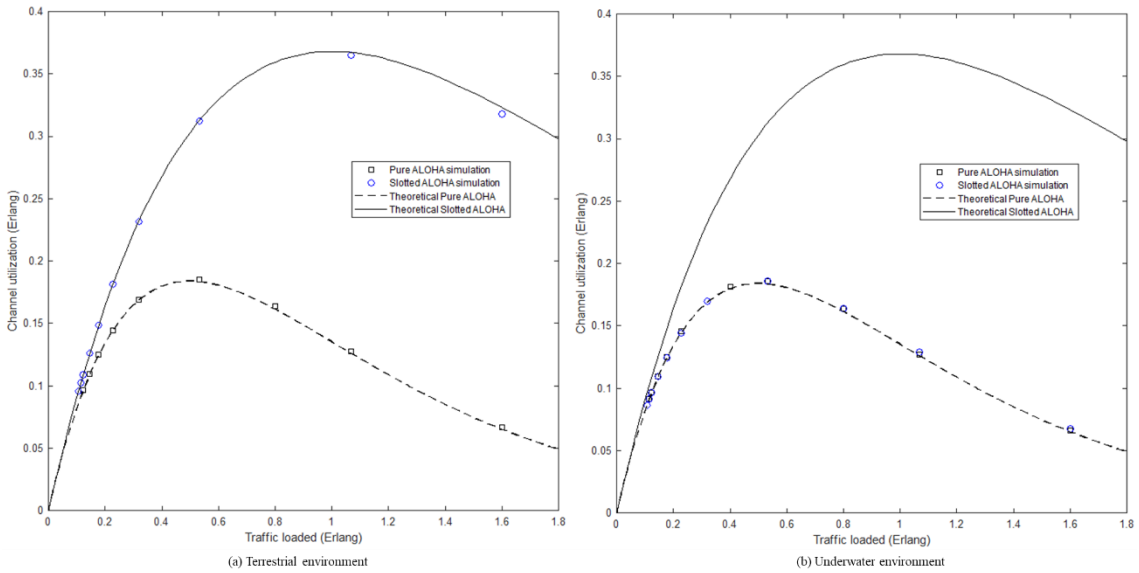


Figure 2-8. Channel utilisation of pure and slotted ALOHA in different environments

Simulation parameters	Value
Number of transmitting nodes	100 nodes
Number of receiving nodes	1 node
Distance from transmitters to receiver	Up to 100 m
Duration of simulations	5 hours
Results collected after	30 minutes
Channel bandwidth	1 kHz
Channel data rate	1,000 bps
Packet size	32 bits
Packet duration	0.032 seconds
Slot duration	0.0324 seconds
Packet size distribution	Constant

Table 2-6. Simulation attributes for Riverbed modeler

Slotted ALOHA is a baseline scheme of a protocol which this thesis proposes. Therefore, the purpose of Figure 2-8 is to provide the baseline performance and to validate the underlying simulation model of the protocol and the reception process with the binary collision model. The reason for selecting ALOHA approach for the baseline protocol rather than opportunistic networks is that the current underwater system is practically not the opportunistic system. The join/leave to/from the network is not as frequent as in terrestrial networks: this difference is discussed in section 2.3.4.3. Moreover, due to the high deploy cost as discussed in section 1.3.5, it is a more practical assumption that usual underwater sensor networks are not commonly opportunistic. Chapter 4 will provide comparison of the underlying random access techniques with the protocol on which the reinforcement learning techniques are built to validate the developed simulation models.

2.2.3.1.3 Framed Slotted ALOHA

Framed slotted ALOHA [36] was designed in 1977 for the satellite system. Frame slotted ALOHA adds the concept of frame to slotted ALOHA. Time is divided into repeating frames and slots. Each node randomly chooses one slot in a frame to transmit one data packet. Figure 2-9 shows an example of a framed slotted ALOHA network in which four nodes are deployed in a random topology network using four slots in a frame. In the example, four nodes are located within a certain size of a network (with radius, R) and they are time synchronised. Framed slotted ALOHA has been used for different applications, for example it is a primary protocol in Radio Frequency Identification (RFID) tag systems [37] because the low computational complexity of framed slotted ALOHA is appropriate to the very limited power requirement of the RFID system.

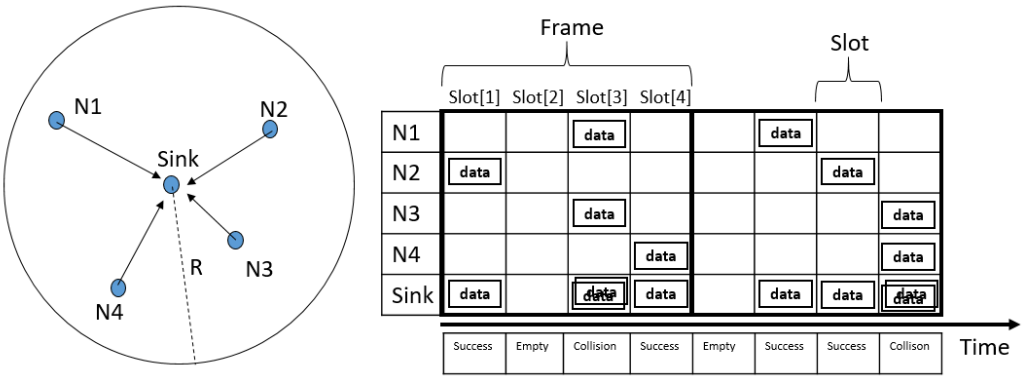


Figure 2-9. An Example of framed slotted ALOHA network and frame structure

The channel utilisation of framed slotted ALOHA can be calculated using Equation (2-7) [36]. We consider of N nodes into S slots. For a given time slot, the number of nodes allocated into the slot is a binominal distribution with N Bernoulli experiments and 1/S occupied probability. Under a condition that the frame size (S) is large enough (i.e. S >> 1), framed slotted ALOHA achieves the maximum channel utilisation of 0.36 Erlangs when the number of generating nodes (N) is equal to the frame size (S).

$$\text{Channel utilisation (U) of framed slotted ALOHA} = \frac{N}{S} \times \left(1 - \frac{1}{S}\right)^{(N-1)} \quad (2\text{-}7)$$

In Figure 2-9, a data transmission is successful only when one node transmits a data packet in a slot and all other nodes do not select the same slot. Since there is no means of coordinating the transmission order of generating nodes, collisions and empty slots occur regularly leading to unreliable and inefficient channel use.

2.2.3.1.4 CSMA

CSMA [38] was suggested in the 1970s by Leonard Kleinrock and Fouad Tobagi. CSMA stands for Carrier Sensing Multiple Access implying that each node senses the medium whenever it is ready to send a data packet. If the medium is sensed as busy, the node waits for a random time and retries (senses) again until the channel is sensed as idle. Therefore, carrier sensing can reduce the probability of collision with respect to the ALOHA schemes in the terrestrial environment. However, due to the long propagation delay in the underwater network, carrier sensing potentially requires a long guard time to sense the signals in the channel properly which deteriorates the achievable channel utilisation, so carrier sensing can be ineffective in underwater networks. Figure 2-10 compares the channel utilisation of pure ALOHA, slotted ALOHA, and CSMA in the terrestrial environment. Figure 2-10 is generated by the Riverbed modeler where pure ALOHA, slotted ALOHA, and CSMA in the terrestrial environment are implemented. MATLAB is used to plot the theoretical results.

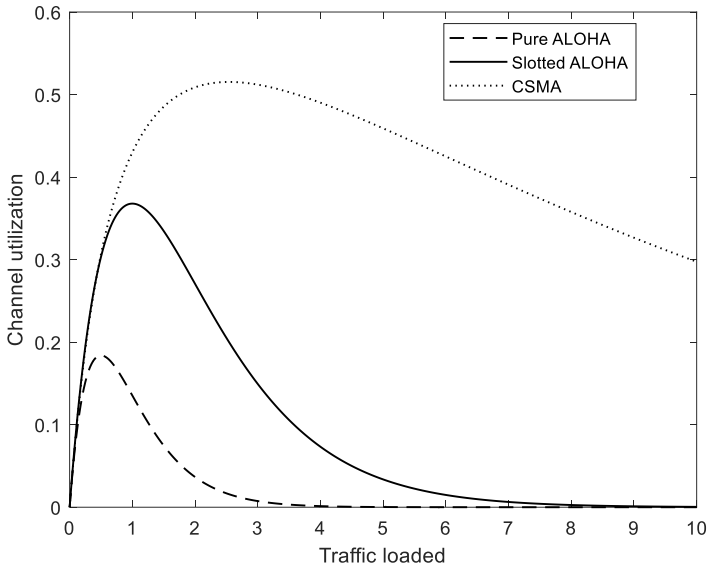


Figure 2-10. Channel utilisation comparison in the terrestrial environment

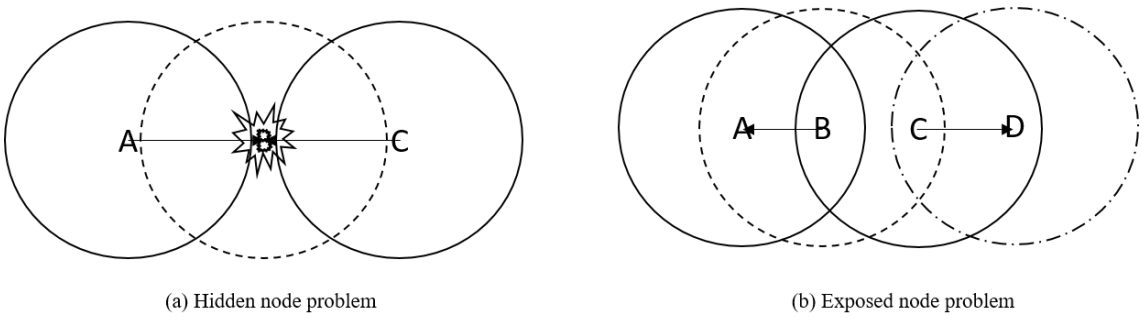
2.2.3.2 Reservation based protocols

The main feature of the random access approach is that the sender decides when to transmit a data packet, making them inherently distributed. The benefits of the random access approach is the simplicity of the protocols which means that they can be used in any type of distributed network. However, they achieve low channel utilisation due to the residual contention in the channel. Therefore, reservation based approaches are designed to avoid collisions in the channel. Instead of sending a data packet when a sender is ready to transmit, the sender reserves a channel first by exchanging small size control packets and then subsequently transmits one or more data packets. This section discusses some problems of CSMA and reviews a representative channel reservation protocol.

2.2.3.2.1 Hidden node problem and exposed node problem

Carrier sensing systems have problems known as the hidden node problem and the exposed node problem. In Figure 2-11 (a), nodes A and C want to send a packet to node B so node A and node C sense the channel before transmitting. The channel is sensed idle because they cannot hear each other, however there can be a collision at B (at the receiver) and this problem is called hidden node problem. In the case of Figure 2-11 (b), node B is sending a packet to node A and node C wants to send a data to node D. When node C senses the channel, the channel seems busy as node C senses the transmission from node B. Although a transmission from node C to node D would

not disturb the reception at node A, node C will defer the transmission, resulting in time being wasted in the use of a channel. This problem is called the exposed node problem.



(a) Hidden node problem (b) Exposed node problem

Figure 2-11. Hidden node problem and exposed node problem

2.2.3.2.2 Handshaking channel reservation

Channel reservation schemes employing handshaking attempt to solve the two problems described in section 2.2.3.2.1. Channel reservation uses short control packets or special control tones before transmitting data packets. The preceding reservation shall reduce the probability of collisions of data packets and thereby compensate for the additional traffic, delay, overheads, and complexity introduced through the use of control packets or tones.

Multiple Access with Collision Avoidance (MACA) [39] proposed a basic concept of handshaking in the 1990's. Instead of carrier sensing, the sender transmits a Request To Send (RTS) packet, and neighbours who hear the sender's RTS remain idle and avoid transmitting on the channel. As soon as the intended receiver receives the RTS, it sends a Clear to Send (CTS) packet, and neighbours who hear the receiver's RTS can also defer their transmissions. Therefore, the chances of a free channel between the sender and the receiver significantly increases and the sender sends a data packet to the receiver and the receiver replies with ACK if the packet is successfully accepted.

MACA was proposed for terrestrial communications and if MACA is used directly for an underwater network, there is the possibility of collisions in the environment due to the long propagation delay. Figure 2-12 shows an example of a possible collision condition if MACA is employed in the underwater environment.

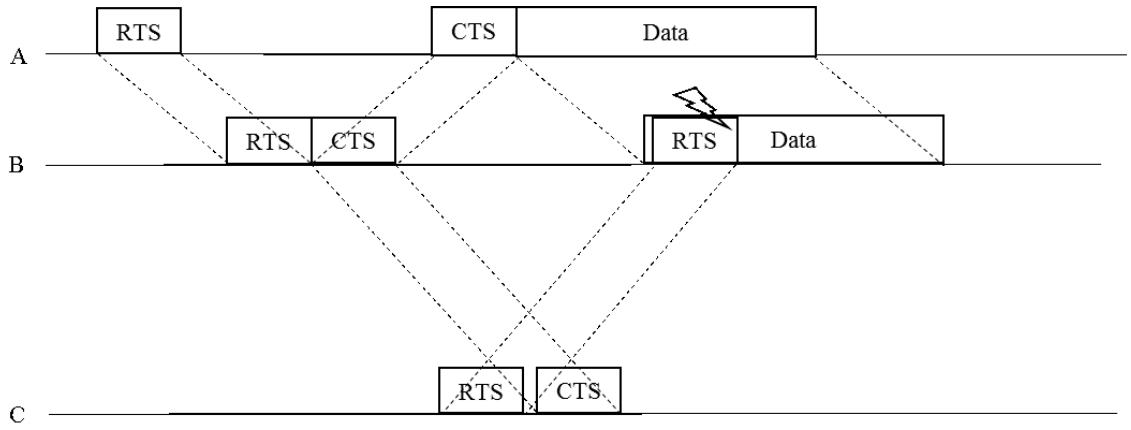


Figure 2-12. MACA problem in the underwater environment

In the example of Figure 2-12, nodes are deployed with different communication distances between them, which results in differing propagation delays. Node A and node C are senders and node B is a receiver. Node A is located closer to node B whilst node C is located further away. Node B sends a CTS responding to the RTS from node A than the transmission of an RTS from node C, however it is possible that node C cannot listen to the CTS due to the long propagation delay. Therefore, node C sends its RTS assuming the RTS does not disturb other transmissions because node C does not hear any RTS and CTS messages, therefore the RTS sent from node C can collide at the receiver node B. There are several studies [40-42] proposed to solve this problem in underwater networks, however their solutions are fundamentally based upon waiting for a longer time to receive control messages (i.e. RTC and CTS), which brings about poorer channel utilisation in the underwater environment.

2.2.4 Discussion

Random access and handshaking based protocols are reviewed in section 2.3.3. Pure ALOHA is the earliest and simplest protocol for wireless packet data transmissions. Slotted ALOHA is the extended version of pure ALOHA and it uses time synchronisation which leads to twice the channel utilisation by reducing the vulnerable period. Framed slotted ALOHA adds the concept of frames and it exhibits the best channel utilisation when the number of slots in a frame is equal to the number of nodes in a network. Carrier sensing is used to reduce collisions in the channel, however it leads to two problems which are called the hidden node problem and the exposed node problem. Therefore, the reservation approach is used to solve the problems and it also reduces the collisions by reserving the channel before data transmission by exchanging control message.

Random access and handshaking based protocols offer a great deal of flexibility such that they can be used for any type of distributed network and are relatively easily modified in terms of the available number of nodes in a network. However in the underwater environment, additional challenges arise due to the slow propagation speed. Carrier sensing requires sufficient guard band time which impairs channel utilisation. Moreover, frequent control packet exchange such as handshaking creates significant idle time in the channel due to the slow propagation speed so that the protocol performance becomes highly dependent on the propagation distance.

2.2.5 MAC protocols for underwater networks

Underwater networks are not limited to fixed deployments and nowadays mobile networks are emerging due to the increasing consideration of AUVs and Unmanned Underwater Vehicles (UUVs) for underwater exploration. The mobility of nodes brings high complexity in a protocol since the mobility factors such as movement patterns, speed, or directions need to be considered in designing a MAC protocol. Therefore, there have been studies [43-48] considering node mobility in underwater networks.

Location based TDMA Mobile MAC (LTM-MAC) [43] is an extension version of Location based TDMA MAC (LT-MAC) [49]. 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 delay, otherwise it cannot operate effectively. This represents a significant overhead with respect to channel utilisation.

Delay-aware Opportunistic Transmission Scheduling (DOTS) [44] is a distributed protocol which is designed primarily for fixed node deployments, but this paper in 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. However, the map quickly becomes out of data 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 [45] 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 [46] 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 [47] also uses RTS-CTS handshaking and assumes multi-channel use.

Asymmetric Propagation Delay aware TDMA (APD-TDMA) [48] is designed for AUV networks and is an extension of Transmit Delay Allocation – without time synchronisation MAC (TDA-MAC) [50] for fixed underwater networks. During the initialisation phase, a centralised node exchanges control packets with mobile nodes until the node obtains location estimates for all mobile nodes. During the transmission phase, the centralised node broadcasts a control message indicating the packet transmission schedule, then mobile nodes transmit data packets according to the timing indicated in the schedule packet. After receiving data packets from all nodes, the central node predicts the future locations of mobile nodes based on the packet reception times and broadcasts the updated control packet indicating the next transmission schedule. Whenever the number of data packet collisions at the sink node is greater than a certain level, the protocol conducts the initialisation phase to get the location information of sensor nodes. This prediction approach for future location of nodes is not appropriate to dynamic movements of mobile nodes since the dynamic changes raise frequent initialisation phases which can significantly reduce the overall channel utilisation, moreover location errors exist because mobile nodes move during the long initialisation phase.

2.2.6 Discussion

Most protocols [44-47] 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 [47], the required multi-channel operation is not easily realisable for underwater acoustic networks since the channel bandwidth is so limited, especially over longer distances. APD-TDMA [48] is a distinct protocol because it estimates the future locations of nodes, however it is not an efficient scheme when nodes moves at variable speeds or directions because it estimates the future locations of AUVs based on the latest data packet arrival time at the central node.

Most of existing protocols [43-48] for mobile underwater networks are extended versions of MAC protocols designed for networks comprising fixed nodes. They add extra functions such as frequent control message exchanges 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 underwater networks. Rather than these supplementary measures to deal with node mobility, the learning approach provides network adaptability, therefore can achieve good channel utilisation, low overheads, and low complexity in the face of changes in the network.

2.3 Reinforcement Learning based MAC

With the proliferation in demand to connect wireless networks, traditional and centralised systems cannot provide efficient solutions for problems such as resource management and mobility management in complex network configurations. As a key technique for enabling Artificial Intelligence (AI), machine learning is capable of solving complex problems. Motivated by its successful application to many practical tasks [51], both industry and academia have advocated the application of machine learning in wireless communication.

Machine learning is generally categorised into supervised learning, unsupervised learning, and reinforcement learning. Supervised learning and unsupervised learning require data sets for training and the outcomes (i.e. optimal strategy) is highly dependent on the data sets. However, agents in reinforcement learning learn through interaction with the environment, therefore the learning result depends on heuristic information obtained by trial-and-error experiences.

The key features of reinforcement learning are 1) it can potentially enable full self-organisation and high adaptability in distributed networks and 2) it does not require a priori knowledge of the operating environment as a model which can hardly be assumed to be available in practice for our purpose. This thesis concerns distributed protocols which provide medium access control without specific network topology or scenario limitations, so that it is desirable that nodes (i.e. learning agents) in a network are capable of learning to adapt to the environment through such interactions.

Therefore, reinforcement learning is appropriate to support the purpose of protocol design hence this section reviews reinforcement learning and how it is applied in MAC protocols in terrestrial and underwater environments.

2.3.1 Q-learning

Reinforcement learning can be categorised into model-based and model-free approaches. This thesis uses Q-learning, one method of model-free reinforcement learning because its approach matches the purpose of this thesis: to provide a scenario agnostic MAC protocol solution for distributed underwater networks. Therefore, model-free reinforcement learning is more appropriate because it does not require estimation of the operating environment model which is necessary for model-based reinforcement learning. Moreover, in the model-free reinforcement learning category, Q-learning is an off-policy method where optimal actions are based on the best possible value estimated through trial-and-error, whilst the learning agent of on-policy methods such as SARSA [52] needs to derive the policy function from trial-and-error experiences.

The fundamental reinforcement learning approach is designed by three functions: a policy function, a reward function, and a value function. The policy function maps the states of the operating environment to actions which need to be taken in the states. States and actions are explained in Equation (2-8). This function is developed through the experience of trying different actions in each state. However, Q-learning is a model-free reinforcement learning method, therefore the policy function is replaced by a Q-table which is estimated by rewards.

Each state-action pair receives a numerical reward which indicates its desirability. Calculating the reward for each state-action pair is handled by the reward function. The value function for Q-learning is a Q-function described in Equation (2-8) which maps each state-action pair to the total discounted sum of rewards. The reward function is relatively easy to design, since it is only concerned with immediate and explicit benefits of taking a certain action in a certain state whilst estimating value function (i.e. Q-function) requires prediction of the future of the system to some extent in order to design the function.

Each Q-learning agent updates a particular state-action pair at time *t* using the Q-function defined in Equation (2-8). The function enables agents to learn an optimal action through trial-and-error interaction in an environment and future actions are determined by prior experience [53].

$$Q(s_{t+1},a_{t+1}) \leftarrow Q(s_t,a_t) + \alpha\,[r_t(s_t,a_t) + r \max_{a_{t+1}} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)\,] \quad (2\text{-}8)$$

- $Q(s_t, a_t)$: the Q-value of the current state-action pair at time t
- $t \in$ T: decision epochs (t = 1, 2, 3…)
- $s_t \in$ S: the current state of the system
- $a_t \in$ A: the action taken in the current state
- $r_t\,(s_t, a_t) \in$ R: the numerical value from the reward function for the current action taken in the current state
- $\max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$: the maximum Q-value out of all actions in the next state s_{t+1}
- $0 \leq \alpha \leq 1$: the learning rate determines to what extent newly acquired information overrides old information. For example, in Equation (2-8), if the learning rate is 0, the agent learns nothing and exclusively exploits prior knowledge, $Q(s_t,a_t)$. However, if the learning rate is 1, the agent considers only the most recent information, $r_t(s_t,a_t) + r \max_{a'} Q(s_{t+1}, a_{t+1})$.
- $0 \leq r \leq 1$: the discount factor determines the importance of future rewards. If the discount factor is 0, the learning becomes myopic by only considering the current reward, $r_t(s_t,a_t)$, whilst the factor value of 1 leads to long term rewards.
- This is the case of a full exploitation policy learning. Exploitation will be discussed in section 2.3.1.3.

2.3.1.1 Convergence

Convergence is a characteristic of Q-learning. Convergence of Q-learning is mathematically proven in the Markov Decision Process (MDP) domain [54]. Figure 2-13 illustrates the concept of network convergence status. This figure is a simulation result which will be discussed in chapter 5. At the beginning of the scenario, learning agents (i.e. sensor nodes) learn the new environment where they are deployed and achieve steady state operation when the network performance graph becomes flat as shown in the figure. After convergence, there are three major changes in the scenario and the learning agents continue trial-and-error learning and hence can provide the steady (converged) performance in the scenario.

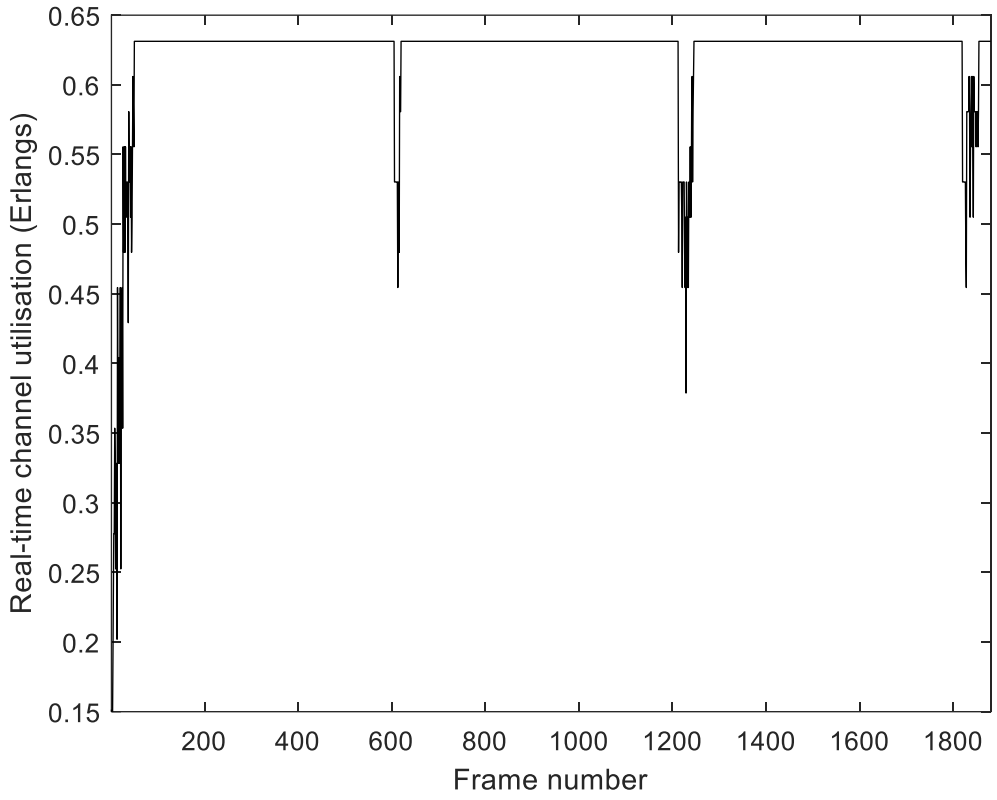


Figure 2-13. Learning and convergence

2.3.1.2 Learning speed

The learning speed primarily depends on the combination of a discount factor (r) and a learning rate (α) in Equation (2-8); there also are studies of algorithms offering rapid learning speed [55, 56]. A faster convergence speed benefits more from the reinforcement learning approach especially when the learning environment changes quickly. However, rapid learning increases levels of complexity in the learning algorithm. First, model-based reinforcement learning requires transition probability vectors of the operating environment since the learning speed can be improved. However, it increases the complexity of the algorithm and is also impractical for some networks where this type of information is unknown in priori. For the best learning performance with model-based reinforcement learning, each node needs to know the other user's state in each time epoch. Exchanging this information incurs significant communication overheads. Secondly, some studies [104, 105] conduct multiple learning in a single epoch to improve the learning speed. However, the complexity of solving a MDP is proportional to the cardinality of its state space S which increases exponentially with the number of nodes. To moderate this problem, the number of nodes in the network needs to decrease, however this solution is not practical.

2.3.1.3 Exploitation and exploration

Exploitation and exploration is another trade-off feature of reinforcement learning. Q-learning exploits the experienced knowledge, known as the greedy policy. However, a learning agent also needs to explore its environment, particularly crucial when the environment is not stationary. The balance between exploration and exploitation is an important factor in reinforcement learning. In each state, reinforcement learning faces a trade-off between exploration and exploitation according to the policy underlined in the below pseudo code.

Q-learning algorithm

Initialise Q-table arbitrarily
while the learning episode has not finished **do**
Detect present state s_t
while present state is not terminal **do**
Choose current action a according to action selection policy
Take this action a_t and observe the reward r_t and next state s_{t+1}
Update Q-table entry for current state-action pair using Equation (2-8)
Store the next state as the present state
end while
end while

Choosing a previously known action which guarantees the best reward amongst all other known actions is the greedy policy. In this greedy policy, the system is exploiting its current knowledge. On the other hand, exploration is choosing a previously unknown action which is likely to have a lower reward than the greedy action but there is also a probability of it being better and becoming the new greedy action. In this case the system is exploring new possibilities.

Table 2-7 compares well known policies used for Q-learning. Greedy selection implies that each agent always chooses the action with the highest Q-value (exploitation) whilst an e-greedy agent generates a random value between 0 and 1 (called ε) and then each agent selects a random action (explores) with the probability of ε, otherwise exploits with that of (1 - ε).

	Exploration	Exploitation
Greedy selection	No	Yes
ε -greedy selection	Probability of ε	Probability of 1- ε

Table 2-7. Balance policy between exploration and exploitation

2.3.2 Discussion

Section 2.3.1 has reviewed Q-learning and explained its main characteristics. Good adaptability, which is a characteristic of model-free and off-policy Q-learning, suits the scenario agnostic purpose of this thesis. Agnostic protocols aim to provide good performance in various distributed networks and scenarios rather than the best (optimised) performance in a specific network scenario or topology. Moreover, Q-learning is more computationally efficient than other reinforcement learning methods due to its simplicity. The next section will provide a summary of relevant literature on how reinforcement learning is applied to wireless communication networks in both terrestrial and underwater environments.

2.3.3 Reinforcement learning based approach for terrestrial networks

In the terrestrial environment, there have been various studies of reinforcement learning based protocols where each node independently acts as a learning agent in order to solve complex problems in the network, for example either improve the energy consumption, channel selection policy, or channel utilisation. The three main functions of reinforcement learning discussed in section 2.3.1 are designed in different ways according to application requirements in the literature. Relevant literature is summarised and reviewed under appropriate sub-headings which focus on the distinct factors of reinforcement learning such as the state, action, and reward function.

2.3.3.1 Energy consumption

This section introduces two reinforcement learning protocols designed for WSNs to improve energy efficiency. Q-Learning based MAC (QL-MAC) [57] is designed to reduce the level of energy consumption for WSNs. The protocol divides the time into frames and slots and reinforcement learning is used for each node (i.e. a learning agent) to decide whether it is better to be in an active or sleep mode during a slot to achieve low energy use.

The study suggests that the ideal protocol designed for energy efficiency should consider the network traffic conditions and then it can calculate the minimum active time to cover that data traffic. Therefore, the reward function of QL-MAC considers network traffic: not only the traffic of a single node, but also the traffic from neighbouring nodes. Hence, the reward function of QL-MAC includes a parameter of the total number of packets to which an agent is able to hear during one slot. The parameter is separated into the number of packets intended for the agent and to its neighbours and then different weights are applied to those two parameters in the reward function. QL-MAC considers two simulation scenarios: a grid topology and a random topology consisting of 16 sensor nodes. The simulation results show that QL-MAC achieves better energy efficiency

than T-MAC [58] and S-MAC [59]. Table 2-8 summarises the reinforcement learning features of QL-MAC.

Reinforcement Learning based MAC (RL-MAC) [60] is also designed for energy efficient WSNs. The agents learn the amount of active time (within a slot) during which each agent is active to receive data packets before it goes into the sleeping mode. Similar to QL-MAC, the amount of traffic is considered, but RL-MAC uses different parameters in the reward function such as the number of packets in a queue at the beginning of each frame. A single-hop star topology with 4 sensor nodes and a sink node, and a multi-hop chain topology with 10 sensor nodes are used for simulation and RL-MAC provides better energy efficiency than S-MAC [59].

The two protocols are designed to save energy by reinforcement learning but there are two main differences in the way the Q-function is designed. First, RL-MAC uses the basic (i.e. state-based) Q-learning discussed in section 2.3.1, which means the Q-value of the protocol is based on the reward function according to state-action pair. However, QL-MAC uses a stateless scheme (which will be discussed in section 3.2), so that the Q-value is updated directly by the reward value of the current action. Stateless reinforcement learning [61] is appropriate to the applications which need to weight instant learning, for example when the operating environment constantly changes as with underwater networks.

Second, QL-MAC does not has discounted rewards since QL-MAC is designed based on stateless Q-learning whilst RL-MAC uses discounted rewards, which implies that RL-MAC weights long term rewards in the learning process, considering the environment where the network is deployed. The Q-function of RL-MAC includes discounted rewards which is a zero or negative value in this paper. A negative reward is called punishment and the discounted punishment of RL-MAC is the discounted sum of the expected number of packets that have failed to be received in the next frame during the non-active duration based on the current action (i.e. the active time in a slot).

	QL MAC [57]	RL MAC [60]	MAC protocol with SARSA algorithm [62]
Medium access	Slotted ALOHA + CSMA/CA	Slotted ALOHA + carrier sensing	Non-persistent CSMA
State	Stateless	State based The number of packets queued for transmission at the beginning of a frame	State based Each agent (cognitive node)
Action	If Q-value is larger than a certain value, an agent (node) is active during the slot s	The active time in a frame	Select a channel to sense and send data transmissions
Learning rate	0.05	0.1	0.5
Reward function	The number of packets one agent can listen to and send in a slot	Depends on current active time, current state, buffer size, etc.	Success or fail result of sensing and transmitting
Current award	Yes	Yes	Yes
Discounted reward	No	Punishment Depends on the expected number of packets failed due to early sleeping	Q-value of previous action (based on channel that the user took during the previous epoch)
Discount factor	No	0.5	0.1
Aim	Active / sleep mode selection for a slot duration	Optimal active time in a frame	Optimal channel to sense and transmit
Performance enhancement	Energy consumption	Energy consumption	Throughput and delay

Table 2-8. Terrestrial reinforcement learning MAC protocols

2.3.3.2 Channel selection

There is a study [62] which applies reinforcement learning to multi-channel and distributed cognitive wireless networks for the purpose of optimal channel selection for data transmission. The considered scenario of the study has 40 channels and 50 cognitive users and each user (i.e. learning agent) conducts carrier sensing amongst the 40 channels to send data transmissions. The protocol applies reinforcement learning in two ways: 1) stateless Q-learning with the learning rate

0 and 2) State Action Reward State Action (SARSA) [52] which is a model-free reinforcement learning method. SARSA is an on-policy method and it is very similar to Q-learning. Q-learning chooses the best action in the next state, $r \max_{a_{t+1}} Q(s_{t+1},a_{t+1})$ for the counted reward (refer to Equation (2-8)) whilst SARSA uses the action actually chosen in the next state, $r\ Q(s_{t+1},a_{t+1})$. Simulation results of non-learning, stateless Q-learning, and SARSA learning are compared and results of reinforcement learning based protocols outperform non-learning protocols in terms of throughput and normalised delay since the learning schemes reduce the probability of collision in the networks.

2.3.4 Reinforcement learning based approach for underwater networks

We have discussed studies applying reinforcement learning schemes to protocols for different wireless networks in the terrestrial environment in the previous section 2.3.3. This section reviews reinforcement learning based protocols designed for underwater acoustic networks. Applying reinforcement learning for MAC protocols is a new research subject for underwater networks as the first related study was published in 2018 [63] by this author. Therefore, there has been limited research into underwater reinforcement learning based MAC protocols whilst more studies have been published for reinforcement learning based routing protocols [64-68] for underwater networks.

2.3.4.1 Reinforcement learning based cross layer protocols in the underwater environment

The majority of communication systems are traditionally designed based on layered architecture, typically based on the principles of the OSI reference model [20]. This layered structure reduces the complexity of design and it has allowed people to work predominantly on one layer with clearly defined interfaces linked to adjacent layers to make communication in design more manageable. However, in practice, communication systems are often designed with fewer layers such as the TCP/IP model [69]. There is some merit in cross layer design and being able to use information at one layer to inform decisions at another layer. There is one study of reinforcement learning based cross layer protocol for WSNs proposed in [70] for extending the lifetime of a network by reinforcement learning. Reinforcement learning is not used for medium access control in this study. The medium access is conducted by a basic slotted CSMA/CA scheme. Each slot is used for a learning epoch, but reinforcement learning is used to decide upon a suitable transmission power (physical layer), a transmission channel (data link layer), and the next neighbour to forward a data packet to (network layer).

This study uses a transition probability vector for learning processes and this type of reinforcement learning is categorised in model-based reinforcement learning which requires estimation of the operating environment in the form of a matrix [71]. Model-based learning improves the learning speed as the insight into the environment is explicitly built upon the knowledge from the transition probabilities as discussed in section 2.3.1.2. However, model-based reinforcement learning is appropriate for a specific network scenario and is less flexible than model-free learning methods.

This study focuses on energy saving, therefore the transmission power value is used as one factor in reward function. Table 2-9 summarises the reinforcement learning parameters of the study.

	Slotted CSMA/CA based reinforcement learning approach [70]
Medium access	Slotted ALOHA + CSMA/CA
State	A node related to packet p
Action	Transmission through a selected sub-channel
Learning rate	1/t (t = 1, 2, 3…)
Reward function	Transmission power factor, neighbour factor, and channel factor
Current reward	Yes
Discount factor	0.5
Aim	Optimal (less) transmission power, optimal sub-channel, and optimal (less number) relay node
Performance enhancement	Energy efficiency
Significant parameter	The number of sub channel and slot duration

Table 2-9. Underwater reinforcement learning approach for energy consumption

2.3.4.2 Discussion

This study assumes multi-channel communication. The lack of bandwidth in acoustic signals is the major challenge of underwater networks as we discussed in sections 1.3.4 and 2.1.3. A multi-channel system in this study shows improved energy consumption, however this protocol is limited to scenarios having a small number of nodes in a network due to the limited number of available channels. Moreover, this paper mentioned that due to the narrow sub-channel, the fragmentation process takes place frequently which leads to long end to end delays in the

underwater network. Therefore, this protocol is limited to use in a network where a small number of nodes is deployed, the application is highly delay tolerant, but requires improved channel utilisation.

Since collisions occur at the receiver, carrier sensing at the transmitter cannot completely eliminate collisions even though the sensed channel is detected to be idle in the underwater network. Due to the long propagation delay, carrier sensing potentially requires long guard times in order to sense the signals in the channel properly which reduces the achievable channel utilisation, otherwise carrier sensing is ineffective in the underwater network.

Every sensor node need to know its neighbours' location information by periodic control message exchanging. This neighbour discovery causes a significant reduction in channel efficiency due to the slow propagation speed of acoustic signals. Moreover, time synchronisation is assumed in this study thus the protocol becomes more complex since the time synchronisation techniques need to be employed.

2.3.4.3 Reinforcement learning based MAC protocols in the underwater environment

This section reviews reinforcement based MAC protocols in the underwater environment. This research subject is a very new subject and only 6 related papers have been found, with two of them published by this author which will be introduced in chapters 3 and 4. This section reviews the remaining studies which were recently published between 2018 and 2020. Reinforcement learning is used for medium access control, i.e. to coordinate scheduling of multiple nodes to efficiently use the shared channel. [70] was published in 2013 however the study uses reinforcement learning for optimal decisions in the physical layer, the data link layer, and the network layer rather than the MAC layer as discussed in section 2.3.4.1.

A conference paper [72] proposes a reinforcement learning based protocol for WSNs. It uses slotted ALOHA as a framework and assumes time synchronisation across the network. A large enough slot is divided into a data transmission phase and an ACK phase as Figure 2-14 shows.

Stateless Q-learning is used for this protocol, however the paper does not reveal the reinforcement learning related parameters such as the learning rate (α) or reward functions. Using reinforcement learning, each node learns the transmission order. Once the order is determined, the protocol omits the ACK phase and can increase channel utilisation up to 0.4 Erlangs. This protocol focuses on only an initialisation stage to decide the transmission order, but does not consider future changes

after the network converges. Therefore, the protocol is highly vulnerable to any future environmental changes in the network and is not appropriate for mobile sensor networks.

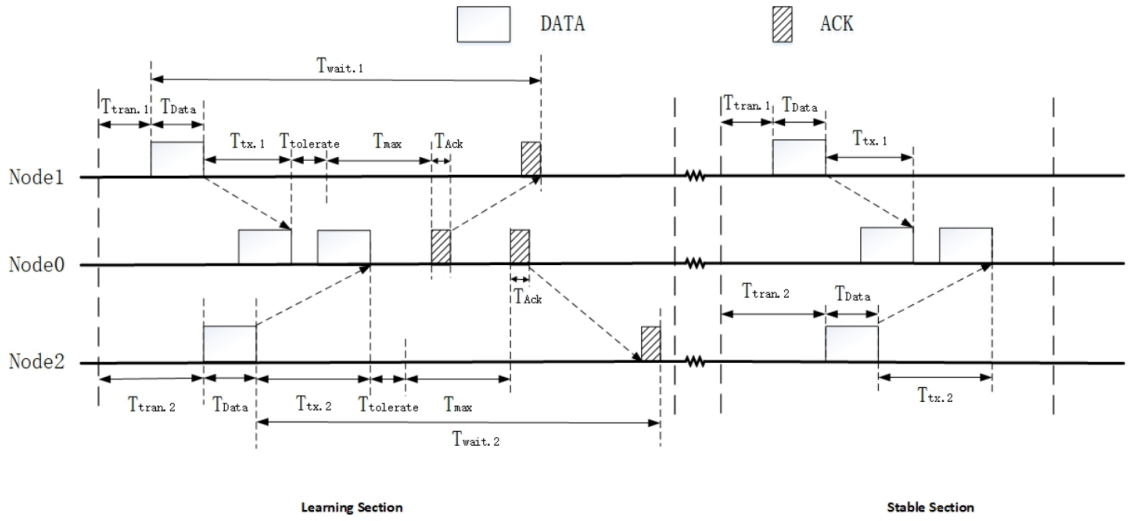


Figure 2-14. Reinforcement learning based MAC protocol for underwater networks: taken from [72]

The simulation scenario has five fixed sensor nodes and one sink node is deployed in the centre of them. The simulation results compare its performance with pure ALOHA and slotted ALOHA and the learning based protocol shows better performance in terms of channel utilisation and end to end delay.

Two conference papers [73, 74] were recently published in the international conference on Underwater Networks & Systems (WUWnet 2019) and both papers were inspired by one journal paper [75] which discusses Deep Reinforcement Learning (DRL) for heterogeneous wireless networks in the terrestrial environment. However, terrestrial and underwater environments are totally different, so problems arise if protocols designed for radio networks are directly applied to the underwater networks as discussed in chapters 1 and 2. These two papers [73, 74] are good examples of those problems.

The study [75] assumes location information, time synchronisation, high propagation speed, and heterogeneous and mobile devices in a radio terrestrial network. Moreover, the mobile nodes frequently trigger the join or leave processes to/from the network as Figure 2-15 shows. Therefore, the author justifies the reason for the use of Deep Neural Networks (DNN) under the complex network circumstances to approximate the Q-function since the state-action space becomes too large and complex.

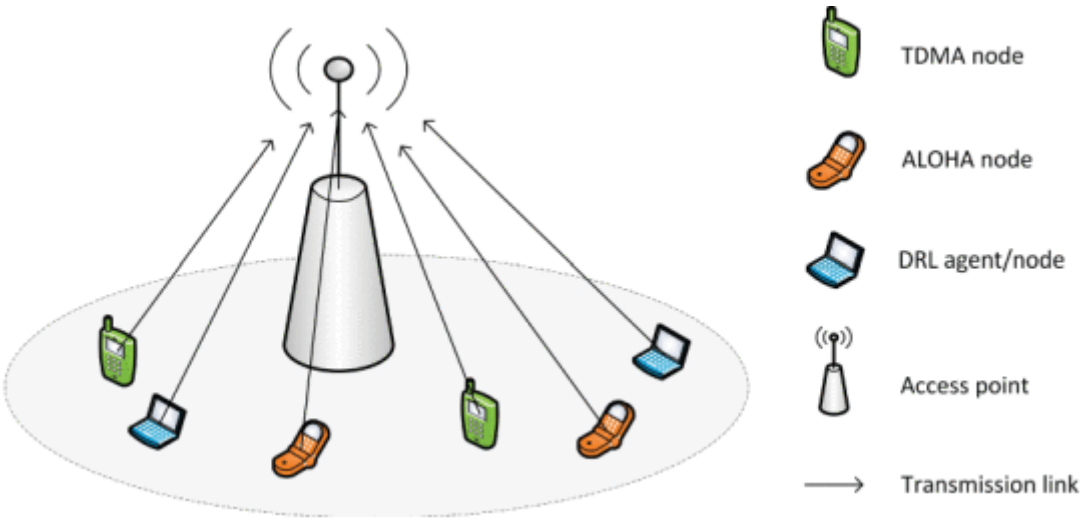


Figure 2-15. Heterogeneous multiple access system: taken from [75]

First of all, both conference papers [73, 74] require the locations of all nodes to be known before the data transmission phase. They obtain the information through control packet exchange, however the propagation speed is slow in the underwater environment, and therefore the channel efficiency of two protocols must be low. Moreover, to receive feedback (i.e. reward or punishment) from a sink node for the learning process, ACKs must be delivered to sensor nodes, and this impairs the performance of those protocols due to the slow propagation speed. To solve this problem, one conference paper [73] ignores the propagation delay of ACK transmissions from the sink node and simulation results show that channel utilisation becomes more than 1 Erlang which is not practical. Another conference paper [74] uses delayed ACKs for the current action to avoid the long propagation delay to get current feedback, which means the scope of the learning history for the current action does not include recent rewards. This workaround only works in fixed node networks, otherwise the delayed ACK approach cannot function in mobile node networks because the previous learning environment has been changed.

Secondly, both studies [73, 74] 1) assume only fixed nodes in the underwater network and 2) do not consider frequent join or leave device processes, which means network configuration is not as complex as in [75]. In the simpler learning environment, using DNN wastes computing resources since it is more efficient when it is used in complicated circumstances. [73] and [74] assume a small size of network for example, where less than 10 fixed nodes are deployed in simulation scenarios and protocols just need to learn the transmission order of the 10 nodes. Lastly, the two papers are basically time synchronised. As we discussed earlier in section 1.3.3,

maintaining time synchronisation challenging in the underwater environment and incurs a lot of overheads.

The most recent study [86] has been published in September 2020. It proposes the use of deep learning for channel selection in a multi-channel underwater system. For the simulation, two fixed nodes are deployed in a single-hop network where a sink node is located in the centre. The protocol uses the slotted ALOHA structure and assumes time synchronisation. The two sensor nodes select one channel among the available three channels in every slot and transmit a data packet. The simulation compares results from 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 during the simulation. The learning approach does not achieve the best throughput at the beginning of the simulation, however it catches up the optimised throughput after sufficient iterations of learning. The study also wastes computing resources for deep leaning in that it merely selects one channel in a slot. Moreover, the acoustic channel is very limited so that the multi-channel system is not ideal for underwater communications.

2.3.4.4 Discussion

The recent studies applying reinforcement learning to the MAC problem for WSNs have been reviewed. These essentially provide some capability to learn appropriate medium access from multiple nodes in the underwater environment using reinforcement learning. However, they still rely on time synchronisation or are designed for very specific purposes.

First, the studies do not deal with the slow propagation speed appropriately. Deleting the feedback phase, ignoring the feedback propagation delay, or excluding very recent rewards causes problems such that networks become vulnerable to any changes in the operating environment.

Secondly, reinforcement learning is not efficiently designed in the existing studies [27-24] [86]. The major benefit of reinforcement learning is that it provides flexibility and adaptability through trial-and-error experiences. The protocols in those studies are not capable of adapting to any future changes once networks are converged.. To improve the underwater channel utilisation, those studies 1) remove feedback (ACKs), 2) ignore the propagation delay of feedback (ACKs), or 3) exclude the most feedback (ACKs), which implies their design does not utilise the benefit of reinforcement learning in an effective way for further changes after network convergence.

Moreover, applying DNN for a relatively simple decision wastes the computational recourses and increases the complexity of the protocols.

Moreover, all existing reinforcement learning based protocols designed for underwater networks consider networks comprising fixed nodes. Reinforcement learning is potentially effective in a mobile networks since it provides inherent adaptability based on continued interaction with an environment. With regard to underwater networks comprised of mobile nodes, we cannot seek convergence and the learning process will continue with oscillation as nodes move. However, the benefit of the reinforcement learning approach is that it still can achieve better performance than non-learning schemes. The effectiveness of such an approach boils down to whether the learning algorithm is able to adapt at a sufficient speed with respect to key environmental changes.

Lastly, the prior studies relay on time synchronisation. Given the challenges of time synchronisation underwater due to lack of GPS and navigation difficulties for long term sensor node deployment, it is interesting to try to develop an asynchronous protocol that will still offer a reasonable QoS.

Based on these observations from literature, this thesis presents new work in chapter 4 that has is geared towards achieving good medium access using reinforcement learning without the need for synchronisation in underwater networks consisting of fixed sensor nodes. Reinforcement learning is more efficient when it responds to changes in networks. Therefore it can be utilised at the MAC layers for mobile networks. Thus, chapter 5 proposes a reinforcement learning based protocol for mobile underwater sensor networks without the need for time synchronisation.

3 ALOHA-Q in terrestrial and underwater environments

3.1 ALOHA-Q

ALOHA-Q [76] is a reinforcement learning based MAC protocol which is designed for Wireless Personal Area Networks (WPANs) in a terrestrial environment consisting of fixed sensor nodes. The main idea of the study is applying reinforcement learning for each node (agent) to learn the preferred slot in a frame and to send a data packet in framed slotted ALOHA structure as depicted in Figure 3-1. The major benefit of ALOHA-Q is that it can achieve a very high channel utilisation (up to 0.95 Erlangs) in the terrestrial environment without any form of centralised scheduling. Channel access starts as framed slotted ALOHA, but nodes are able to learn to avoid each other on the channel and a scheduled outcome can be achieved through the fully distributed reinforcement learning process.

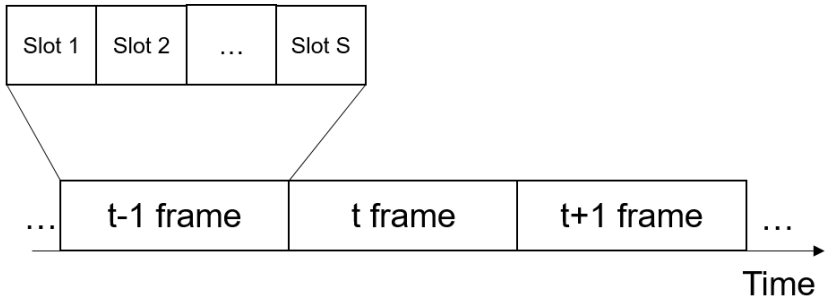


Figure 3-1. ALOHA-Q frame and slot flow in time

An extended version of ALOHA-Q [77] was published for energy efficiency and a new scheme called Informed Receiving (IR) is added to ALOHA-Q. IR estimates the expected number of epochs that a node will continue to use the same slot in a frame based on the known information being embedded by the transmitter in the data packet. One of the assumptions of ALOHA-Q is that the number of nodes deployed in a network (N) and the number of slots in a frame (S, called frame size) are known. ALOHA-Q sets the number of slots per frame (S) to equal the number of sensor nodes (N) for the best throughput (as framed slotted ALOHA in section 2.2.3.1.3), which is reasonable for a single-hop system. However for more complicated deployments for example, when the number of nodes within interfering range is unknown or nodes are successively deployed in a more ad-hoc fashion, it would be difficult to predict the appropriate frame size (S) for the network. Using too many slots in a frame wastes channel capacity because each node of ALOHA-Q is designed to use one slot per frame. On the other hand, using too few slots results in residual contention and cannot converge in these fixed systems.

The key benefit of ALOHA-Q is providing very high channel utilisation, however, it is reliant on knowing the number of nodes (N) to be deployed in a network. Therefore [78] developed a mechanism to additionally allow learning of the frame size (S). This study specifically looks at using reinforcement learning to adapt the frame size (S) rather than assuming a fixed frame structure which cannot be known in advance. Further works [79] were undertaken to do some practical experiments to demonstrate ALOHA-Q in hardware and proposed solutions for the practical issues for example by improving the reward factors of the Q-function. Note that [78] and [79] comes from the same research group of the author.

ALOHA-Q is the fundamental protocol used to propose a new underwater protocol in this thesis, hence details will be discussed in this section. ALOHA-Q uses framed slotted ALOHA (refer to the section 2.2.3.1.2) as a framework, therefore all nodes of ALOHA-Q are time synchronised. Time is divided by repeating frames and slots in the same manner as framed slotted ALOHA does. Figure 3-1 shows the frame and slot structure of ALOHA-Q.

The data transmission flow of ALOHA-Q is shown in Figure 3-2 where a frame consists of two slots, i.e. frame size (S) = 2. One slot duration (T_s) is sufficient to accommodate a data packet (T_{dp}), propagation delay (τ_p), an ACK (T_a), and a guard time (T_g). A slot duration (T_s) can be calculated by Equation (3-1). An individual slot is designed to support the transmission of a data packet to a receiver and to receive the ACK back. To achieve this, the slot duration (T_s) needs to account for the maximum propagation delay from a fixed sensor node to a receiver in both directions. This small guard band (T_g) is merely for the case that the maximum delay is underestimated and to account for potential clock drift.

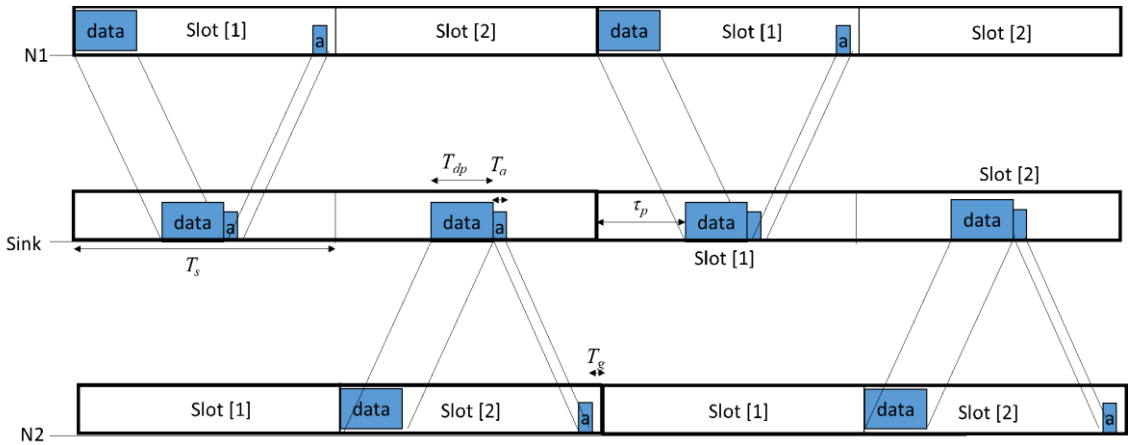


Figure 3-2. An example of ALOHA-Q when frame size and the number of nodes are two

$$T_s = (T_{dp} + T_a + T_g) + 2 \times \tau_p \quad (3\text{-}1)$$

ALOHA-Q uses ACKs for two reasons. ALOHA-Q aims to provide reliable data transmission, therefore ACKs are used to ensure the data packet is delivered. 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 retransmission must be initiated. Moreover, the ACKs are used to determine the reward or punishment in the Q-learning process of ALOHA-Q.

Therefore, reserving enough time to receive ACKs in a slot are important for ALOHA-Q and it is not a problem for ALOHA-Q in the terrestrial environment since the propagation delays are very small. The most significant element which impacts on slot duration (T_s) is the propagation delay (τ_p) in Equation (3-1). Therefore, in the underwater environment, the network needs long slots to accommodate the long propagation delay which introduces a lot of idle time to the slots and results in low channel utilisation in the underwater network, as shown in Figure 3-2.

3.2 Stateless Q-learning

Stateless Q-learning [61] is used in the ALOHA-Q protocol. Stateless Q-learning was proposed for some problems where an environment does not have to be represented by state. The learning agents are stateless and only the action space and a one-dimensional Q-table is considered. The job of reinforcement learning becomes simpler and the aim of stateless Q-learning is to estimate an expected value (Q-value) of a single reward for each action available to the learning agent.

The advantage of stateless Q-learning is the significant reduction in the number of Q-values that need to be estimated by the learning agent. Therefore there is a potentially dramatic reduction in the number of trials needed for it to learn an optimal action. Such a significant increase in the learning speed directly translates into the higher adaptability of reinforcement learning.

The principle of ALOHA-Q is a slotted structure in time. Please note that ALOHA-Q uses time synchronisation since it is easy to obtain in the terrestrial environment. Nodes make decisions on which slot to pick at the start of each frame. The baseline scheme is one of random selection, essentially a framed version of slotted ALOHA. However, more intelligent selection of a slot can be carried out based on Q-learning. ALOHA-Q uses stateless Q-learning at each node as a means of determining which slot should be selected by maintaining weight values, one per slot, updated every frame based on the experience in the last frame. The principle is for nodes to select a transmission slot having a largest Q-value based on the stored Q-value in the Q-table. All nodes

have a Q-table which contains the individual Q-value for each slot in a frame as Figure 3-3 shows. In this example, four different sensor nodes (called N1 to N4 respectively) are deployed at random locations within communications range of the sink node. Four nodes collect data and transmit the collected information to a sink node which is located in the centre of a network. The standard implementation of ALOHA-Q uses frame size (S) equal to the number of nodes deployed in a network (N). For example, Figure 3-3 uses four slots in a frame (frame size, S=4) for the four nodes in a network (N=4). Therefore, one node has a chance to transmit collected data once in a frame and needs to select one slot in a frame to transmit a data packet.

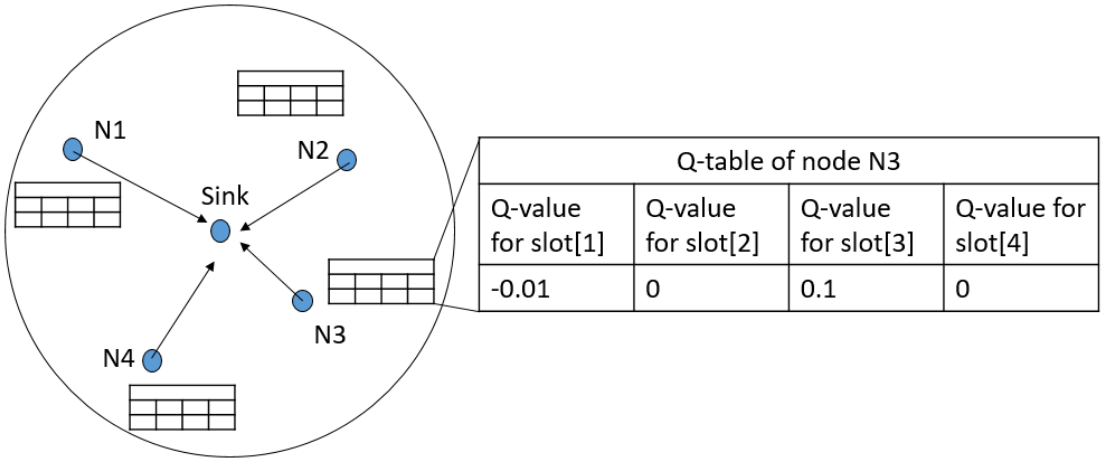


Figure 3-3. Example Q-table for ALOHA-Q for a four node network

All Q-values are initialised to 0 so it is initially random access, however Equation (3-2) defines how the weights are updated based on that initial trial. If a node transmits in a slot that it has picked, and it successfully gets a positive reward, it subsequently increases the weight value for that slot so it will then pick the same slot in the next frame. Alternatively, if the node does not get ACK, it receives a negative reward and the weight associated with that slot will then become negative such that the node will then pick one of the other slots (those with a weight value of 0) at random in the next frame. This process continues according to Equation (3-2) which defines the stateless Q-learning process used to determine how Q-values are updated in ALOHA-Q when the *i*th node has sent a data packet in the *s*th slot in a frame:

$$Q_{t+1}(i, s) = Q_t(i, s) + \alpha\,(r - Q_t(i, s)) \tag{3-2}$$

where, Q_t is the Q-value at time t, t is a time epoch (i.e. a frame), α is a learning rate, and r is reward. A standard implementation of ALOHA-Q [76] uses $\alpha = 0.1$ and $r = 1$ if the transmission

is successful, (i.e. a transmitting node successfully receives an ACK) otherwise, r = -1. More detailed information is described in Equation (2-8).

Considering the example in Figure 3-4 in more detail, since all Q-values in the Q-table are initially zero in Figure 3-4, a 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 through the application of Equation (3-2). 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.

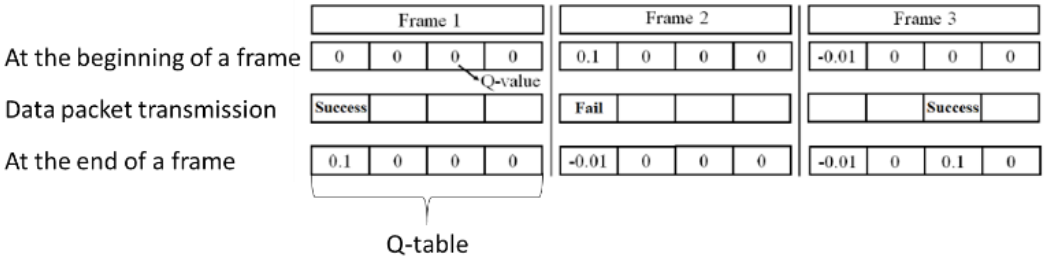


Figure 3-4. Example of Q-table update process of N3 for the first three frames

At the start of the second frame, the node transmits 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 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.

Table 3-1 summarises reinforcement learning features of ALOHA-Q. Agents (i.e. sensor nodes) learn a distinct slot in a frame, hence the protocol can reduce collisions based on its historical learning experiences.

	ALOHA-Q [76]
Medium Access	Framed slotted ALOHA
State	Stateless
Action	Select a slot in a frame
Learning rate	0.1
Reward function	Reward: 1 (Transmission success) Punishment: -1 (Transmission failure)
Current reward	Yes
Discounted reward	No
Discount factor	No
Aim	Slot allocation
Performance enhancement	Channel efficiency
Significant parameter	Slot duration (T_s)

Table 3-1. ALOHA-Q learning summary

3.3 ALOHA-Q in the terrestrial environment

Figure 3-5 and Table 3-2 shows the status when an ALOHA-Q network converges: the node transmission order has been determined by learning and therefore, there are no collisions or empty slots at the sink node (compared to Figure 2-9). Node 3 uses the third slot in Figure 3-5 because the Q-value of the third slot has the highest Q-value in its Q-table (Table 3-2). Before the convergence, the node 3 experience the changes of Q-value as Figure 3-4 describes for the first three frames at the beginning. Then the Q-value for the third slot increases as 0.1 → 0.19 → 0.271 → 0.3439 → 0.40951 → 0.468559 → 0.5217031 → 0.56953279 → 0.612579511 → 0.65132156 → 0.686189404 → 0.717570464 → 0.745813417 → 0.771232075 → 0.794108868 → 0.814697981 → 0.833228183 → 0.849905365 → 0.864914828 → 0.878423345 at every successful transmission based on Equation (3-2). Figure 4-8 will show this increase in graph at the learning rate of 0.1. Please note that ACK transmissions are omitted in Figure 3-5.

This author implemented ALOHA-Q using Riverbed Modeler for validation purposes and the typical parameters of ALOHA-Q used for the implementation of ALOHA-Q shown in Table 3-3. Channel utilisation is measured during the simulation at the sink node using Equation (3-3):

$$\text{Channel Utilisation (U) when a network is not converged} = \frac{D}{r \times T} \quad (3\text{-}3)$$

where, D is the total amount of data in bits received by the sink node, r is the data rate in bps in Table 3-3, and T is the total period of time in seconds during which to calculate the channel utilisation (U). ALOHA-Q achieves 0.95 Erlangs of channel utilisation and the remaining 0.05 Erlangs accounts for the fixed overheads in the frame (i.e. T_a and T_g).

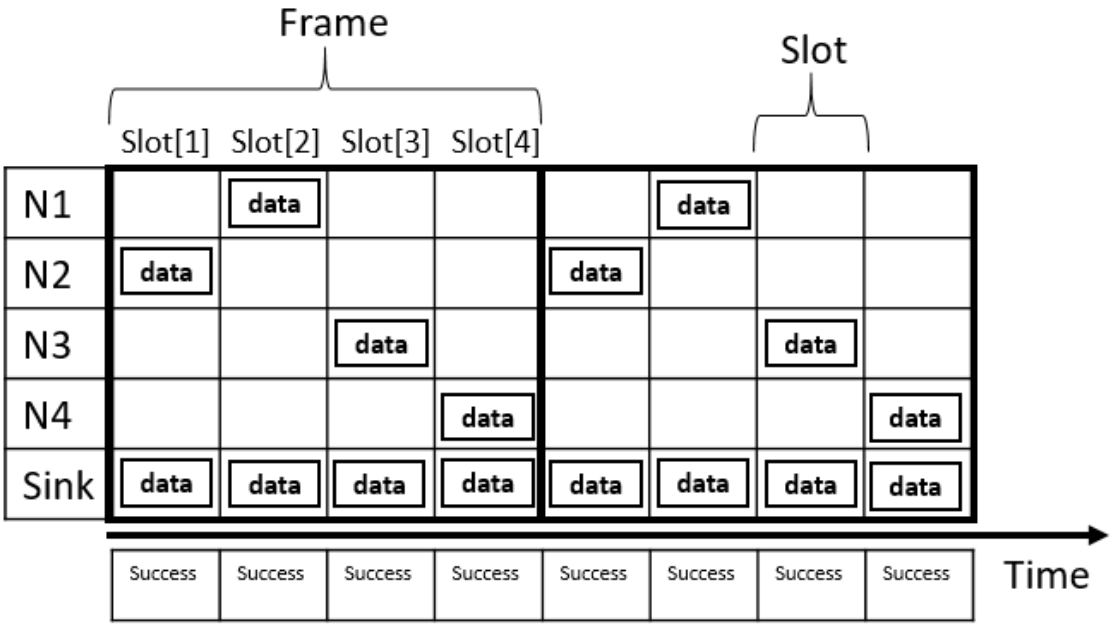


Figure 3-5. Concept of ALOHA-Q when network converges

Slot1	Slot2	Slot3	Slot4
-0.01	0	0.878	0

Table 3-2. Example of the Q-table of node 3

Assuming the propagation delay (τ_p) is negligible, the expected theoretical maximum channel utilisation of ALOHA-Q can be calculated by Equation (3-4) which means that the network achieves the steady state (converged) and hence channel utilisation of one slot can represent the overall channel utilisation of ALOHA-Q under convergence since the identical slot is repeated at the sink node as shown in Figure 3-5.

$$\text{Theoretical channel utilisation (U) of ALOHA-Q when network converges} = \frac{T_{dp}}{T_s} \quad (3\text{-}4)$$

Parameters	Value
Duration of a data packet of 1044 bits (T_{dp})	4.176 ms
Duration of an ACK of 20 bits (T_a)	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
Network size in radius (R)	12.9 m
Tx and Rx data rate (r_{tr})	250,000 bps
The number of generating nodes in a network (N)	50 nodes
Propagation speed (v_{tr})	3×10^8 m/s
Propagation delay (τ_p)	Negligible
Topology	Single-hop, star topology

Table 3-3. Typical ALOHA-Q parameters for terrestrial use

3.4 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. Moreover, ALOHA-Q uses stateless Q-learning which provides high adaptability and simplicity and this characteristic of ALOHA-Q matches that of the scenario agnostic protocols. Therefore it is of interest to explore the possibility that ALOHA-Q can be used in the continuously changing underwater environment. However, this section will show that ALOHA-Q provides low channel utilisation in the underwater environment because the design of ALOHA-Q did not consider the characteristics of underwater acoustic communications.

An initial simulation of ALOHA-Q in terrestrial and underwater networks has been undertaken using Riverbed Modeler. The purpose of the initial simulation is to compare the performance of ALOHA-Q in both terrestrial and underwater environments. The considered network comprises 50 fixed sensor nodes in a single-hop star 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 same as does the standard ALOHA-Q [76]. Table 3-4 shows the simulation parameters used for ALOHA-Q in the underwater environment.

The identical simulation parameters in Table 3-3 are used in so far as possible, but 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 [80] is considered.

In order to initially demonstrate and validate ALOHA-Q implementation in Riverbed Modeler and also to provide some initial comparisons between simulation results of ALOHA-Q in terrestrial and underwater environments, only key parameters related to the underwater environments (i.e. r_{uw} and v_{uw}) have been changed as mentioned above. Therefore, not all parameters are realistic for a practical underwater deployment, for example 12.9 m network size (R) in Table 3-4, which is taken and exactly the same as published by [76]. Beyond this initial comparison, realistic parameters are used for underwater network simulations.

Parameters	Value
Duration of a data packet of 1044 bits (T_{dp})	16.704 ms
Duration of an ACK of 20 bits (T_a)	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
Network size in radius (R)	12.9 m
Tx and Rx data rate (r_{uw})	62,500 bps
The number of generating nodes in a network (N)	50 nodes
Propagation speed (v_{uw})	1,500 m/s
Propagation delay (τ_p)	8.6 ms
Topology	Single-hop, star topology

Table 3-4. Typical ALOHA-Q parameters for underwater use

The result of the simulation is that ALOHA-Q can be operated in the underwater environment but the protocol only provides a channel utilisation of 0.48 Erlangs when network converges, much lower than the 0.95 Erlangs which achieved by the same protocol within a terrestrial environment when the network converges [76]. Channel utilisation is measured at the sink node using Equation (3-3)

The slow propagation speed of acoustic signals is the primary cause for low channel utilisation since it makes the duration of slot (T_s) much greater in the underwater networks. Equation (3-1) shows the calculation for the duration of a slot (T_s) and the propagation delay (τ_p) is the significant element. During the propagation of the data packet and ACK, the channel remains in an idle state which consequently causes a decrease in achievable channel utilisation. Therefore, chapters 4 and 5 propose new methods to improve the performance of ALOHA-Q in the underwater environment.

4 UW-ALOHA-Q for fixed underwater sensor networks

This chapter proposes the new protocol UW-ALOHA-Q [81] which to the best knowledge of the author, is the first reinforcement based MAC protocol for the underwater environment. Three improvements are proposed in this chapter to address the limitation of low channel utilisation which was discussed in section 3.4. The new protocol: 1) includes asynchronous operation to eliminate the challenges associated with time synchronisation under water; 2) offers an increase in channel utilisation through refinement of the frame size; 3) achieves collision free scheduling by incorporating a new random back-off scheme. UW-ALOHA-Q is discussed in detail in this chapter. Note that UW-ALOHA-Q assumes that 1) the number of nodes deployed in the network (N) and 2) the network distance (R) in terms of radius in meters are known in advance.

4.1 Asynchronous operation

The cost of synchronisation is considerable in the underwater environment, especially for long term deployments, because GPS signals are not available and distributed synchronisation introduces notable overheads and complexity. Therefore, this thesis considers the situation where all nodes are not time synchronised. This section describes how the asynchronous operation can be applied to ALOHA-Q for the purpose of underwater communication.

In the terrestrial environment, asynchronous operation reduces the channel utilisation in networks because of the vulnerable period as shown in Figure 2-7. To avoid collisions, a packet being received should not overlap with any others at the receiving node for twice the duration of one slot (Figure 2-7(a)). However, in the time synchronised network in Figure 2-7(b), the vulnerable period decreases by half due to the benefit of synchronisation and the probability of collisions in the network is halved. The difference the throughput between pure ALOHA (0.18 Erlangs) and slotted ALOHA (0.36 Erlangs) shows this clearly in section 2.2.3.1.

Figure 4-1 compares the simplified structures of (a) the standard (synchronised) ALOHA-Q protocol and (b) the protocol where asynchronous operation is applied. All nodes in the synchronised ALOHA-Q network use a global time reference. Using global time synchronisation in the network, the protocol can reduce the probability of collisions by reducing the vulnerable period. In terms of the underwater environment, however, global time synchronisation is difficult and costly to achieve. Therefore, this section considers asynchronous operation as depicted in Figure 4-1(b). Each node has its local clock, so it calculates its own slot and frame times. However, because its clock is not globally synchronised, each node's frame starts at different times and each

node works independently, which means that there is no need for nodes to know their neighbours' clock information. Nodes are assumed to start the frame at a uniform randomly distributed time within the range of zero to the duration of one frame.

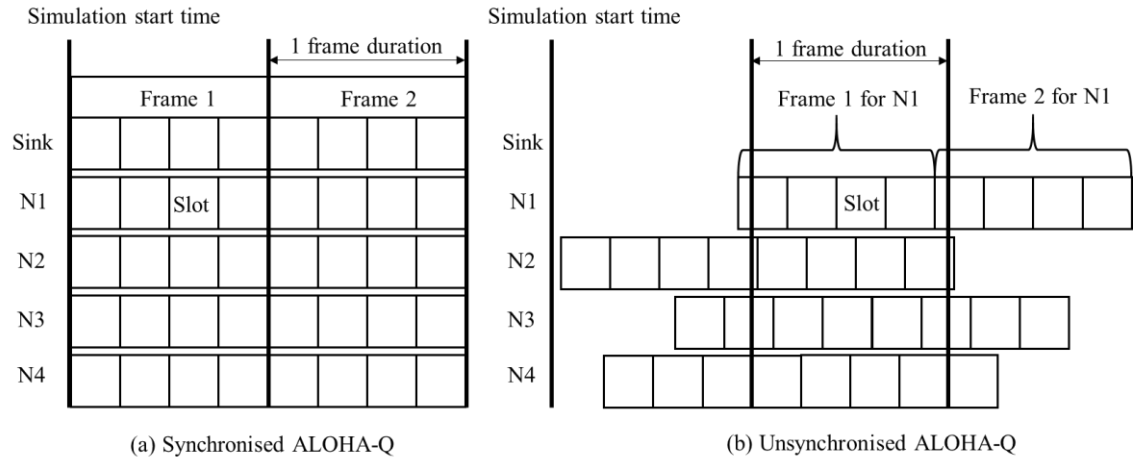


Figure 4-1. ALOHA-Q with and without time synchronisation

Figure 4-2 shows the flow of the asynchronous operation in UW-ALOHA-Q. The two generating nodes N1 and node N2 start their frame in different moments and a sink node does not need to work within a frame structure compared to Figure 3-2. UW-ALOHA-Q does not use duty-cycle and is in a half-duplex mode.

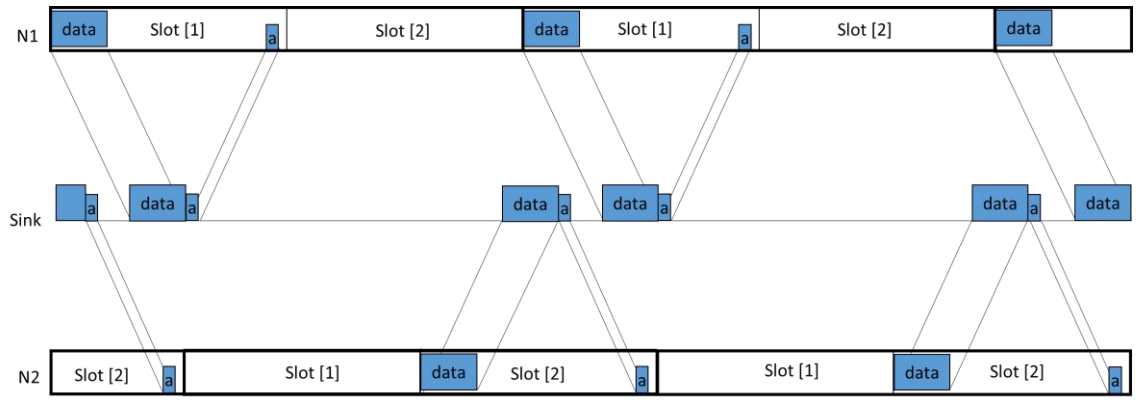


Figure 4-2. Data transmission process with asynchronous operation

Figure 4-3 compares the difference in reception patterns of data packets at sink nodes of the ALOHA-Q protocol in the two different environments. In Figure 4-3(a), the terrestrial set-up with minimal propagation delay is efficient when synchronised, which is also easy to achieve. Therefore, data packets arrive close in time to each other at the sink node and channel utilisation is high. However, if asynchronous, the system does not work well due to the increased vulnerable

period as discussed in section 2.2.3.1.2. In the underwater environment, however, the long propagation delays result in long slots and significant idle time (Figure 4-3(b)). This makes the system inefficient when synchronised so channel utilisation becomes low. When asynchronous operation is considered, however, the idle time is sufficient to avoid overlapping reception, so the protocol is not prone to experiencing collisions. Even if packets overlap at the receiving sink node, nodes can learn and find the distinct slot number in the frame by reinforcement learning to avoid such overlap.

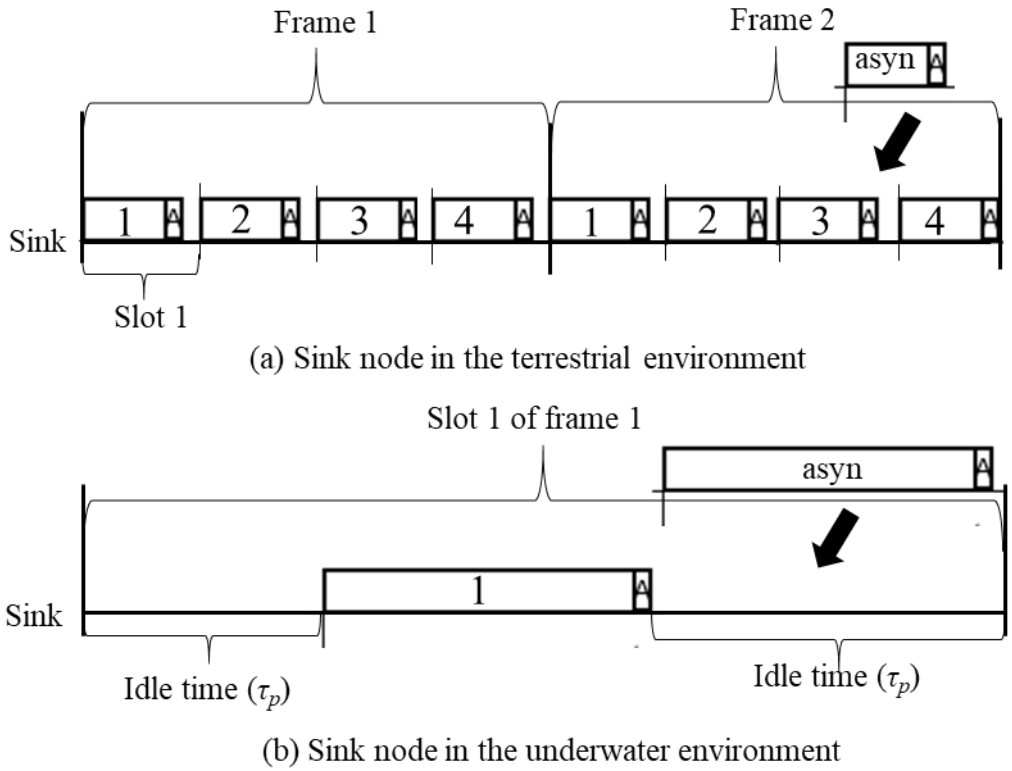


Figure 4-3. Slot reception at the sink node in two different environments

Figure 4-4 shows an example of the way that reinforcement learning enables nodes to avoid collisions without time synchronisation in an UW-ALOHA-Q network. Four generating nodes are deployed in a star topology and one sink node is located in the centre of the star topology. The standard ALOHA-Q protocol uses framed slotted ALOHA as a baseline and hence ALOHA-Q uses the number of slots in a frame equal to the number of nodes deployed in a network for the maximum channel utilisation as discussed in section 2.2.3.1.3. In this example, four sensor nodes are deployed in a network so one frame principally consists of four slots. Four nodes have to choose one slot among slot1, slot2, slot3, or slot4 for their transmission in each frame. They are not synchronised, so the frame start timing for each node is different. In the first frame, node 1

randomly chooses slot2 and transmits a data packet in the slot, node 2 in slot1, node 3 in slot3, and node 4 in slot2. At the sink node, packets from node 1 and node 2 overlap with each other and therefore collide in their first frame. The two nodes do not therefore receive an acknowledgement (ACK) from the sink node. As a result, the Q-values of the slots in the Q-table are updated to negative values (i.e. -0.1 according to the Equation (3-2)), thus the two nodes change slot numbers for the next transmission, based on the operation of the Q-learning algorithm and the slot selection policy with the highest weigh value at the start of each frame. If more than two Q-values are identical and highest in the Q-table, the node randomly selects one slot as explained in section 3.2 in particular in the example in Figure 3-4. Hence node 1 chooses the slot1 and node 2 chooses the slot2.The new order does not incur overlapping data packets from node 1 and node 2 anymore. On the other hand, node 3 and node 4 continue to use the same slots that they used for their initial transmissions, because they were successful. By repeating the learning scheme, the four nodes are able to learn which slot number they need to use and finally all four packets can arrive at the sink node without interfering with transmissions from other nodes in the network.

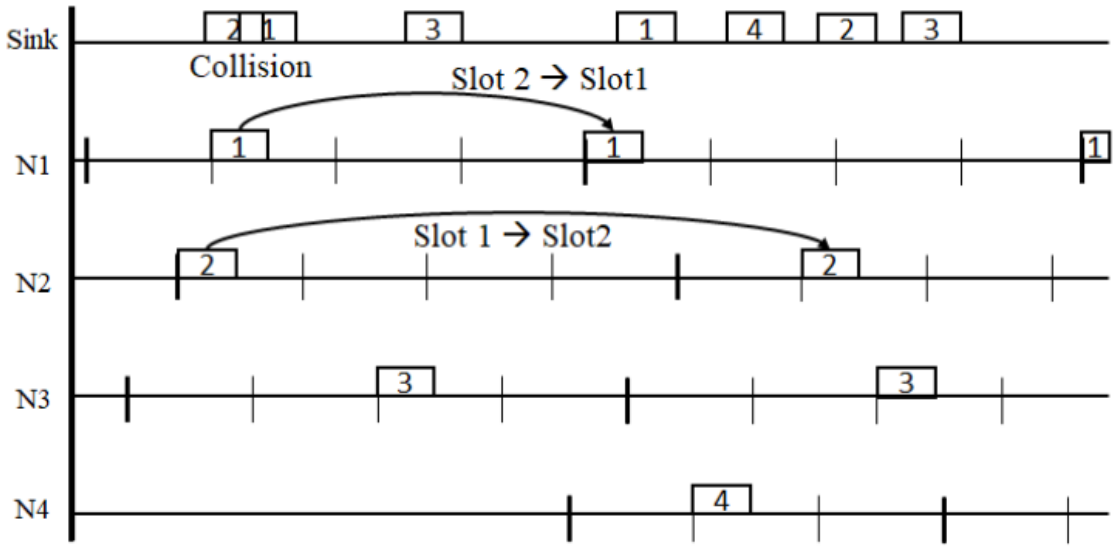


Figure 4-4. Asynchronous operation for UW-ALOHA-Q in the underwater environment

4.2 Discussion

The UW-ALOHA-Q protocol does not necessarily experience any reduction in channel utilisation despite nodes operating in an asynchronous fashion in the underwater environment. Utilising the idle time at the sink node caused by the slow propagation speed which is depicted in Figure 4-2 and Figure 4-3, UW-ALOHA-Q can still achieve collision free reception by reinforcement

learning as illustrated in Figure 4-4. The idle time at the sink node is sufficient to avoid overlapping reception and reinforcement learning allows nodes to find the distinct slot which can fill in the gap (the idle time at the sink node) so that a data packet can be successfully received by the sink node. Table 4-1 compares the relevant utilisation of the standard ALOHA-Q protocol with and without time synchronisation in the two different environments. The second column in the table (i.e. With time synchronisation) summarises simulation results in sections 3.3 and 3.4 and related parameters are defined in Table 3-4 and Table 3-4. ALOHA-Q in the terrestrial network can achieve very high utilisation (0.95 Erlangs) due to negligible propagation delays whereas the underwater network cannot (0.48 Erlangs) because of the slow propagation speed which means that the network has large slots to accommodate a data packet, an ACK and large idle times.

Unit: Erlangs	With time synchronisation	Without time synchronisation	Ratio
Terrestrial	0.95	0.64	≈ 0.67
Underwater	0.48	0.48	1

Table 4-1. ALOHA-Q and UW-ALOHA-Q channel utilisation in different environments with and without time synchronisation

Now this thesis looks at and considers the channel utilisation that is achievable without time synchronisation. Using the same simulation settings and configurations in section 3.3 and 3.4, only the asynchronous operation is applied. Asynchronous operation reduces channel utilisation of the ALOHA-Q protocol in the terrestrial environment because the vulnerable period increases and the probability of collisions at the sink node thereby increases as discussed in section 2.2.3.1.2. The channel utilisation should be half of the synchronised ALOHA-Q, however, the learning scheme of the protocol increases the channel utilisation (Table 4-1: 0.95 vs 0.64). On the other hand, asynchronous operation in the underwater environment does not result in any loss of channel utilisation (Table 4-1: 0.48 vs 0.48).

However, the ALOHA-Q protocol in the underwater environment continues to have a lower utilisation (Table 4-1: 0.64 vs. 0.48). Therefore, a new scheme for refining the frame size (S) in a network is proposed in the next section 4.3 to increase the channel utilisation in underwater networks.

4.3 Reduced frame size

The standard ALOHA-Q protocol shows the best performance when the frame size (S) and the number of generating sensor nodes (N) in the network are the same in the terrestrial environment as discussed in section 2.2.3.1.3. However, if the frame size (S) is identical to the number of nodes (N) in the underwater environment, the sink node has significant idle time (Figure 4-3(b)) caused by the long propagation delays (τ_p). To increase the channel utilisation, this thesis proposes a new scheme which fills the spare space (in time at the sink node) by reducing the number of slots per frame (S).

Figure 4-5 shows an illustrative example of how convergence could be achieved, corresponding to all nodes finding transmission times which avoid any overlap/collision at the sink node. In this example, two slots are used per frame to support four generating nodes (the optimal and converged case in this example). Comparing Figure 4-4 and Figure 4-5, it is obvious that channel utilisation is better when a small number of slots is used because the amount of idle time at the sink node is decreased. For this collision free reception to achievable in practice, it is additionally important for nodes to be able to adjust their (initially random) frame start times and this will be discussed in section 4.6.

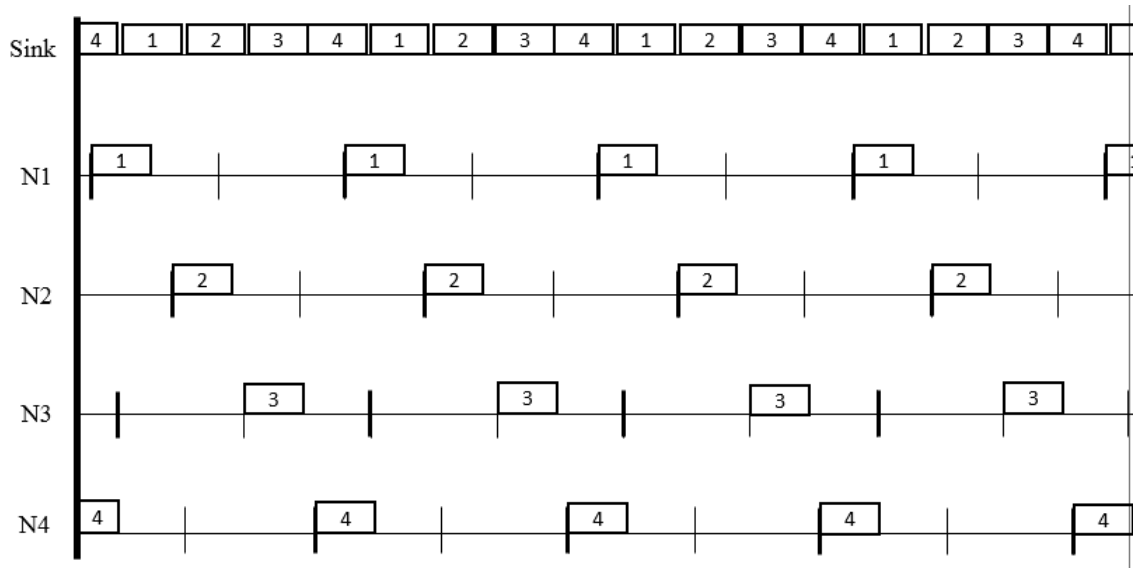


Figure 4-5. Reduced frame size (S) for UW-ALOHA-Q to improve channel utilisation

4.4 Simulation

This simulation is based on the full set of parameters shown in Table 3-4 and with channel utilisation measured by Equation (3-3). In addition, the asynchronous operation is added. Figure 4-6 compares the channel utilisation between the standard synchronised ALOHA-Q and UW-

ALOHA-Q protocols depending on the frame size (S) in the underwater network. As a baseline scheme, the performance of framed slotted ALOHA is also shown for comparison.

The standard (synchronised) ALOHA-Q protocol shows the highest channel utilisation when the frame size (S) is the same as the number of nodes in the network (N = 50). As the frame size (S) is reduced from 50, the utilisation also decreases because there are then insufficient slots for individual nodes to have an independent slot, therefore there has to be residual contention and associated retransmissions. However, the UW-ALOHA-Q protocol can achieve its highest channel utilisation when the frame size (S) is 38, with 50 generating nodes in the underwater network, because asynchronous operation and reinforcement learning make the protocol able to utilise the idle time more effectively at the sink node.

Compared to the baseline protocol (i.e. framed slotted ALOHA), UW-ALOHA-Q shows significant improved channel utilisation. This is a key message of this thesis that, in the identical network configuration, applying the learning approach has great potential to improve the channel utilisation for underwater networks.

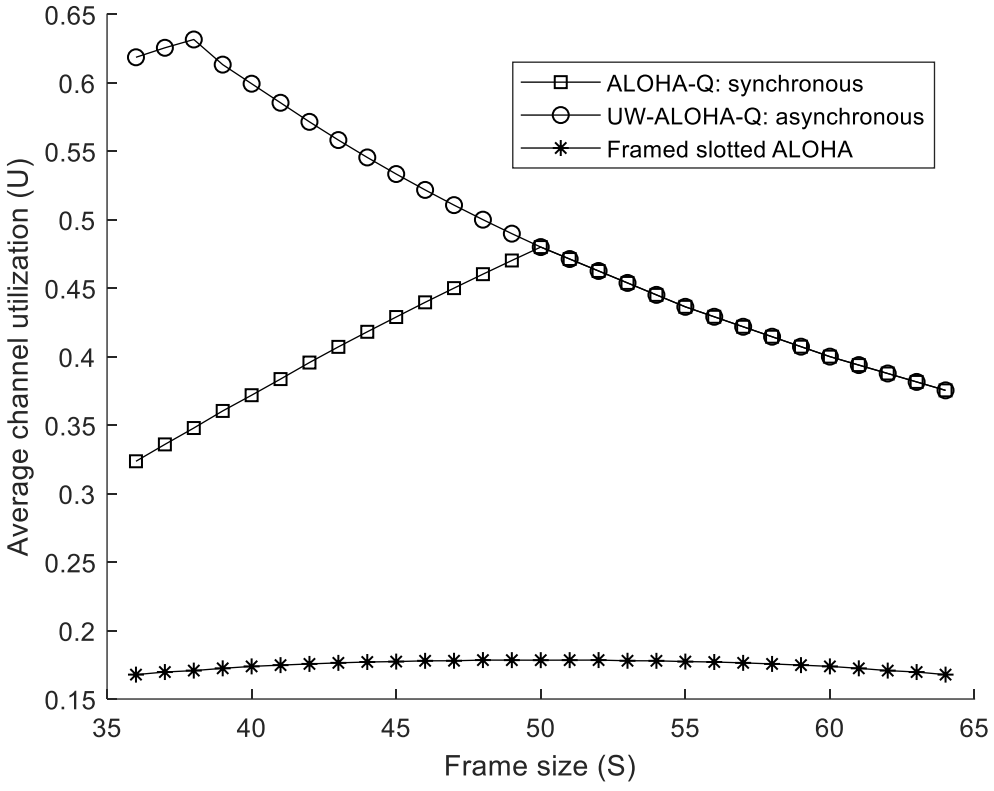


Figure 4-6. Channel utilisation according to frame size (S) with 50 sensor nodes

An appropriate frame size (S) of UW-ALOHA-Q can be determined for a network depending on the application requirements. With the parameters given in Table 3-4, the data packet corresponds to approximately half a slot in duration, such that around half the capacity is unused. Mathematically the frame size (S) can be reduced by up to N/2, where N is the number of nodes in the underwater network. N/2 comes from the ratio of ≈ 1:1 between the packet duration (T_{dp} = 16.704 ms) and the idle time in one slot ($2 \times \tau_p$ = 17.2 ms) in one time slot (T_s = 34.8 ms). Therefore, a frame size (S) between [N/2 and N] is feasible in the UW-ALOHA-Q protocol.

When more slots in a frame are used, channel capacity is wasted because each node of ALOHA-Q and UW-ALOHA-Q is designed to transmit a data packet in one slot per frame. Therefore the channel utilisation of three protocols decreases when the frame size (S) is larger than the number of nodes (N = 50) as seen in Figure 4-6.

4.5 Discussion

By refining the frame size (S), UW-ALOHA-QM can achieve a higher channel utilisation than the standard ALOHA-Q protocol in the underwater environment, and a similar level of utilisation when asynchronous operation is applied (for example, 0.63 Erlangs in Figure 4-6). Despite good channel utilisation one significant problem occurs and it is almost impossible for UW-ALOHA-Q to achieve network convergence and collision free scheduling in a network where fixed nodes are deployed. The reason of the problem is a reduced frame size (S) since a possibility arises that the network cannot converge due to the randomly inherited frame start time which cannot be changed. This limits the use of the protocol because it cannot be used for applications which require a good level of QoS regardless of the high channel utilisation in average achieved. Therefore, it is necessary to design an underwater protocol which can allow network convergence in a fixed environment. The next section will discuss this problem deeply and propose a new scheme which can provide the network convergence for underwater networks comprising fixed or pseudo static sensor nodes.

4.6 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 frame size (S), it is highly probable 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 [82] 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 UW-ALOHA-Q with the two proposed improvements in the underwater environment, the possibility of non-convergence continues to exist since the structure of frames and slots still uses a fixed frame start time. Therefore, a uniform random back-off scheme is proposed in this section. This scheme operates independently from the slot learning process (which is described in section 3.2) and allows the nodes to adapt their frame start times. Using this scheme, subsequent to 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 the preferred slot to use in successive frames.

Operation of the proposed uniform random back-off scheme is illustrated in Figure 4-7. This new random back-off scheme technically provides chances for node to adjust their transmission timing and find a gap for successful transmission. The two sensor nodes in this example start their frames at different times due to the asynchronous operation. Node 1 sends a data packet in slot1 of a frame X and node 2 in slot 2 of its frame X-1. The two packets collide at the sink node. The two nodes therefore invoke the uniform random back-off scheme which adjusts their frame times. Node 1 and node 2 do not change their slot number as it is assumed that the currently selected slots still retain the highest Q-values in the Q-table, despite the collision. This explains how the learning process and the back-off scheme work independently. After moving their frame time, node 1 and node 2 can find an appropriate frame start time to transmit their data packet successfully in frame X+1 and in frame X respectively. Node 2 will not adjust its frame time for frame X+1 since the node should receive an ACK successfully from the sink node in frame X. For the same reason, it is predicted that node 1 maintains its current frame cycle for the next frame X+2.To understand Figure 4-7 easier, it is helpful to point out the weakness of UW-ALOHA-Q: 1) the number of nodes in the network (N) and 2) the network size in terms of the radius in meters (R) need to be known in advance. Therefore, future work 6.3.7 will discuss the frame adaptation scheme.

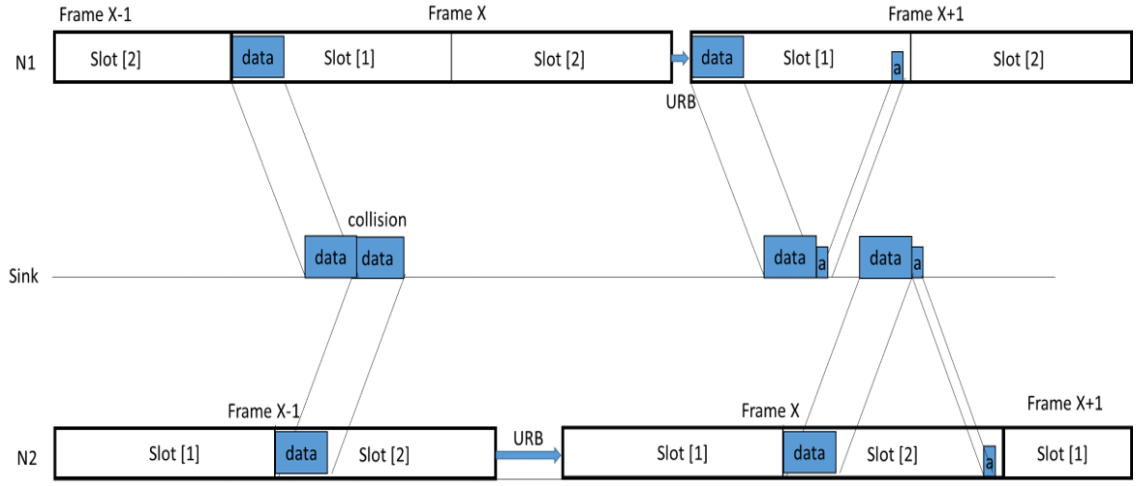


Figure 4-7. Uniform random back-off scheme for UW-ALOHA-Q

Inclusion of this scheme leads to collision free scheduling and permits convergence in UW-ALOHA-Q underwater acoustic networks comprising fixed nodes under the assumption that any underestimation of maximum delay can be covered by the guard duration (T_g).

This section 4.6 and Figure 4-7 does not discuss simulation results. This section illustrates the concept of the uniform random back-off scheme. Therefore, there are no specific parameters to be described for Figure 4-7. Nodes of UW-ALOHA-Q are designed to send one data packet per frame, which means that the periodicity of the data transmission always equals the frame duration: $T_s \times S$. Convergence will be discussed in the future sections 4.8.2 in particular Table 4-3 and Table 4-4. Section 4.8.2 shows that the number of slots (S) in a frame significantly impacts on convergence. As the number of slots per frame increases, the chance of network convergence increases and vice versa. In summary, the proposed reinforcement learning based UW-ALOHA-Q scheme can achieve high channel utilisation and network convergence using very low overheads without the need of time synchronisation and any centralised controller in the underwater environment. The next section 4.7 will provide more detailed analysis and investigation of Q-values. Following this, the simulations in section 4.8 will demonstrate the behaviour of UW-ALOHA-Q with different network configurations and topologies and serve to validate the envisaged channel utilisation capability of the protocol.

4.7 Impact of reduced frame size (S) on Q-value

For simulations in this thesis, ALOHA-Q and UW-ALOHA-Q systems are implemented in Riverbed Modeler. The Q-value significantly impacts on network convergence measurement for

simulation results. Therefore, it is important to understand and predict the trend of Q-value changes during the operation of learning based MAC protocols.

Previous studies of ALOHA-Q [83] for terrestrial radio networks use a frame size (S) equal to the number of nodes (N), which is appropriate when trying to achieve maximum channel utilisation in single-hop wireless networks. When the frame size (S) is equal to the number of generating nodes (N) in a network, once a Q-value of one slot becomes negative at the first data transmission attempt (based on Equation (3-2)), the slot is unlikely to be selected again in a greedy learning policy. In this case, a sufficiently large Q-value in a Q-table can be a criterion of simulations to judge nodal convergence in the implemented ALOHA-Q system. Therefore, the simulation of ALOHA-Q [83] makes a judgement as to whether the implemented system is converged when all nodes maintain the use of a particular slot for at least 20 consecutive frames.

The way in which the Q-values of ALOHA-Q are adjusted is shown Figure 4-8 where the Q-value rapidly rises because the number of slots (S) is sufficient for the number of nodes (N). Using Equation (3-2), the Q-value reaches 0.878 after 20 consecutive successful transmissions with the learning rate (α) = 0.1 and the Q-value is reasonable enough to judge nodal convergence for the ALOHA-Q implementation. The previous study [84] carried out the analysis of Q-value in a radio wireless network and shows the same result.

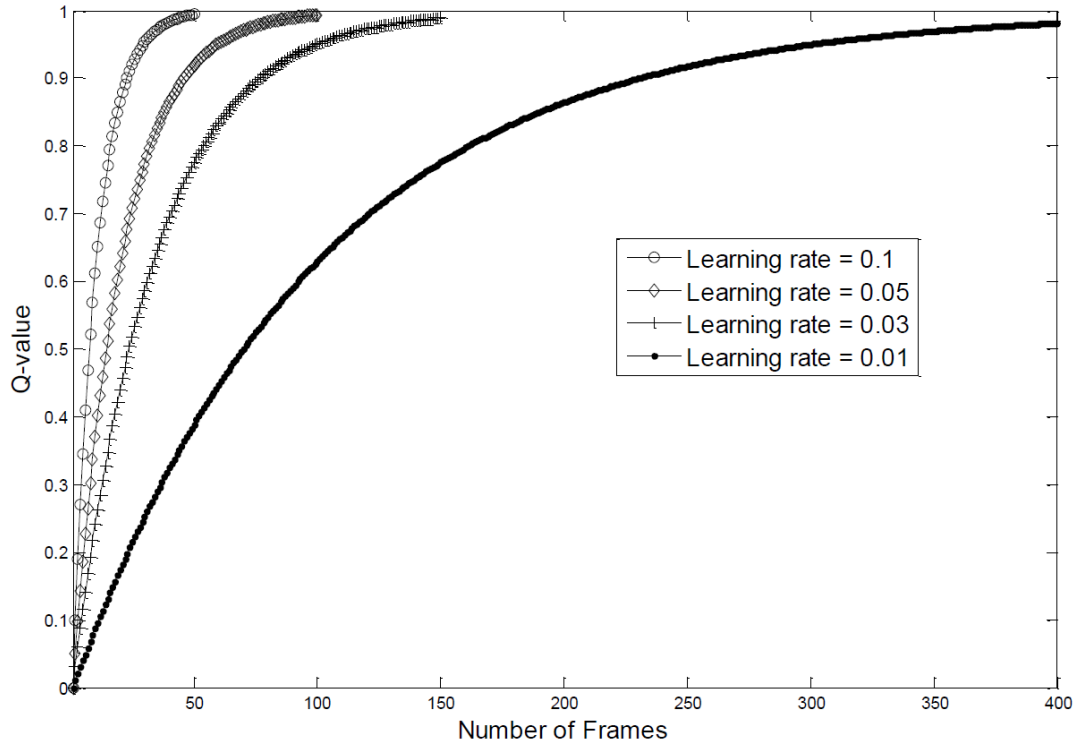


Figure 4-8. Q-value increase without collision

However, due to the specific requirements of asynchronous UW-ALOHA-Q schemes from a channel utilisation perspective, it is important to have a small number of slots (S) for a given number of nodes (N). Consequently nodes compete for the small number of slots and hence the level of contention is high, which leads Q-values of UW-ALOHA-Q nodes rapidly becoming negative (i.e. punishment).

Therefore the criterion of terrestrial ALOHA-Q to judge network convergence (i.e. 20 consecutive uses of a particular slot) is not appropriate for the implemented UW-ALOHA-Q system which have a smaller frame size (S) than the number of nodes (N) in the network since the number of available slots is not enough for all learning nodes. Figure 4-9 shows an example of one node's Q-table when the reduced frame size (S) is 2 for 10 fixed generating nodes, which are deployed in a network with the Uniform Random Back-off (URB) scheme. The line with squares is the changes in the Q-value of slot1 for one node in the network and the line with stars is the Q-value of slot2 of the same node. Compared to Figure 4-8, the trend of Q-value changes of UW-ALOHA-Q is much more arbitrary and random due to the lack of slots for all nodes in the network and the URB scheme. The two Q-values actually get negatively reinforced quickly at the initial stage because of contention in the network. The fluctuation of the Q-value represents the learning process where the node is trying to find the particular slot and an appropriate frame time through trial-and-error experiences.

If the same measure of ALOHA-Q is used in the simulation of UW-ALOHA-Q, nodal convergence would be measured much earlier, which would lead to a wrong simulation result. Due to the significant difference in the Q-value changing features between ALOHA-Q and UW-ALOHA-Q, an appropriate criterion is required for UW-ALOHA-Q to judge system convergence for the network having a smaller number of available slots than there are learning nodes. Therefore, this thesis proposes to use the absolute value for UW-ALOHA-Q rather than the number of consecutive transmissions in a distinct slot. UW-ALOHA-QM simulations judge nodal convergence when one Q-value is greater than 0.9 in the Q-table and network convergence when all nodes achieve nodal convergence.

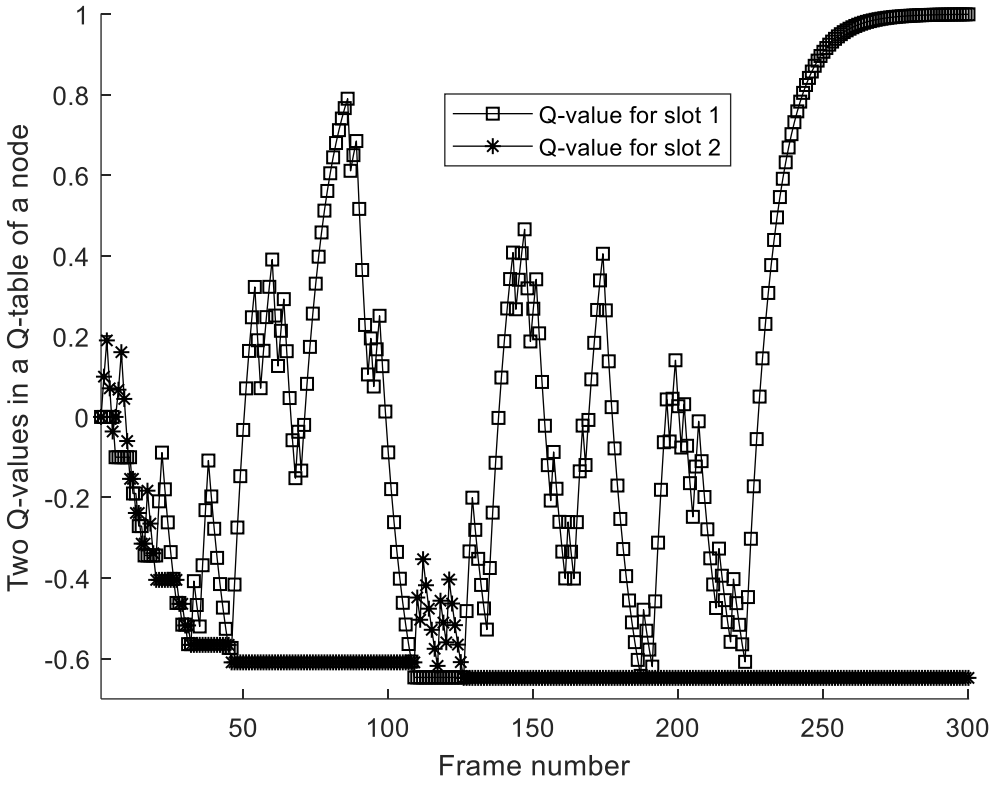


Figure 4-9. Example of Q-value changes of one node in UW-ALOHA-Q

4.8 Simulation

Simulations have been carried out to understand the baseline channel utilisation of UW-ALOHA-Q. Referring to the EPSRC funded project named 'USMART' [85], simulations in this section consider underwater networks comprising either 25 or 50 nodes, as well as with propagation distances varying from 100 m to 1,000 m networks.

This section first introduces the parameters and performance measures and then discusses an important trade-off of the UW-ALOHA-Q system. Simulation results showing channel utilisation and end to end delay are analysed according to the different network sizes. Moreover, the convergence features of UW-ALOHA-Q are looked at in detail. Lastly, the performance of UW-ALOHA-Q in a random topology is shown.

4.8.1 Parameters and performance measures

Mostly, channel utilisation (U) is measured which is evaluated as the fractional amount of time in which data traffic is successfully received at the sink node and is calculated by Equation (3-3).

We also define three parameters for simulation analysis. They are related to frame size (S) since the frame size is the important parameter which has a significant impact on the performance of UW-ALOHA-Q.

- Scvg: the convergence frame size which permits convergence to be achieved (as defined) for a certain size of network
- Smax: the frame size which can achieve the maximum channel utilisation for a certain size of network
- Index B: this ratio represents the theoretically available space at the sink node to be used for reception of data packets related to the total duration of the data carrying capacity in a frame as described in Equation (4-1). Because reducing frame size is proposed in section 4.3, the frame size is smaller than the number of nodes (i.e. S < N). We consider cases where the index (B) is greater than 1.

$$B = \frac{S \times (2 \times \tau_p + T_{dp})}{N \times T_{dp}} \quad (4\text{-}1)$$

where, the potential range of frame size (S) considered in this thesis is 0 < S < N.

Figure 4-10 provides the concept for the index B. If there are 4 sensor nodes in the network, the aim of the network is to successfully receive 4 data packets in a frame. The 4 data packets must successfully arrive at the sink node. At the sink node, the available time to be used for the data packet reception can be calculated. The sink node cannot receive the data packet during the time that it sends ACK packets because it is a half-duplex mode. Moreover, the guard time is reserved for the case in which the maximum delay is underestimated and to account for potential clock drift so that the guard time must be remain idle. Therefore, the rest of the time in a frame of the sink node can be used to receive data packets from 4 sensor nodes. The rest of the time is the denominator of Equation (4-1) during which the 4 data packets from 4 sensor nodes need to successfully arrive at the sink node for good channel utilisation. If the index B increases, the available time at the sink to be used for data packet receptions increases, therefore the potential number of collisions decreases but channel utilisation decreases due to the idle time at the sink node and vice versa. One aim of this simulation section is to find an appropriate index B between the trade-off of channel utilisation and network convergence.

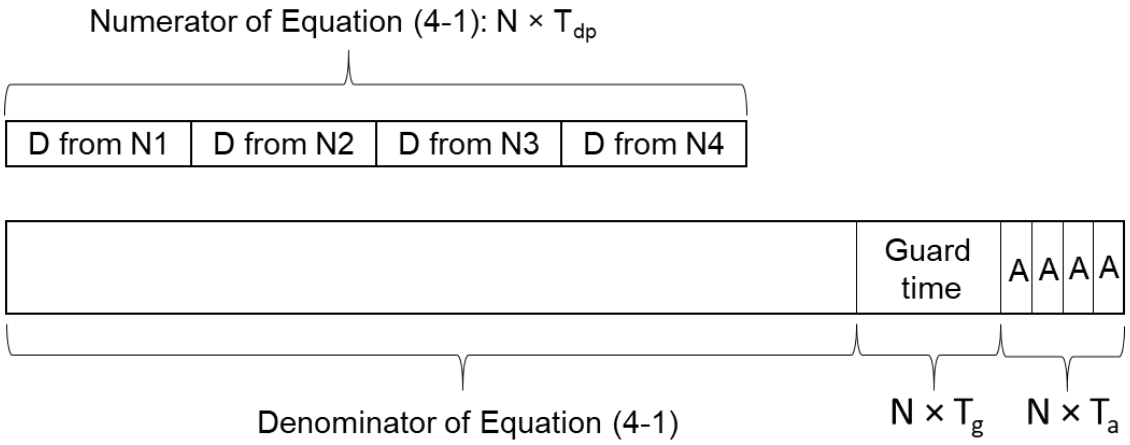


Figure 4-10. The ratio index B of the sink node

4.8.2 The trade-off between channel utilisation and convergence

This section provides simulation results of channel utilisation of UW-ALOHA-Q according to the frame size (S) and highlights a trade-off between channel utilisation and the probability of convergence. This section particularly shows simulation results with and without the uniform random back-off scheme to examine the capability of slot learning process with and without the scheme. Table 4-2 lists details of simulation parameters in this section.

Parameters	Value
Duration of a data packet of 1044 bits (T_{dp})	16.704 ms
Duration of an ACK of 20 bits (T_a)	0.32 ms
Duration of a guard time of 36 bits (T_g)	0.576 ms
Duration of a slot (T_s)	150.933 ms
Network size in radius (R)	100 m
Tx and Rx data rate (r_{uw})	62,500 bps
The number of generating nodes in a network (N)	25 nodes
Propagation speed (v_{uw})	1,500 m/s
Propagation delay (τ_p)	66.67 ms
Topology	Single-hop, star topology

Table 4-2. Simulation parameters

Table 4-3 shows the simulated channel utilisation of UW-ALOHA-Q when the frame size (S) varies, for a network comprising 25 generating nodes (N) equally spaced around a 100 metre radius star topology with a central receiver. The simulations in Table 4-3 include the first two improvements (in sections 4.1 and 4.3) and exclude the uniform random back-off scheme (in section 4.6) in order to particularly understand the impact of changing the frame size (S). For each value of frame size (S), 100 simulations are carried out and one simulation comprises 5,000 frames to ensure sufficient time to converge.

The index ratio (B) in Table 4-3 is defined by Equation (4-1). When 4 slots in a frame (S) is used for 25 nodes (N) in Table 4-3, the index (B) is 1.44 which means that the idle time at the sink node has 44% more time compared to the total 25 data packet durations.

As the frame size (S) increases, the idle time at the sink node increases (so its index ratio (B) increases as well), and therefore the channel utilisation measured at the sink node is reduced due to the long idle time. However, the idle time at the sink node is used to receive data packets in the network, such that the chance of network convergence increases because generating nodes are likely to find an appropriate gap at the sink node through reinforcement learning. If the frame size (S) decreases, channel utilisation increases to a certain level as the available time at the sink node is reduced, but the possibility of collisions would increase since nodes cannot find a distinct slot in a frame. Therefore, there is a trade-off between average channel utilisation and the chance of network convergence according to frame size (S).

Frame size (S)	Index ratio (B)	The number of simulation trials where the network converges (times)	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
…	…	…	…
25	8.98	100	0.11

Table 4-3. Channel utilisation according to the frame size (S)

For example, during the simulations in Table 4-3, 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 neighbours. Increasing the frame size (S) up to 8 slots 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, despite a relatively low channel utilisation of 0.34 Erlangs. However, as shown in the results, a trade-off is observed when the frame size (S) is lowered from 8 to 5, 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 frame size (S) varies.

As stated earlier, the simulations in Table 4-3 do not include the new back-off scheme to examine the impact on the learning capability in relation to the frame size (S). As shown in Table 4-3, the network fails (base on the convergence criteria of 0.9 discussed in section 4.7) to converge on 3 occasions out of 100 trials when a frame size (S) of 8 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.

Applying the uniform random back-off 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. 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 4-4 compares simulation results with and without the uniform random back-off scheme. It shows that UW-ALOHA-QM overcomes the non-convergence issue by applying the uniform random back-off scheme with the frame size of 8 (Scvg) in this network configuration (i.e. 25 nodes in 100 m size network). Applying the uniform random back-off scheme, UW-ALOHA-QM can achieve network convergence 100 times out of 100 trials so the frame size of 8 is the convergence frame size (Scvg) in this network. However, more learning iterations are required to learn not only a distinct slot but also the appropriate frame start time.

Frame size (S)	URB scheme	The number of simulation where the network converges	Average channel utilisation (Erlangs)	Average number of frames used when the network converges (frames)
8	No	97	0.34	20.04
8	Yes	100	0.35	158.51

Table 4-4. Simulation results with and without the uniform random back-off scheme

Simulations of UW-ALOHA-Q have also been carried out using 25 nodes with different propagation distances varying from 100 m to 1000 m. Two key observations were found. First, an identical trade-off is observed under the condition that the index ratio (B) is greater than 1.5 in simulation results of all different network sizes. This also implies that the highest average channel utilisation of UW-ALOHA-Q is achievable under a condition of the index ratio (B) equal to 1.5. Therefore, we call this frame size Smax which is the smallest frame size meeting this condition.

Moreover, simulation results show that UW-ALOHA-QM achieves network convergence when the index ratio (B) is greater than 2.6 in all different size networks having 25 nodes. UW-ALOHA-Q is fundamentally able to achieve network convergence in a fixed node network and therefore it is important to understand the required amount of time at the sink node for network convergence in relation to the total amount of data packet duration in the network. Importantly these simulation results show that 2.6 times more space is required at the sink node for UW-ALOHA-Q to achieve network convergence.

4.8.3 Channel utilisation as a function of network size

In terms of network deployment, the size of a network (R) and the number of nodes (N) 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 (R) in order to define the baseline performance which UW-ALOHA-Q can provide for a range of different applications. Network configurations in Table 4-5 are used for the simulations and the uniform random back-off scheme is applied for network convergence. Convergence is measured when all nodes have at least one Q-value greater than 0.9 as discussed in section 4.7.

Figure 4-11 illustrates the simulated channel utilisation of UW-ALOHA-Q following convergence in a star topology where the network size varies from a 100 m to 1000 m radius with 25 nodes. These results present the detailed UW-ALOHA-Q behaviour based on the index ratio (B). As discussed in section 4.8.2, Scvg is the smallest frame size (S) under the condition where

the ratio index (B) is greater than 2.6. Network convergence is achievable when the index ratio (B) is greater than 2.6 as Figure 4-11 specifies. The convergence frame size (Scvg) varies from 1 to 8 as the network size decreases.

Parameters	Value
Duration of a data packet of 1044 bits (T_{dp})	16.704 ms
Duration of an ACK of 20 bits (T_a)	0.32 ms
Duration of a guard time of 36 bits (T_g)	0.576 ms
Duration of a slot (T_s)	Varies according to network size (R)
Network size in radius (R)	100 m to 1000 m
Tx and Rx data rate (r_{uw})	62,500 bps
The number of generating nodes in a network (N)	25 nodes
Propagation speed (v_{uw})	1,500 m/s
Propagation delay (τ_p)	Varies according to network size (R)
Topology	Single-hop, star topology

Table 4-5. Simulation parameters

In longer distance networks the propagation delay (τ_p) primarily accounts for one slot (T_s) as referred to by Equation (3-1) and during the propagation delay (τ_p), the channel is idle. Therefore increasing or decreasing frame size (S) significantly impacts channel utilisation due to the idle time. For example, in the longer distance networks such as those with a 900 m and 1000 m radius, the amount of idle time in one slot (T_s) 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 frame size (S) is 1 (Scvg). However, the amount of available time (τ_p) 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 frame size (S) equals 2 (Scvg). Adding one more slot in a frame, however, causes a significant drop in channel utilisation due to redundant idle time (τ_p). This change in channel utilisation is termed 'the effect of a slot'. This effect becomes smaller in smaller networks because the propagation delay accounts less for a slot and therefore there are small drops although the frame size (Scvg) increases in the smaller networks.

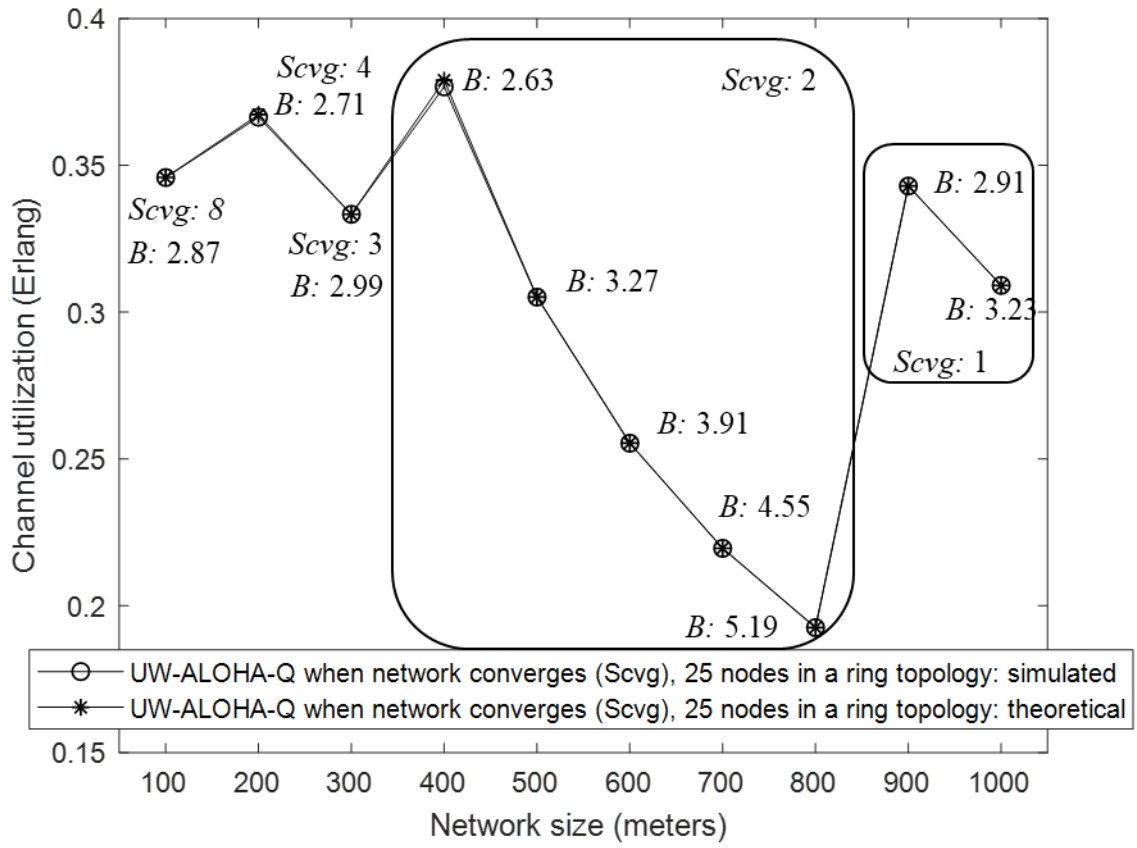


Figure 4-11. Channel utilisation of UW-ALOHA-Q at a variable network size

For the same reason, when the convergence frame (Scvg) continues to be identical (for example, networks between 400m to 800m), the channel utilisation decreases as the network size increases because the greater propagation delay accounts for a slot as the network size increases.

Overall, an UW-ALOHA-Q network converges when the ratio (B) is greater than 2.6. When the number of nodes (N) in a network remains 25, channel utilisation is higher if the index (B) is smaller as it indicates less idle time at the sink node whereas the larger ratio (B) implies a greater amount of idle time at the sink node resulting in lower channel utilisation.

Once a network has converged, all nodes use the same slot numbers 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 utilisation under network convergence can be determined by considering the proportion of time available for data transmission in just a single frame, as given by Equation (4-2). The theoretical maximum channel utilisation shown in Figure 4-11 is calculated using Equation (4-2) and it can be seen that a very close match is obtained.

Theoretical channel utilisation (U) of UW-ALOHA-Q when converged = $\frac{N \times T_{dp}}{S \times T_s}$ (4-2)

Figure 4-12 illustrates simulation results of channel utilisation of UW-ALOHA-Q using 50 nodes in a star topology and shows a similar trend to the channel utilisation results obtained when 25 nodes are used. The convergence frame size (Scvg) varies from 2 to 17 as the network size (R) decreases and is achieved when the index ratio (B) 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.

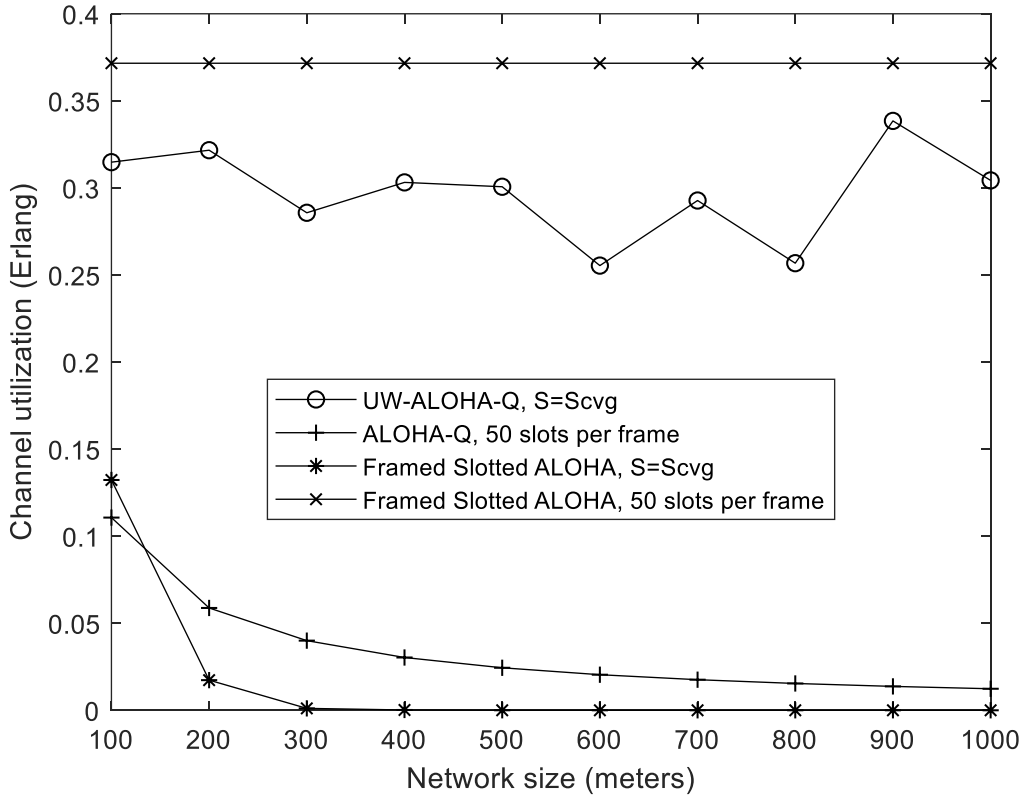


Figure 4-12. Channel utilisation with 50 nodes in variable network size star topology

For a comparative analysis, simulation results of framed slotted ALOHA and ALOHA-Q are also shown in Figure 4-12 when the frame size (S) of 50 and convergence frame size (Scvg) are used. Framed slotted ALOHA and ALOHA-Q are chosen for comparison since they are the baseline schemes for UW-ALOHA-Q. UW-ALOHA-Q achieves a much higher channel utilisation compared to ALOHA-Q when the frame size (S) is equal to the number of nodes (i.e. 50). 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 (refer to section 2.2.3.1.3), 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 convergence frame size (Scvg), most cases show almost zero channel utilisation.

4.8.4 End to end delay

One of characteristics of UW-ALOHA-Q is the reduced frame size (S), whereas ALOHA-Q and framed slotted ALOHA use a frame size (S) which is equal to the number of nodes (N) as shows. The frame size (S) for 25 nodes of ALOHA-Q and framed slotted ALOHA is shown as an example in Figure 4-14 and 50 nodes equivalently need 50 slots per frame. The reduced frame size results in a better end to end delay of UW-ALOHA-Q compared to framed slotted ALOHA and ALOHA-Q.

In any size of network, because of the longer frame size, one node of ALOHA-Q needs to wait for a much longer time for the next transmission than is the case with UW-ALOHA-Q and this becomes more serious in the underwater environment. For example, in a 1,000 m network consisting of 25 nodes, a slot duration (T_s) is 1.35 seconds calculated by Equation (3-1). As shown in Table 4-6, UW-ALOHA-Q uses only one slot (Scvg) to accommodate 25 nodes in a frame, 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 4-6. The table shows the average end to end delay and channel utilisation of 100 simulation trials for each result.

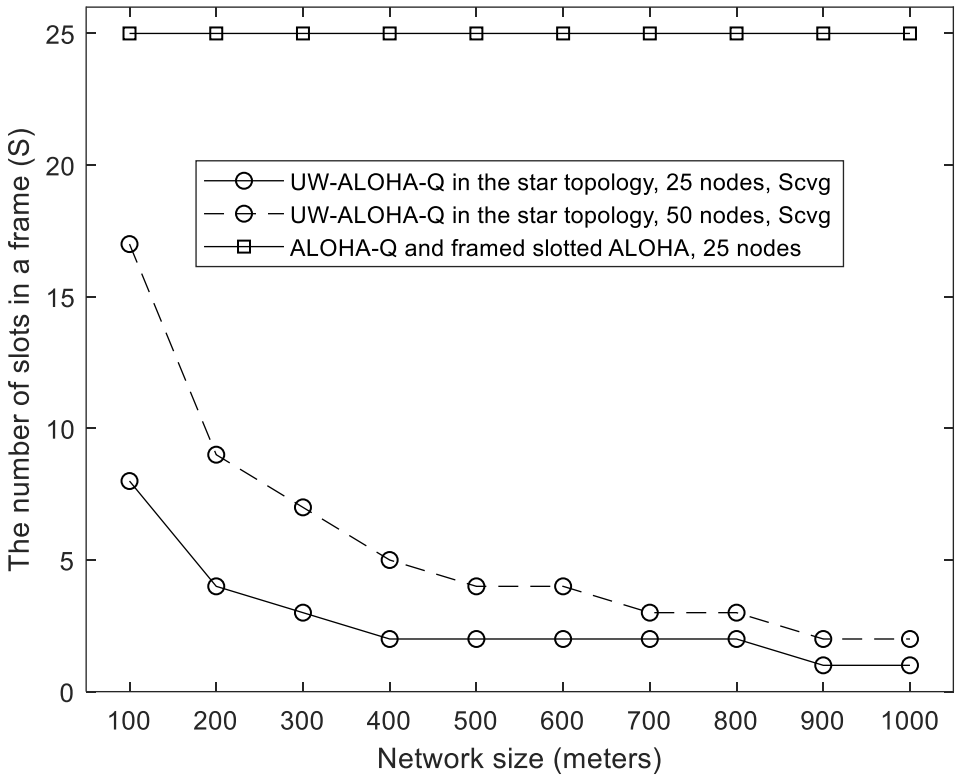


Figure 4-13. The frame size (S) used for UW-ALOHA-Q in different sizes of network (R)

Protocol	Frame size (S)	Network size (m)	The average end to end delay of successfully delivered data packets (seconds)
UW-ALOHA-Q	1 (Scvg)	1,000	271
ALOHA-Q	25 (S=N)	1,000	6787

Table 4-6. End to end delay of UW-ALOHA-Q and ALOHA-Q when 25 nodes are deployed

When 50 nodes are deployed, this benefit of UW-ALOHA-Q is magnified as shown in Table 4-7. UW-ALOHA-Q uses 2 slots in a frame (Scvg) for a 1,000 m size network whilst ALOHA-Q needs 50 slots in a frame (S).

Protocol	Frame size (S)	Network size (m)	The average end to end delay of successfully delivered data packets (seconds)
UW-ALOHA-Q	2 (Scvg)	1,000	555
ALOHA-Q	50 (S=N)	1,000	13576

Table 4-7. End to end delay of UW-ALOHA-Q and ALOHA-Q when 50 nodes are deployed

By reducing the frame size (S), UW-ALOHA-Q improves channel utilisation and decreases the end to end delay. Moreover, the protocol guarantees network convergence using uniform random back-off scheme with very low overheads (i.e. ACKs) and without the need of time synchronisation. Notably, greater benefits can be obtained in longer distance networks using a greater number of nodes in a network. These results demonstrate that UW-ALOHA-Q becomes more efficient in large scale networks where high propagation delay exists.

4.8.5 Network convergence

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. Figure 4-14 shows the channel utilisation as a function of time of UW-ALOHA-Q with and without the uniform back-off scheme and compares with ALOHA-Q in a 200 m network where 25 nodes are deployed. Please note that each graph in Figure 4-14 shows a typical example of four individual results rather than the average of multiple simulation trials. Three asterisk marks in Figure 4-14 indicate the times at which the network converges. Channel utilisation is measured using Equation (3-3) from the first frame at the end of every frame.

UW-ALOHA-Q without the uniform random back-off scheme shows fast 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. As shown in section 4.8.2, the back-off scheme solves this problem.

Given the network configuration in this section (i.e. 200 m network having 25 nodes), UW-ALOHA-Q using 4 slots per frame (Scvg) needs more iterations (i.e. more frames) to converge since the uniform back-off scheme disturbs nodes which achieves nodal 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 (Scvg) in Figure 4-14 fluctuates when the simulation starts and this fluctuation shows that nodes are learning the optimised frame start time and a distinct slot through trial-and-error learning processes. Once the network converges, the result shows an increase in channel utilisation due to collision free scheduling.

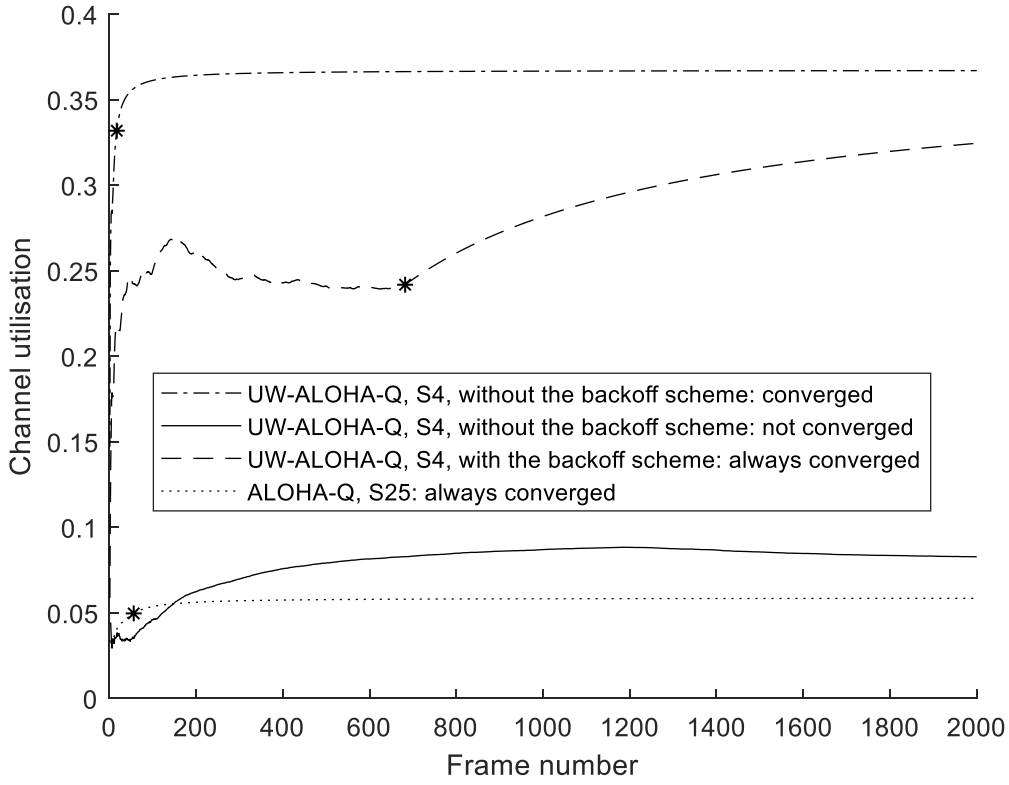


Figure 4-14. Real time channel utilisation as a function of time (25 nodes used)

The fluctuation implies that UW-ALOHA-Q is able to learn and operate in the time-varying environment. If environmental changes occur, the channel utilisation and the end to end delay performance fluctuate temporarily but UW-ALOHA-Q is capable of adapting and maintaining a good level of performance overall.

UW-ALOHA-Q achieves much higher channel utilisation than standard ALOHA-Q in the underwater environment when it converges, and its channel utilisation performance remains superior to ALOHA-Q even in a situation where it does not converge. Standard ALOHA-Q using the frame size (S) which equals the number of node (N) exhibits low channel utilisation due to the propagation delay (τ_p), however ALOHA-Q achieves network convergence in a short time since the sufficient number of slots allow the network to converge easier.

This section validates 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 utilisation following convergence is therefore a more important metric hence

performance metrics during the learning process are not considered, such as collision ratio during the learning process.

4.8.6 Random topology

The position of each sensor node can be determined depending on the application purposes and requirements. 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 of radius (R). Simulation results show that UW-ALOHA-Q achieves convergence using the appropriate frame size (Scvg) described in Figure 4-13. This is the interesting benefit of UW-ALOHA-Q since the protocol can provide the identical baseline performance in the random topology. Figure 4-15 shows channel utilisation of UW-ALOHA-Q when 25 nodes are deployed in different sizes of networks. Full simulation parameters are described in Table 4-5 and the duration of a slot (T_s) varies according to the network size (R) since the duration of slot (T_s) is calculated by Equation (3-1).

A successful data packet transmission is determined by an ACK 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. Each graph of Figure 4-14 shows the average value of 100 simulation runs and the simulation results are measured from network convergence, which means the network is in a steady-state such as the TDMA system. Therefore, once network is converged, UW-ALOHA-Q provides the constant channel utilisation in the fixed network. 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.

Figure 4-16 shows the real time channel utilisation of ALOHA-Q and UW-ALOHA-Q in the random topology (200 m network, 25 nodes). 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 star topology. The results demonstrate that UW-ALOHA-Q is robust and tolerant in networks with varying inter-node distances.

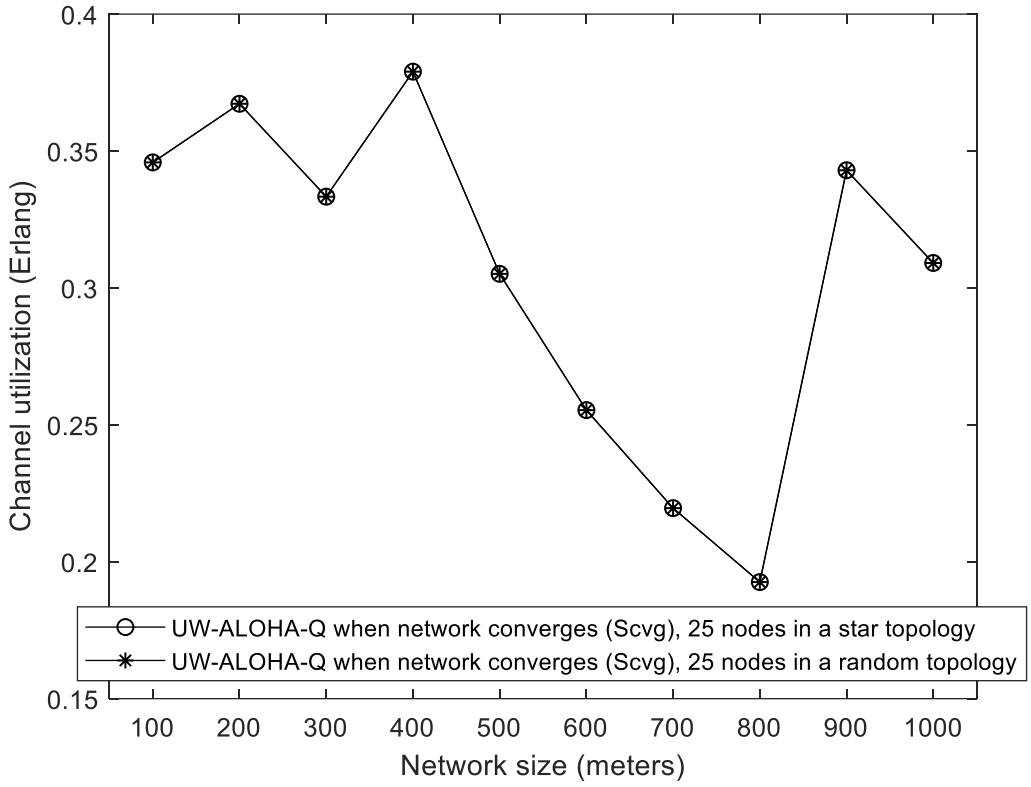


Figure 4-15. Channel utilisation of UW-ALOHA-Q in two different topologies

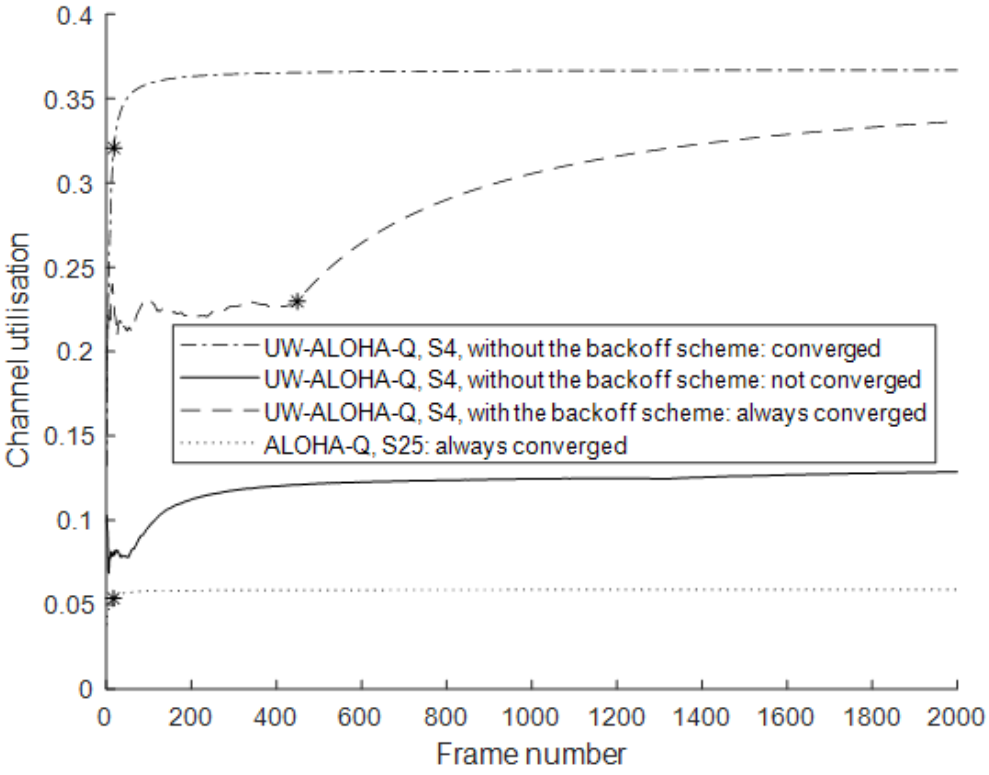


Figure 4-16. Real time channel utilisation in a random topology (25 nodes used)

4.9 Discussion

This chapter has 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 chapter has transformed the protocol to UW-ALOHA-Q for use in underwater acoustic networks comprising fixed (or pseudo static) nodes. 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. Simulation results show that UW-ALOHA-QM achieves network convergence, desired channel utilisation, and good end to end delay in the fixed underwater network.

End to end learning is achieved by the interaction using ACK packet reception between generating nodes and a sink 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 signalling overheads and complexity. Due to the very low overheads and complexity, hardware 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.

5 UW-ALOHA-QM for mobile underwater sensor networks

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 [87].

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 [88]. 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. 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. For example, in Figure 5-1, 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. Figure 5-1 shows the example where there are

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

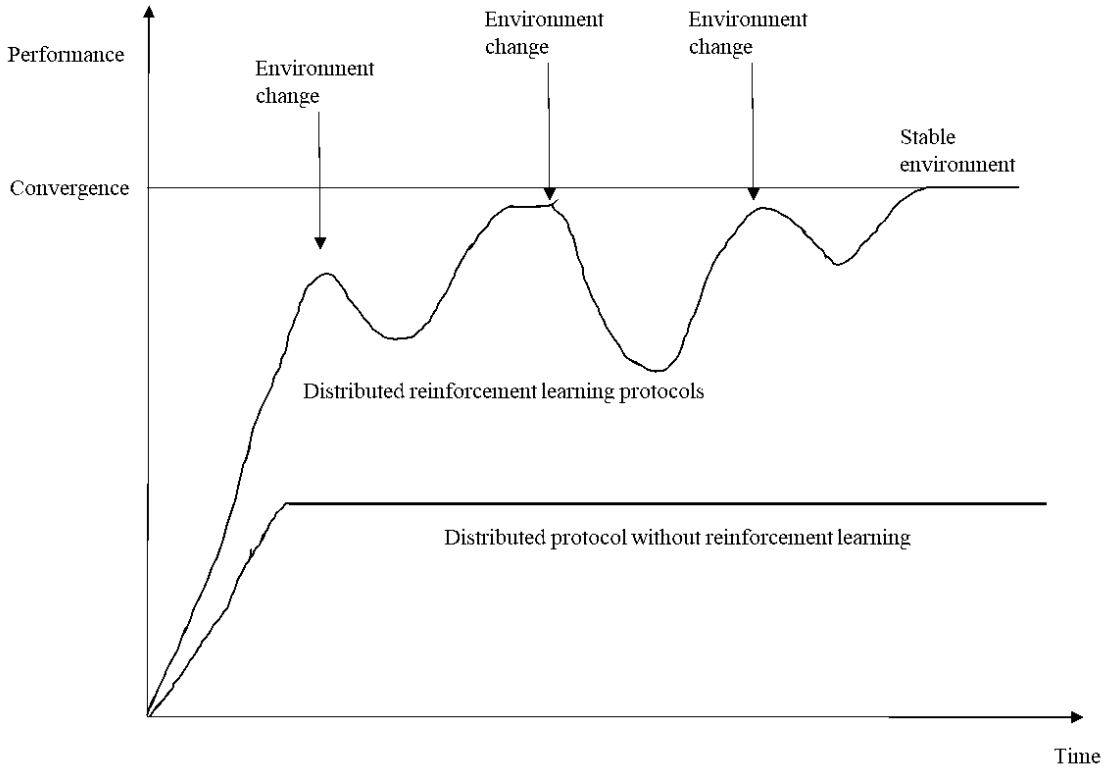


Figure 5-1. Network resilience

5.1 7 - Uniform random back-off

When there are sufficient slots in a frame (i.e. S=N), UW-ALOHA-Q nodes need to learn only a distinct slot in a frame in order to avoid collisions at the sink node. However, having a reduced frame size, nodes additionally need to learn the appropriate frame start time to fill in the gap at the sink node. Therefore, URB is proposed in section 4.6 to help nodes adjust their frame start time, but it significantly decreases the channel utilisation in the network comprised of mobile nodes 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 and transmitting a data packet in the selected slot can generate collisions in the mobile network. Moreover, mobility makes learning of UW-ALOHA-Q 'myopic' [89]. 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. Therefore in the mobile network, URB (the new frame start time) causes a situation that all nodes must learn the new environment from scratch in every frame resulting in inefficient and unnecessary learning processes.

Therefore, the URB design needs to be modified for mobile networks to achieve 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 Equation (3-2), we can calculate when a node needs to trigger a new learning process. Assuming a node experiences a collision 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 after the 7th consecutive collision. Seven consecutive failures causes the Q-value to return to ≈ zero at a learning rate (α) of 0.1 as shown in Figure 5-2. The previous study [84] carried out the analysis of Q-value in a radio wireless network and shows the same result.

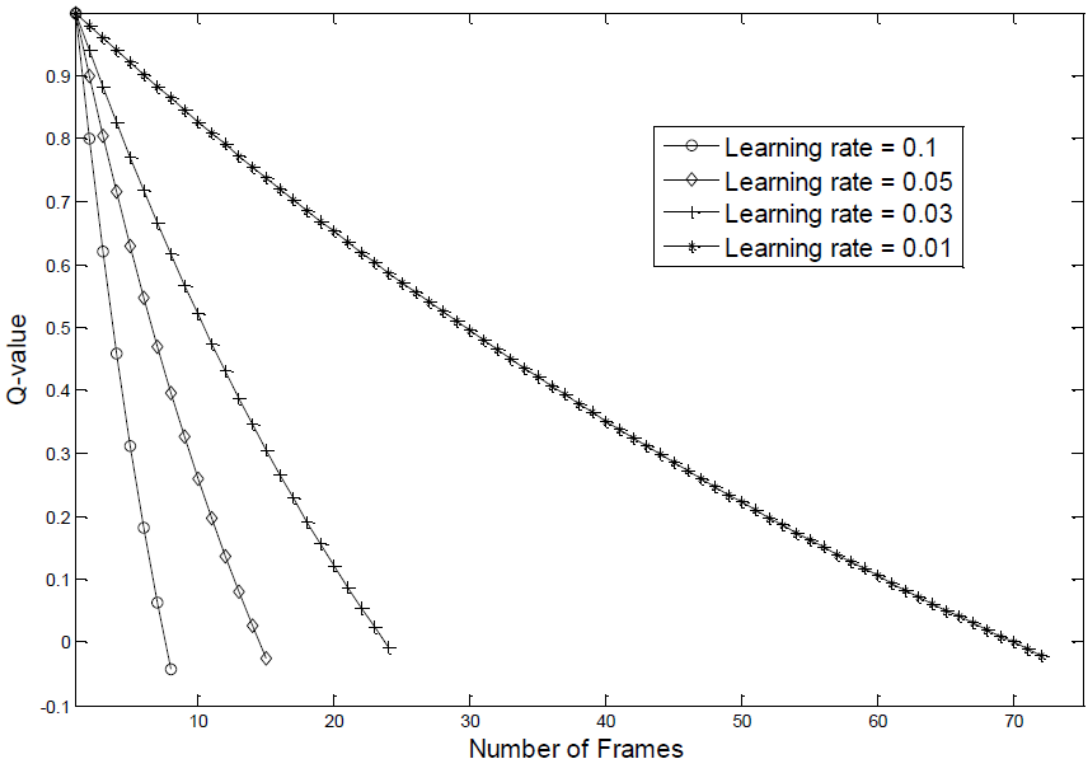


Figure 5-2. Level of resilience to loss of convergence

Therefore, based on the results of verified studies about Q-learning in ALOHA-Q [84], this thesis proposes the 7-Uniform Random Back-off (7-URB) scheme which invokes the URB scheme after

seven consecutive collisions for the mobile network. 7-URB utilises the experienced Q-value and removes unnecessary learning processes. [78] analyses Q-value changes in a Q-table for ALOHA-Q, which is designed for WPANs, however underwater properties (for example the slow propagation speed) do not impact on the Q-value changes. The stateless Q-learning function, Equation (3-3), does not include any variables related to terrestrial and underwater environments. Therefore, the Q-value study [78] can be applied to UW-ALOHA-QM.

In summary, a technique for nodes to adjust their frame start times is required for UW-ALOHA-Q because there is a much smaller number of slots in a frame than the number of nodes (N) in the networks. Therefore, URB is considered for nodes to find appropriate timing to send a data packet so the packet can be received by the sink node when it is idle. However, URB at every frame in a mobile network causes a serious problem, very high collisions due to the unnecessary learning processes and myopic learning. Therefore, 7-URB is proposed to make sure the current best Q-value cannot represent optimal action anymore. 7-URB reduces the unnecessary learning processes and adjust the frame start time for data packets to arrive during the idle time at the sink node. Mobile nodes can learn a best timing and slot number in a frame by efficient learning iterations.

5.2 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 purpose of these four different scenarios is provide evaluation of UW-ALOHA-QM for wide range of different characteristics of mobile scenarios.

These scenarios broadly are

- Moored or anchored sensor network
- Free floating sensor network [44]
- AUV assisted network [43]
- AUV sensor network [48].

The first scenario is a reference scenario for illustration of the fundamental operation of UW-ALOHA-QM with typical parameters. The subsequent three scenarios and their corresponding parameters are defined in other MAC protocol studies, [44], [43], and [48] which are reviewed in section 2.2.5. These four scenarios have been considered to provide a comprehensive evaluation of UW-ALOHA-QM and have been chosen for two primary reasons: 1) to provide very distinct mobility setups and cases, and also to provide a wide evaluation of the capability of UW-ALOHA-QM and 2) to allows direct comparison of UW-ALOHA-QM with other state of the art schemes based on results presented in their papers. For each scenario, results from the respective paper from which the 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 order to evaluate UW-ALOHA-QM and the results figures show all these combined. Channel utilisation of UW-ALOHA-QM is measured at the sink node using Equation (3-3).

5.2.1 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 due to wave motion with the assumption that nodes are spatially correlated. Spatial correlation is generally used as a fundamental assumption for studies of underwater node localisation [28-30] and it means that when one node moves, the other nodes also move in a related pattern. The parameters for UW-ALOHA-QM used for this scenario are listed in Table 5-1. Data packet size, ACK size and guard time size in bits are derived from previous studies [76]. For practical underwater environment settings, the data rate (r_{uw}) of 13,900 bps is chosen by referring to an underwater modem which is currently on the market [90].

Parameters	Value
Duration of a data packet of 1044 bits (T_{dp})	75.108 ms
Duration of an ACK 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)	212.27 ms
Network size in radius (R)	100m
Tx / Rx data rate (r_{uw})	13,900 bps
The number of nodes in a network (N)	25 nodes
Propagation speed (v_{uw})	1500 m/s
Propagation delay (τ_p)	66.667 ms
Frame size (Smax)	14
Maximum theoretical channel utilisation	0.631 Erlangs
Node speed	2-4 m/s

Table 5-1. Typical UW-ALOHA-QM parameter for underwater use

A number of studies have been undertaken to measure tidal currents at sea, such as those devoted to tidal energy research devoted to discovering sites of fast current movements for harnessing effective tidal energy resources. One example is [91] which refers to a velocity profile at tidal-stream energy sites 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 as Figure 5-3 shows. This data is used to provide realistic mobility levels for sensor nodes in the moored and anchored scenario.

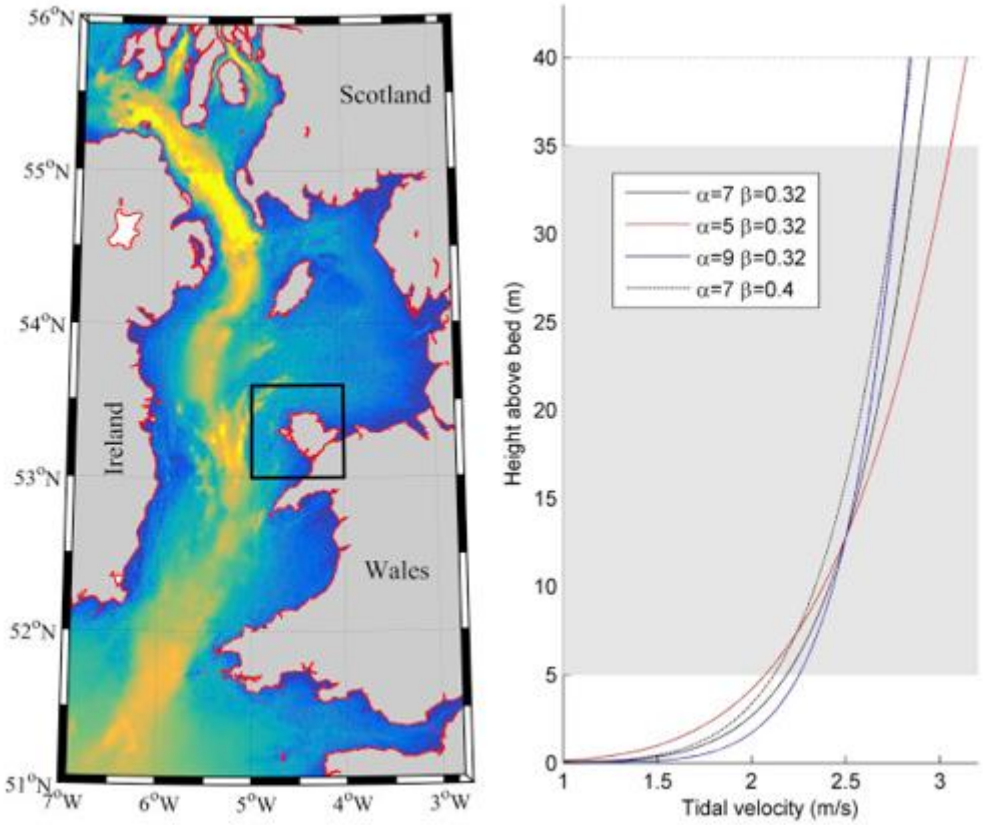


Figure 5-3. Tidal velocity profile of the Irish Sea

Figure 5-4 shows the trajectories of nodes in this scenario. There are 25 sensor nodes in a single-hop topology in a circular area with one sink node located centrally. Nodes will start at a uniformly distributed random position, within the 100 m radius circle (R). All nodes are considered to be within the interfering range. We assume packets are generated according to a random Poisson arrival process as the baseline model. Referring to the previous ALOHA-Q [76], each transmitter is designed to generate constant size packets with exponentially distributed packet inter-arrival times at the same rate as all the other transmitting nodes. At any instant in time, each node has no more than one data packet to transmit and the queue size of each node is 200. All lost packets are due to the packet collisions. To provide a worst case model, any overlap in packet reception times causes the complete packet to be lost.

In Figure 5-4, each movement happens every 30 min, i.e. at 30 min (at frame 605), 60 min (at frame 1,210), and the last one at 90 min (at frame 1,805). 30 min between each movement is sufficient to allow the network to converge. Nodes move in a random direction 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 $2/\pi$ radius.

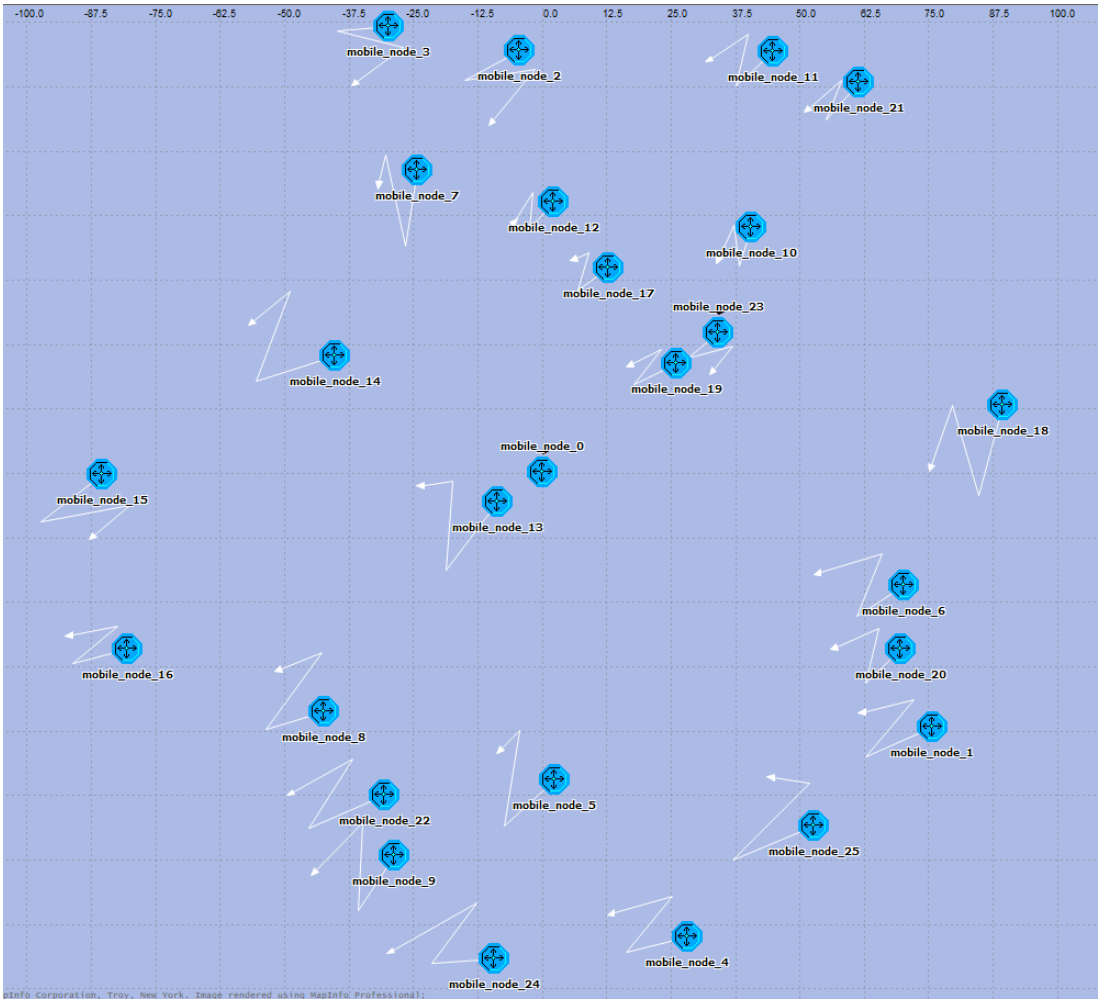


Figure 5-4. Discrete movement of 25 nodes in a random topology of 100 m network size

Figure 5-5 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 and this leads to changes in node locations and hence 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 UW-ALOHA-QM is able to learn and adapt to changes in the network without a coordinating node or additional control message exchanges.

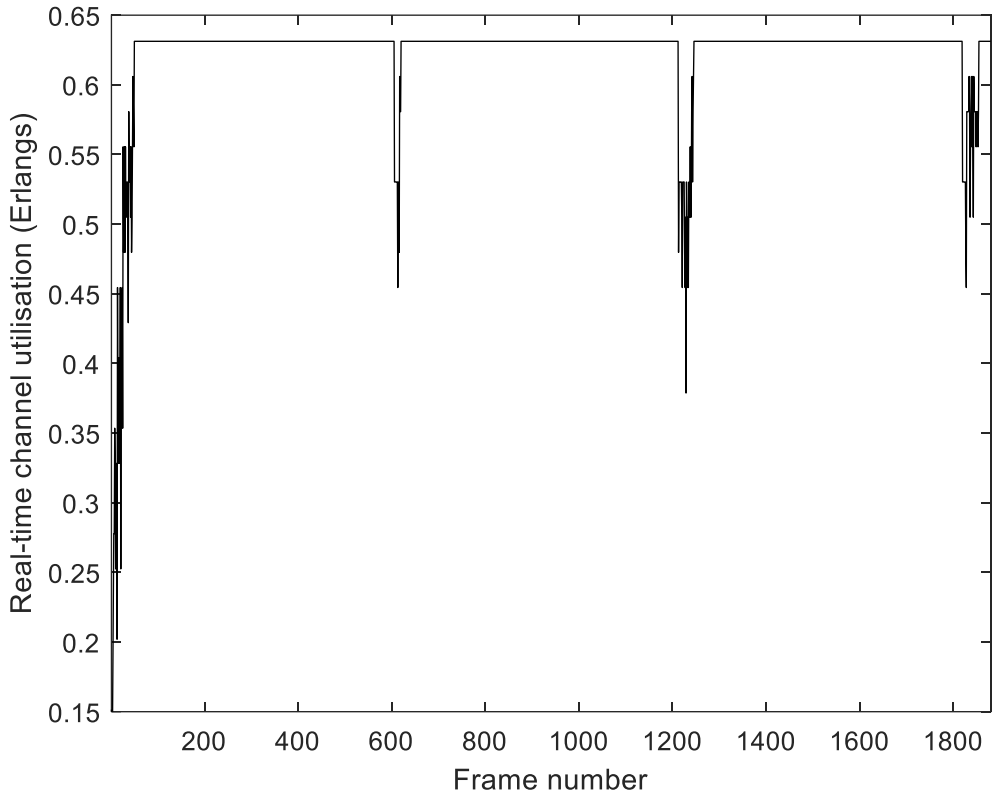


Figure 5-5. Real time channel utilisation of UW-ALOHA-QM

5.2.2 Free floating sensor networks

This type of mobile network is characterised by floating by current or wave movements. UW-ALOHA-QM is evaluated and compared to DOTS [44] which is designed for free floating sensor networks. DOTS uses time synchronisation and the Meandering Current Mobility (MCM) model [92] for node movement. More details of DOTS are described in literature review in section 2.2.5.

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 then each node can build a delay map of its one-hop neighbours and calculate the expected time for a response back to the sender of the packet overheard. Parameters defined by DOTS are described in Table 5-2. The maximum node speed is restricted to 0.3 m/s [92]. The study shows that DOTS achieves 0.2 Erlangs of channel

utilisation when the offered load is greater than 1. For a fair comparison, UW-ALOHA-QM is simulated using the parameters suggested by DOTS [44].

Scenario	Free floating
Defined by	DOTS [44]
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
Maximum channel utilisation	0.2 Erlangs
Simulation time	50 runs × 1 hour

Table 5-2. Parameters used for free floating scenario evaluation

Figure 5-6 compares the simulated channel utilisation of UW-ALOHA-QM to the other protocols as reported in [44]. 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 [81] but the protocol achieves 0.617 Erlangs due to the node mobility. The small difference in channel utilisation stands out given that one network comprises mobile nodes moving at the slow speed. Nodes of UW-ALOHA-QM are designed to transmit one data packet in a frame which implies a periodic data transmission, and its period time is fixed to the duration of frame (T_f). Although the loaded traffic from the application layer increases, the nodes does not change the periodicity therefore the channel utilisation continues to remain rather to increase.

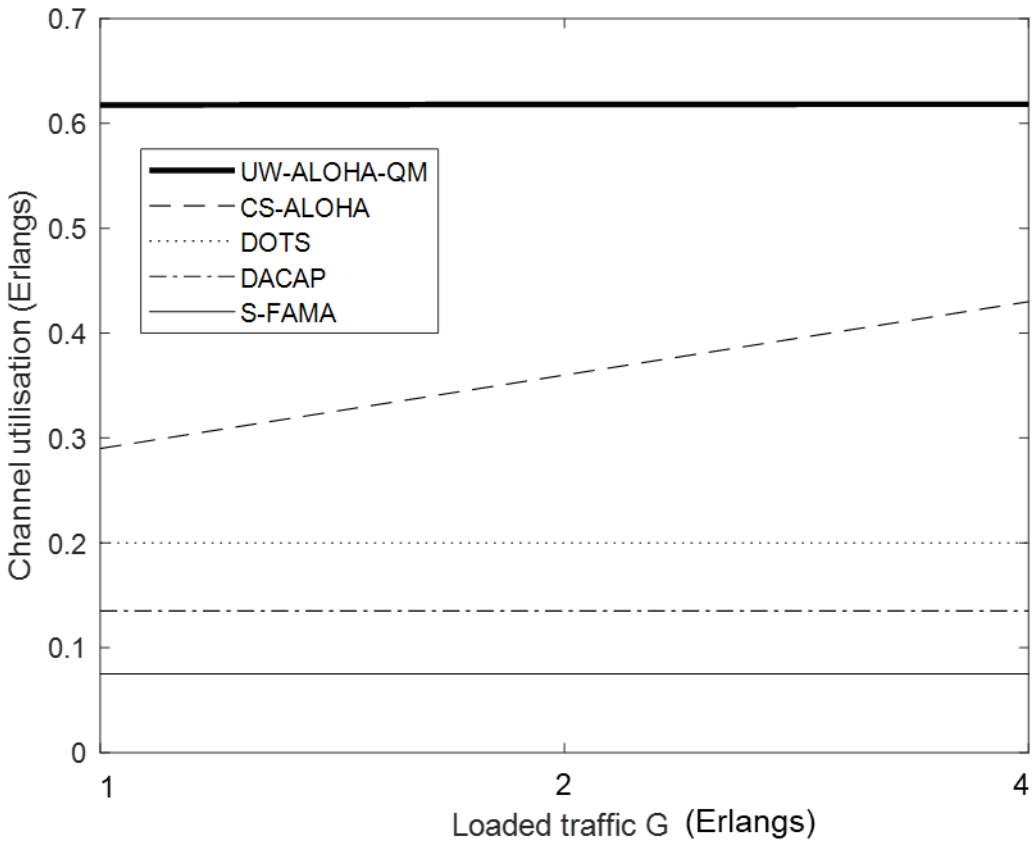


Figure 5-6. Channel utilisation according to different traffic (G)

Table 5-3 provides parameters for UW-ALOHA-QM in the network configurations defined by [44]. Given a network of 430 m size with 10 nodes [44], the smallest frame size under a condition that B is greater than 1.5 is 2 (Smax) for the maximum channel utilisation. In this network configuration, B is 1.6, which means a sink node has 60% more capability than a total of 10 data packet durations. In other words, the sink node is able to receive 16 data packets if the network is time synchronised and scheduled.

Duration of a data packet (T_{dp})	0.08192 seconds
Duration of a slot (T_s)	0.0656373 seconds
Duration of a frame (T_f)	1.312747 seconds
Frame size (Smax)	2
Index ratio (B)	1.6

Table 5-3. UW-ALOHA-Q parameters for free floating scenario evaluation

Figure 5-7 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 Figure 4-2, but it is illustrated in Figure 5-7 for an easy understanding of the theoretical concept of UW-ALOHA-QM. 10 sensor nodes send a data packet to the sink node and they are not time synchronised, therefore 10 packets arrive at the sink node in a random time. When the node speed is 0.3 m/s, a sensor node moves 0.39 meters during a frame which results in 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 that one data packet accounts for 6.24% of a frame. Therefore, the 60% more 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 Figure 5-7, N1 moves away from the sink node at 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 toward the sink node and their packets are collided 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 the proper gap at the sink node. Therefore, with slow node movement (0.3 m/s), UW-ALOHA-QM can maintain good channel utilisation.

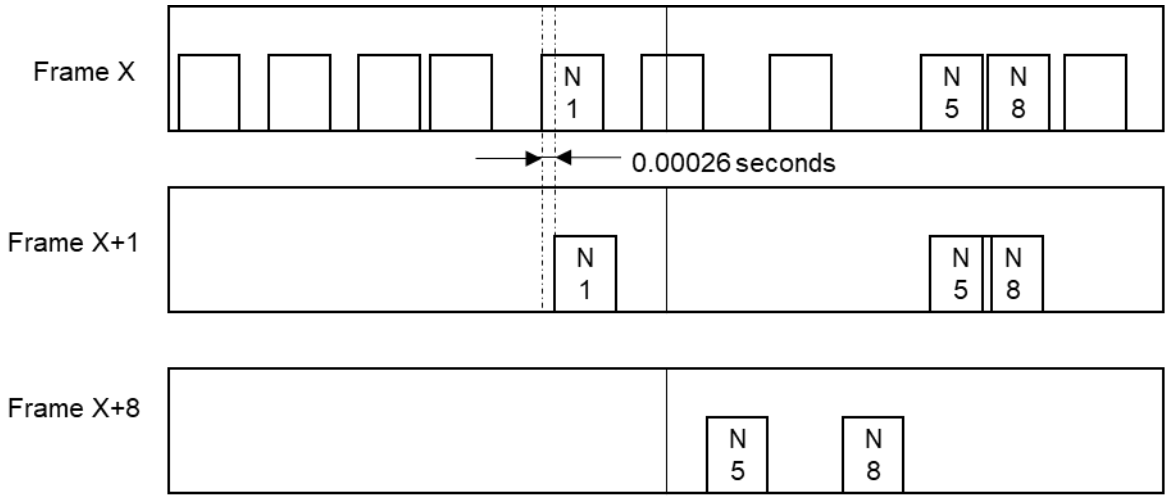


Figure 5-7. Data packet reception at the sink node

Figure 5-8 compares the channel utilisation of the different protocols using various node speeds from 0.3 m/s to 3 m/s. All nodes always move during the simulation time for one hour. DOTS shows 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 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.

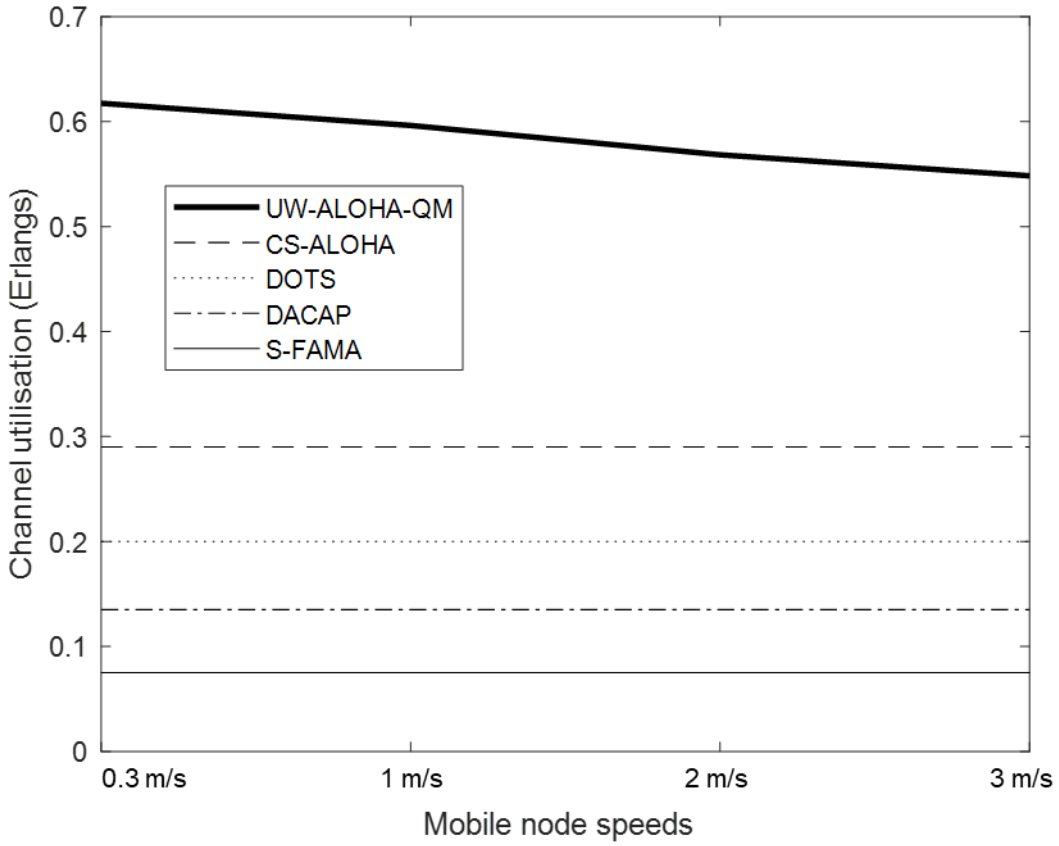


Figure 5-8. Channel utilisation according to different node speeds

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.

	0.3 m/s	1 m/s	2 m/s	3 m/
Average number of times 7-URB is 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

Table 5-4. The average number of times 7-URB is triggered

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 therefore the channel utilisation is low. However, UW-ALOHA-QM uses the learning approach where all nodes independently learn and fine a distinct slot and appropriate frame time through interacting with a time-varying environment, which brings about better adaptability and higher channel utilisation than other existing protocols.

Figure 5-6 and Figure 5-8 also can explain features of different MAC approaches. S-FAMA [109] is a synchronized underwater MAC protocol based on RTS/CTS handshaking. The main idea of S-FAMA is to time slot exclusive access to the channel medium so that the time duration of each slot is long enough to ensure that any frame transmitted at the start of the slot will reach the destination before the slot duration ends. CS-ALOHA [110] with ACK is ALOHA adapted for the underwater environment, where each node transmits whenever the channel is idle after performing carrier sensing without the handshaking process.

CS-ALOHA uses random access and the channel utilisation is therefore heavily dependent on traffic load (G) but is not dependent on the node speed. The DOTS, DCAP, and S-FAMA protocols conduct handshaking before the data transmissions and their performance is not effected 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 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 impacts on 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 adaptability. Therefore, the channel utilisation of UW-ALOHA-QM can be significantly higher than other protocols.

5.2.3 AUV assistant networks

These networks consist of fixed sensor nodes and one or more AUVs. The LTM-MAC [43] and Load adaptive CSMA/CA [46] protocols are designed for this type of mobile network.

LT-MAC [49] was proposed for small-scale static underwater networks and LTM-MAC [43] is an extended version for the extra AUV in the 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 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, those 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 described in [43], however asynchronous operation is applied. With a frame size (Smax) of 6, the theoretical maximum channel utilisation of UW-ALOHA-QM is 0.58 Erlangs with a saturated traffic model in this scenario [81]. Table 5-5 summarise the parameters used in the AUV assisted scenario and they are defined in [43].

Scenario	AUV assisted
Defined by	LTM-MAC [43]
The number of nodes (N)	7 static node + 1 AUV
Network size (R)	1,500 m
Data rate	2,000 bps
Packet size	500 bytes
Node sped	1 to 3 knots (0.51 m/s to 1.54 m/s)
Simulation time	1,000 seconds

Table 5-5. Parameters used for AUV assisted scenario evaluation

Figure 5-9 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 if there is a speed of 3 knots used 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).

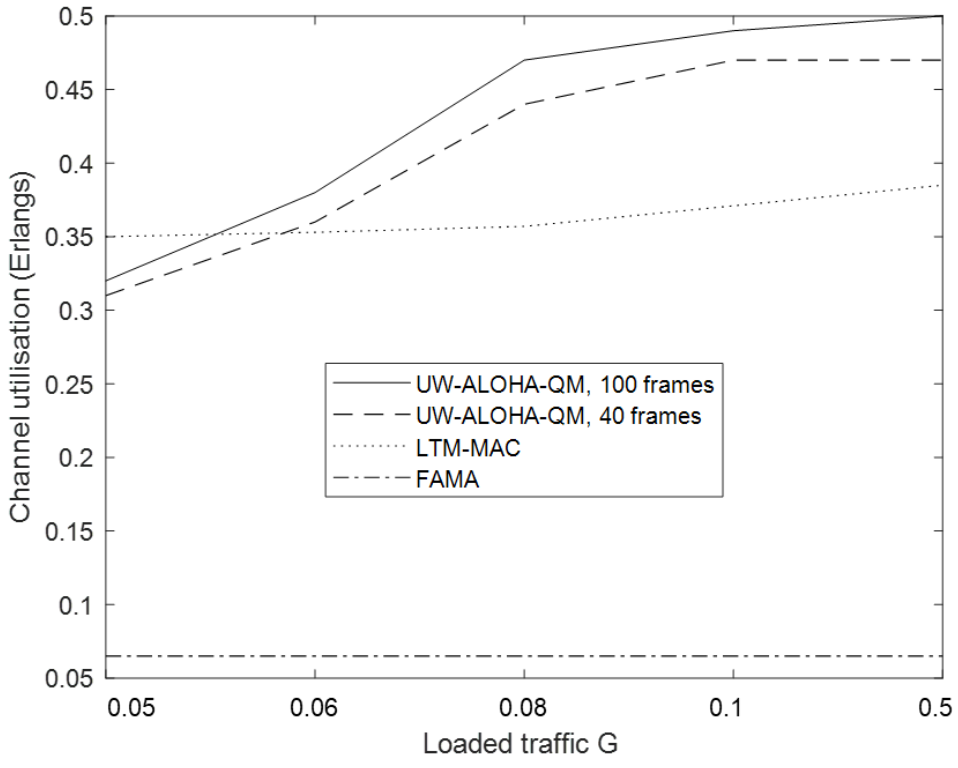


Figure 5-9. Channel utilisation according to different traffic (G)

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.

5.2.4 AUV sensor 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 [7]. Therefore, the movement models are different depending on the application requirements. The AUV speed is usually limited in order to save the energy needed for the propulsion of the AUVs. The speed of AUVs varies typically from 1 to 5 knots (5 knots: 2.572 m/s) [93]. The scenario described in APD-TDMA [48] is the AUV network and UW-ALOHA-QM is compared with the protocol in the scenario.

APD-TDMA [48] is designed for AUV sensor networks and it is an extension of the TDA-MAC protocol [50] 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 5-6 provides AUV network configurations defined by APD-TDMA and Figure 5-10 compares channel utilisation of 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.

Scenario	AUV network
Defined by	APD-TDMA [48]
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
Maximum node speed	5m/s
Simulation time	10,000 cycles (frames) × 10

Table 5-6. Parameters used for AUV network scenario evaluation

The theoretical maximum channel utilisation of UW-ALOHA-QM is calculated by Equation (4-2). The channel utilisation of UW-ALOHA-QM depends primarily on the network size (R) and also the optimum frame size (Smax). The growth of theoretical channel utilisation of UW-ALOHA-QM shapes the step increases in a given network size (R) and the number of nodes (N) because the frame size (S) is significantly impacts on the theoretical channel utilisation. Therefore, UW-ALOHA-QM shows the zig-zag style shape in Figure 5-10 which is typical feature as explained in section 4.8.3. Table 5-7 provides theoretical channel utilisation of UW-ALOHA-QM in different settings in the AUV network scenario.

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 manners.

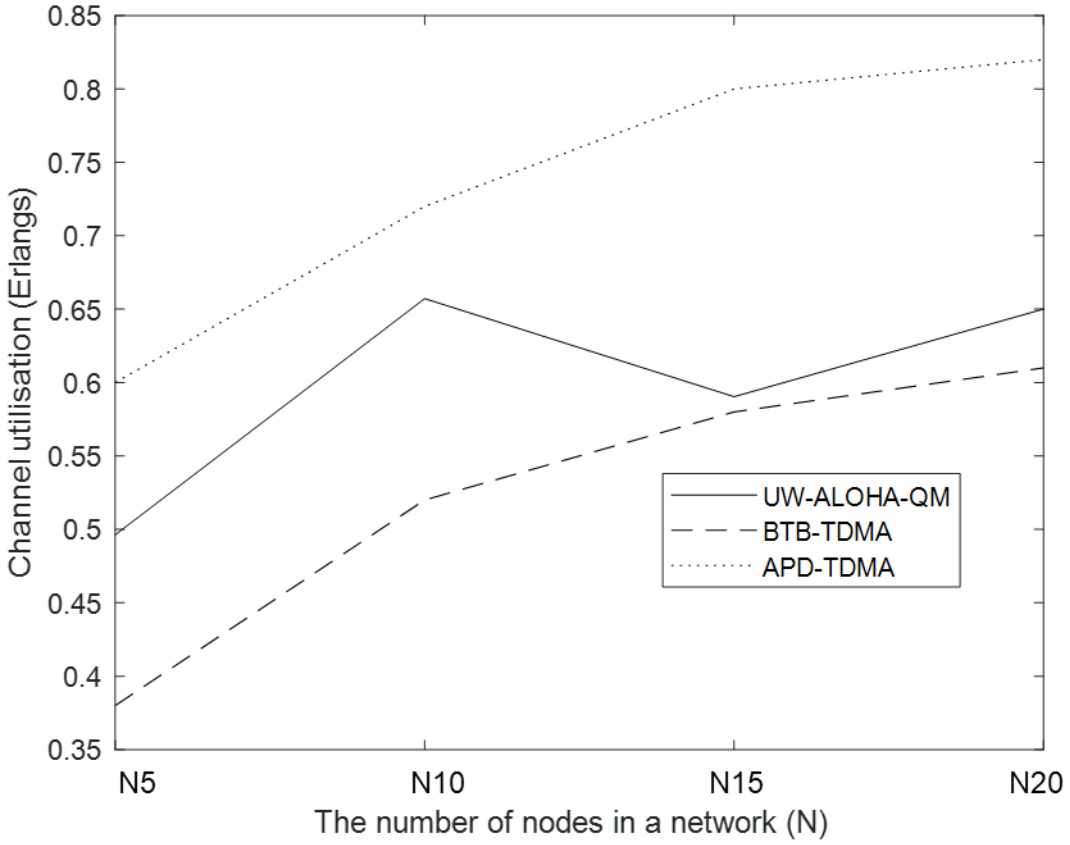


Figure 5-10. Channel utilisation in an AUV network

Network size (R)	The number of AUVs (N)	Optimum frame size (Smax)	Maximum theoretical channel utilisation (Erlangs)
1,500 m	5	2	0.5
1,500 m	10	3	0.66
1,500 m	15	5	0.6
1,500 m	20	6	0.66

Table 5-7. Theoretical maximum channel utilisation of UW-ALOHA-QM with different N and Smax

5.3 Discussion

UW-ALOHA-QM shows good performance for a range of different mobility scenarios. Most of published MAC protocols are typically designed to provide their best performance in a particular scenario or a certain type of network configuration and environment since each protocol aims at a specific application requirement. It does not necessarily follow that such protocols can provide a good performance in other scenarios.

UW-ALOHA-QM can be flexible and adaptable in its application to different scenarios, topologies, or network configurations since it is not geared to one specific type of scenario. UW-ALOHA-QM does not rely on specific scheduling, traffic, or channel assumptions. This scenario agnostic feature is achievable by the reinforcement learning approach.

The potential weakness of UW-ALOHA-Q is that two parameters need to be known: the number of nodes in the network (N) and network size in distance (R). In the underwater environment this assumption can be practical because sensor nodes are usually taken by ship to the sea and then deployed and therefore the two parameters are known before the network is deployed.

Due to the energy limitations, the speed of underwater nodes is also commonly limited by 5 knots: 2.572 m/s [93]. UW-ALOHA-QM shows lower channel utilisation as the node speed increases; however, UW-ALOHA-QM still outperforms non-learning protocols in the underwater environment.

If the number of node increases, the probability of collisions will increase as the number of packets which are attempted to be transmitted on the channel increases. This leads more learning iterations to find an optimal decision for each node. Therefore, as the number of nodes in the network increases, UW-ALOHA-QM will take more time to achieve the desired channel utilisation in the mobile network or network convergence will take longer in the fixed networks.

Having looked at the simulation results for four different scenarios, it is clear that UW-ALOHA-QM offers benefits and is an effective for topology agnostic solution. It does not solve all problems and it is not necessarily as effective in low load conditions, but overall it works effectively in many different scenarios.

Most existing protocols designed for mobile underwater networks are built upon their original protocols which were designed for fixed networks. Many use RTS-CTS handshaking which impairs channel utilisation, or carrier sensing which requires large guard bands. ADP-TDMA [48]

uses a unique approach without requiring time synchronisation but it requires multiple re-initialisation process and is appropriate only for scenarios where nodes move at a constant speed in a constant direction.

UW-ALOHA-QM can be used for any networks, in particular fully distributed networks where centralised scheduling is not feasible and frequent control message exchanges are inhibited by long propagation delays. Moreover, UW-ALOHA-QM is also appropriate for networks where node communication takes place over different propagation distances since nodes potentially do not move in a constant manner. Pure ALOHA (refer to section 2.2.3.1.1) is the common option which can be used in any network without any specific requirements. UW-ALOHA-QM can replace the role of pure ALOHA but at the same time provide much better performance in any network scenarios.

This thesis does not argue that UW-ALOHA-QM always provide high channel utilisation in any types of networks. UW-ALOHA-QM cannot provide high channel utilisation in some scenarios for example, the scenario where the traffic load is very low. In this scenario, UW-ALOHA-QM has insufficient learning iterations, and can therefore only provide a lower channel utilisation capability. However, UW-ALOHA-QM provides a flexible and adaptable scenario agnostic solution using the reinforcement learning approach which can be effective for a wide range of scenarios.

The main idea and simulation results of this chapter are in preparation for journal submission.

6 Summary and future work

6.1 Summary

This thesis has presented a detailed description of the Ph.D. research undertaken from 2016 to 2020. The primary concern of this work is concerned with providing a topology agnostic approach to medium access control by improving network resilience, adaptability, and flexibility using reinforcement learning.

Chapter 1 discusses challenges in underwater communication. The major issue is the ineffective utilisation of limited acoustic bandwidth due to the slow propagation speed of acoustic signals. Chapter 2 provides a detailed literature review. First, section 2.1 explains that acoustic signals are the most viable means for underwater communication since they propagate longer distance than radio and optical signals. However, using acoustic signals raises issues in that existing techniques for terrestrial communication cannot be directly apply to underwater networks. Section 2.2 introduces medium access techniques and the essentials of MAC protocols. It highlights that MAC protocols play a key role in underwater networks since the aim of MAC layer is efficient use of a shared channel and the achievable utilisation efficiency is governed by the underlying MAC protocol. This section also explains node mobility which is a major issue in the underwater networks. The literature review of MAC protocols designed for mobile underwater networks show that existing protocols are built up on methods used for static networks and that many propose additional means such as frequent control message exchanges to deal with node mobility rather than increasing network adaptability. Section 2.3 outlines reinforcement based learning protocols and shows that reinforcement learning is capable of interacting with network environment and improve the network adaptability through trial-and-error iterations. However, most existing reinforcement learning based protocols assume a fixed network or time synchronisation.

Therefore it is necessary to design a new reinforcement learning based MAC protocol for underwater acoustic networks in different scenarios since the reinforcement learning approach can replace the existing methods which significantly impair channel utilisation in the underwater environment. Chapter 3 compared the use of the ALOHA-Q protocol in the terrestrial and underwater environment and simulation results show a significant decrease in channel utilisation in the underwater environment primarily due to the slow propagation delay of acoustic signals. Therefore, chapter 4 proposes a new MAC protocol called UW-ALOHA-Q. This protocol applies

novel approaches on top of reinforcement learning taking account of the characteristics of the underwater environment. This new protocol achieves good channel utilisation and can guarantee network convergence by applying the learning processes in a network consisting of static nodes. The most notable benefits of UW-ALOHA-Q are that the protocol does not require time synchronisation and it achieves good performance regardless of the topology configuration with low overheads.

However, UW-ALOHA-Q exhibits poor performance in a mobile network due to the uniform random back-off scheme since it wastes the heuristic learning results obtained through learning. Therefore, chapter 5 proposes a new back-off scheme, called 7-URB which arises only when an agent is expected to lose convergence hence, UW-ALOHA-QM can reduce unnecessary learning processes. As a result, UW-ALOHA-QM achieves much higher channel utilisation compared to existing protocols by increasing network resilience and adaptability through reinforcement learning in different types of scenarios of mobile networks. Moreover, the new protocol is not limited by movement direction or speed.

6.2 Conclusion

Reinforcement learning techniques provide a means of scenario agnostic solution for medium access control in underwater acoustic wireless sensor networks. Through the ability to interact with the learning environment, reinforcement learning provides adaptability to UW-ALOHA-QM. Using stateless Q-learning, UW-ALOHA-QM achieves respectable channel utilisation without time synchronisation with low overheads compared to existing protocols which employ inefficient underwater solutions to deal with the slow propagation delay and node mobility. The limitation of UW-ALOHA-QM is that it requires enough iterations to learn the operating environment, however considering the typical long deployment time of underwater networks, the required iteration duration is not significant.

6.3 Novel contributions

The conference paper [63] by this author was the first attempt to apply reinforcement learning for channel access scheduling in underwater sensor networks. The UW-ALOHA-Q [81] built up on [63] is the first published MAC protocol aiming to provide good channel utilisation and convergence in underwater networks comprising fixed networks. UW-ALOHA-QM is the extended protocol from UW-ALOHA-Q [81] for mobile networks and can provide scenario agnostic solutions and high channel utilisation. UW-ALOHA-QM utilises the benefits of

reinforcement learning and considers the properties of the underwater acoustic channel and the underwater environment. Intensive simulation results provide an understanding of how the protocol operates and shows that the reinforcement learning approach outperforms the existing traditional schemes.

6.4 Recommendations for future work

6.4.1 Learning and node movement analysis

A theoretical analysis of the relationship between learning factors and the speed of node movement would be helpful to an understanding of UW-ALOHA-QM. Node movements bring unpredictable changes to the network such as changes in propagation distances/delays and the timing of transmissions/receptions. Therefore UW-ALOHA-QM needs to learn through interactions and to adapt quickly enough to these changes. This thesis shows that UW-ALOHA-Q is capable of achieving such adaptation for a range of typical mobility levels but a more analysis is necessary to understand the capabilities of such an approach and to provide a more informed choice of learning related parameters.

For example, further research could include an analysis of nodal channel utilisation, fairness, the number of necessary interactions (frames), and 7-URB (how fast the slot timing is changing) according the different node speeds, different learning rates (α), different the discount factors, and different reward policies. These values are expected to continue to fluctuate due to the node mobility, and such research would provide useful analysis in terms of upper or lower bounds.

6.4.2 Full duplex UW-ALOHA-QM

Recently, studies have been undertaken into the feasibility of a full duplex physical layer [94-97] for underwater networks. However, there are only a couple of studies on MAC protocols designed to support full duplex [98, 99]. UW-ALOHA-QM is designed for distributed networks and its use is not limited by the feature of full duplex or half duplex scenario. However, it is predicted that the performance of UW-ALOHA-QM would not significantly increase with a the full duplex physical layer, since a very low level of control messages (ACKs) are used in UW-ALOHA-QM. Therefore, further analysis and enhancement design are required to utilise the channel resource of the full-duplex efficiently in the underwater networks and the full duplex UW-ALOHA-QM is expected to have potential benefits in distributed peer-to-peer underwater networks where data needs to travel bi-directionally.

6.4.3 Power consumption

Energy saving is a critical issue in underwater networks as we discussed in section 1.3.2. Therefore, UW-ALOHA-QM needs to also consider energy efficiency. There is a previous study [77] which introduces Informed Receiving (IR) for ALOHA-Q to provide information about how many times each node plans to use the current slot based on Equation (3-2). However, this study assumes a static network and this approach is not appropriate to mobile networks since the highest Q-value is not always optimal due to the node mobility. Therefore, new energy consumption models are required for appropriate sleep, listen, and transmission rules of sensor nodes in mobile networks. Energy efficiency issue can be investigated by applying two different modes to UW-ALOHA-Q: periodic data transmission and event driven transmission for different application requirements.

6.4.4 Exploration and exploitation in learning of UW-ALOHA-QM

The balance of exploration and exploitation is an important feature of reinforcement learning as we discussed in section 2.3.1.3, especially in underwater networks in which the environment constantly changes. Potential future work in this area is very visible, since three balancing methods for exploration and exploitation were investigated for ALOHA-Q [79] in the terrestrial environment: greedy, ε -greedy, and decreasing e-greedy as Table 6-1 shows.

	Exploration	Exploitation
Greedy selection	No	Yes
ε -greedy selection	Probability of ε	Probability of 1- ε
Decreasing ε -greedy selection	Probability of Q_{value} before the convergence Probability of $Q_{convergence}$ after the convergence	Probability of 1-Q_{value} before convergence Probability of 1-$Q_{convergence}$ after convergence

Table 6-1. Balance policy between exploration and exploitation

ALOHA-Q [76] employs a greedy selection which means each agent always chooses the action with the highest Q-value (exploitation). Using an ε -greedy policy, an agent generates a random value between 0 and 1 called ε and explores with the probability of ε and exploits with probability 1- ε. When ALOHA-Q is implemented in hardware sensor devices, due to the limited capacity of the device, the sink node cannot immediately transmit ACK packets after receiving a data packet, which lead to the loss of ACK packets. Therefore, a study [79] proposes the decreasing ε-greedy policy to solve this problem and can achieve network convergence using the new policy.

6.4.5 Frameless protocol – one slot in a frame

UW-ALOHA-QM is based on framed slotted ALOHA. The disadvantage of the protocols is that the number of nodes (N) in a network needs to be known in order to determine an appropriate frame size (S) based on the application requirements. To alleviate this problem, the framed slotted ALOHA framework can be replaced by the slotted ALOHA protocol (refer to section 2.2.3.1.2). We consider a large enough slot called a super frame which meets the protocol requirement of B > 1.5. Theoretically, all nodes can learn where to send their data packet in a slot by repeating learning processes. With this new approach, UW-ALOHA-QM is supposed not to have the step shape channel utilisation for example in Figure 4-11 and frameless UW-ALOHA-QM is expected to provide the theoretical maximum channel utilisation achievable in the fixed network. In a fixed network convergence is achieved by more learning iteration than frame based UW-ALOHA-QM since the number of available actions (a_t) can go to infinity. However, it is still necessary to know the number of nodes (N) beforehand to determine the super frame duration.

6.4.6 Join and leave frequent scenarios

The advantage of UW-ALOHA-QM is that there is no need to execute any processes to join/leave to/from a network – no registration or withdrawal control message exchanging because all sensor nodes work independently. Frequent join/leave operations are practical assumption since the underwater environment dynamically changes for example, floating sensor nodes move in/out of range or nodes may become lost. It is predicted that UW-ALOHA-QM is robust against to the small number of new joining and leaving nodes since its design aims the scenario agnostic. However, it is necessary to study some extreme cases where the ratio (B) becomes very small or large when the channel is cannot efficiently used to figure out the limit of UW-ALOHA-QM.

6.4.7 Frame size adaptation

A new approach to adapt frame size (S) is required for the problem that the number of nodes (N) needs to be known prior to deploy UW-ALOHA-QM networks. For WSNs, distributed frame size selection is proposed [78], which uses a distributed method to select the size (S) for ALOHA-Q. The concept of a window is introduced in the protocol. A window consists of a number of frames and frames have the same number of slots in one window. At the end of second last frame, the algorithm decides to increase or decrease the number of slots for the next window until the frame size (S) reaches steady state. Using this solution, results show that agents can adapt their frame size (S) for the best channel utilisation so that it is not required to know the number of node (N) beforehand. However, the solution [78] assumes negligible propagation delay and hence the study

can propose to use a large window size (e.g. initial window size is 200) and this can cause very slow adaptation speed in the underwater environment due to the slow propagation speed. Therefore, a new design of a fast frame size adaptation scheme is required for future work considering underwater characteristics. Frame size adaptation can provide the solution of the issue described in section 6.4.5.

6.4.8 Multi-hop scenario

Multi-hop is more complex than single-hop since the routing feature needs to be considered into the design of MAC protocols. This thesis provides simulation results of single-hop networks, however UW-ALOHA-QM can fundamentally support the multi-hop scenario as well since its design focuses on being scenario agnostic rather than a specific single-hop data communication. However, in the multi-hop scenario, channel utilisation of UW-ALOHA-QM is expected to be impaired due to more interference signals by one-hop range neighbours. Therefore, it is necessary to investigate UW-ALOHA-QM in different multi-hop scenarios in different topologies such as tree topology or chain topology. There are previous studies [100] which propose multi-hop ALOHA-Q in the terrestrial environment and their results are promising.

6.4.9 Heterogeneous networks

Underwater devices can vary in a network and the different groups of sensor nodes (for example, static, mobile, moored or drifting nodes) can collect different types of data (for example, pictures, videos, or text type database) according to application requirements. Therefore, heterogeneous network studies are necessary to deal with different kinds of data transmission requirements in a single network. Traffic load, periodicity, priority, or data size are needed to be considered in UW-ALOHA-QM.

6.4.10 Other scenarios

This thesis considers scenario where sensor nodes collect data and send to the sink node. However, there can be other application requirements for communication types for example, any node to any node communication in a network. UW-ALOHA-QM fundamentally can support different types of communication scenarios, but future work is required to conduct more simulations to understand the underlying performance of UW-ALOHQ-QM in different types of communication.

6.4.11 Practical underwater channel environment

It is important to apply the practical channel environment to simulations to understand practical limitations and results of a protocol. Although simulations for UW-ALOHA-QM account for the

1,500 m/s propagation speed of acoustic signals in the underwater communication environment (refer to Table 3-4), other practical channel features are not considered. For example, we discuss Sound Speed Profile (SSP) in section 2.1.3.1, and the depth dependent SSP causes refraction of the acoustic waves resulting in curved propagation trajectories [18]. The trajectories can be obtained using the BELLHOP which is a beam tracing model for predicting acoustic pressure fields in ocean environment [101]. Also there are studies [102, 103] to collaborate BELLHOP and Riverbed Modeler (refer to the Appendices) which could be used to obtain more practical simulation results by applying realistic channel conditions.

6.4.12 More simulation results according to the node speed

This thesis focuses on channel utilisation because the limited bandwidth and the inefficient channel use is a major challenge in underwater networks. UW-ALOHA-QM cannot guarantee network convergence in mobile underwater networks. In this case the packet loss and hence the end to end delay are important factors to evaluate the network performance in particular, according to the node speed. Moreover, current simulation results need to be extended to make more generalisable results. For example, the optimal values for the learning process such as the learning rate (a) and reward values (r) need to be researched and simulated to optimise settings according to the node speed rather than fixed values for UW-ALOHA-QM.

Appendices

A. Protocol simulation using Riverbed Modeler ®

Most results presented in this thesis have been obtained through simulations in Riverbed Modeler. It used to be called Opnet Modeler until it was acquired in 2012. Riverbed Modeler is a network design and protocol simulation tool developed by Riverbed Technologies ® and supports modelling and simulating wired systems, satellites, mobile, and fixed radio systems.

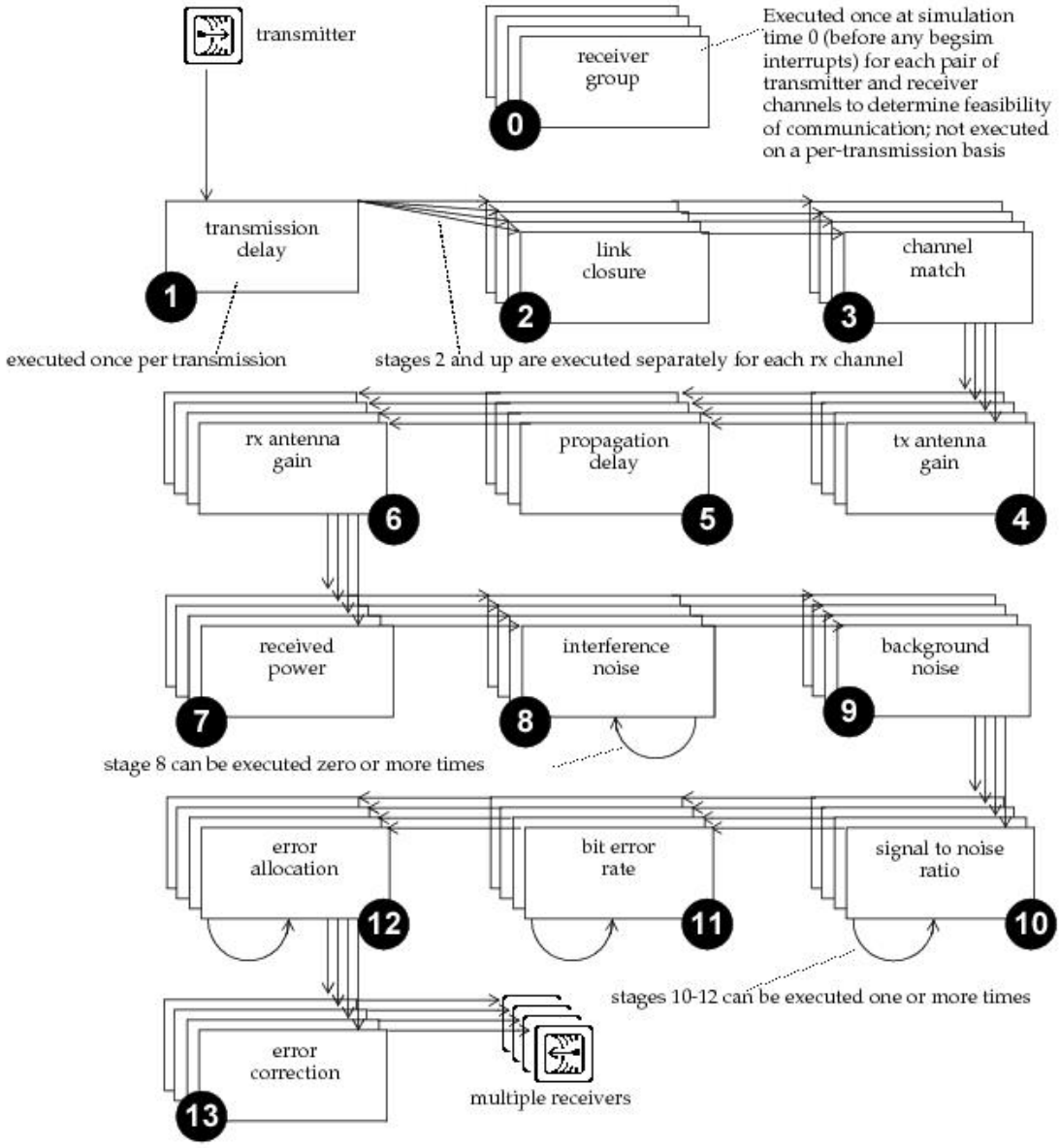


Figure (1) The radio transceiver pipeline

In order to evaluate packet transmission over a radio channel, Riverbed modeler executes a series of computational stage which constitute the radio transceiver pipeline. The pipeline consists of a

14 stage radio link model as shown in Figure (1). Details of the pipeline can be found in Riverbed tutorial [108] and this thesis explains stages which need to be modified for UW-ALOHA-QM simulation modelling.

First, receiver group needs to be updated at every time due to the node mobility. In a static network, the receiver group is executed once at simulation time 0 as described in Figure (1). However, the receiver group needs to keep updated due to the node mobility at the beginning of every frame to handle node mobility.

In order to set the slow propagation delay, pipeline stage 5 needs to be modified to the acoustic propagation speed of 1,500 m/s.

```
/* dra_propdel.ps.c */
/* Default propagation delay model for radio link Transceiver Pipeline */
/* propagation velocity of ~~radio~~ acoustic signal (m/s) */
#define PROP_VELOCITY      ~~3.0E+08~~          1500
```

For the Reception model at the sink node, pipeline 13 stage is modified for collision-based error model. This model arises a packet collision if there is an overlapped moment between data packets received at the sink node. The code from the default ecc stage is shown in normal font and the additional code added to create the modified ecc stage is highlighted in italic.

```
/* dra_ecc.ps.c */
/* Default error correction model for radio link Transceiver Pipeline */
#include <opnet.h>
void
pdm_ra_ecc (pkptr)
        Packet*         pkptr;
  {
        int             pklen, num_errs, accept;
        Objid           rx_ch_obid;
        double          ecc_thresh;
        /** Determine acceptability of given packet at receiver. **/
        FIN (pdm_ra_ecc (pkptr));
        /* Do not accept packets that were received */
```

```
        /* when the node was disabled. */
        if (op_td_is_set (pkptr, OPC_TDA_RA_ND_FAIL))
        {
           accept = OPC_FALSE;
        }
        else
    {
           /* Obtain the error correction threshold of the receiver. */
           ecc_thresh = op_td_get_dbl (pkptr, OPC_TDA_RA_ECC_THRESH);
           /* Obtain length of packet. */
           pklen = op_pk_total_size_get (pkptr);
           /* Obtain number of errors in packet. */
           num_errs = op_td_get_int (pkptr, OPC_TDA_RA_NUM_ERRORS);
           /* Test if bit errors exceed threshold. */
                 if (pklen == 0)
                 {
                    accept = OPC_TRUE;
              }
                 else
              {
                 if( op_td_get_int(pkptr, OPC_TDA_RA_NUM_COLLS) > 0 )
                 {
                    accept = OPC_FALSE;
                 }
                 else
                 {
                    accept = OPC_TRUE;
                 }
          }
         }
        /* Place flag indicating accept/reject in transmission data block. */
        op_td_set_int (pkptr, OPC_TDA_RA_PK_ACCEPT, accept);
        /* In either case the receiver channel is no longer locked. */
        rx_ch_obid = op_td_get_int (pkptr, OPC_TDA_RA_RX_CH_OBJID);
        op_ima_obj_attr_set (rx_ch_obid, "signal lock", OPC_BOOLINT_DISABLED);
        FOUT;
  }
```

B. Performance validation metrics

To verify performance of UW-ALOHA-QM, several performance metrics are collected and measured during simulations.

Channel utilisation (U)

Channel utilisation is a very important matric to underwater networks since acoustic channel struggles with the low efficiency due to the slow propagation speed and limited bandwidth. Therefore, this thesis primarily focuses on improving channel utilisation of underwater mobile networks and measures it using a unit called Erlang. Erlang corresponds to the fractional proportion of time during which data traffic is usefully received at a sink node. Therefore, 1 Erlang represents the fundamental capacity of the channel. The concept of channel utilisation can be described as:

$$\text{Channel utilisation (U)} = \frac{amount\ of\ time\ where\ the\ data\ traffic\ is\ usefully\ received\ (seconds)}{the\ total\ amount\ of\ time\ where\ channel\ utilisation\ is\ measured\ (seconds)}$$

The numerator can be calculated as the total number of data packets successfully received at the sink node (D) during a simulation • 1044 bits / data rate (r_{uw}) bps.

The denominator can be expressed by the total simulation time since a simulation time is counted by the number of frames in this thesis. Therefore, it can be calculated as the total number of frames in a simulation (M) • frame size (S) • slot length (T_s).

Various parameters are involved to calculate channel utilisation, hence channel utilisation can be expressed in different ways according to the network configuration and the environment.

- ALOHA-Q in the terrestrial environment when network converges

The propagation delay is negligible in the terrestrial environment, therefore the propagation delay (r_{tr}) in a slot can be ignored, which brings about high channel utilisation of radio networks. Moreover, ALOHA-Q uses the identical frame size (S) for the number of nodes (N) in a network, thus when a network converges, D and (M × S) become identical. Therefore, the theoretical maximum channel utilisation of ALOHA-Q can be simply measured:

$$\text{Channel utilisation (U) of ALOHA-Q} = \frac{T_{dp}}{T_s} = \frac{1044\ bits}{1100\ bits} = 0.95 \text{ Erlangs in steady state}$$

Data packet duration (T_{dp}) in seconds = $\frac{1044\ bits}{data\ rate\ (r_{tr})\ bps}$

Slot duration (T_s) in seconds = $\frac{1100\ bits}{data\ rate\ (r_{tr})\ bps}$

- ALOHA-Q in the underwater environment when network converges

In the underwater environment, the propagation delay (r_{wu}) is significant, therefore it needs to be considered to calculated channel utilisation.

Channel utilisation (U) of ALOHA-Q = $\frac{T_{dp}}{T_s}$= depends on network size (R)

Data packet duration (T_{dp}) in seconds = $\frac{1044\ bits}{data\ rate\ (r_{uw})\ bps}$

Slot duration (T_s) in seconds = $\frac{1100\ bits}{data\ rate\ (r_{uw})\ bps} + 2 \times \frac{network\ size\ (R)}{propagation\ speed\ (1500\ m/s)}$

- UW-ALOHA-Q when network converges

UW-ALOHA-Q reduces the frame size (S) to improve channel utilisation therefore, a slot cannot represent the overall channel utilisation of the system. However, a frame is repeated under network convergence status, thus channel utilisation of a frame can represent the overall channel utilisation of UW-ALOHA-Q in a fixed network. The frame size (S) is decided by the network size (R) and the number of nodes (N) deployed in a network. Channel utilisation depends on the network size (R), frame size (S), and the number of nodes (N).

Channel utilisation (U) of UW-ALOHA-Q = $\frac{N \times Tdp}{S \times Ts}$

This equation is for the theoretical maximum channel utilisation of UW-ALOHA-QM. UW-ALOHA-QM is designed for mobile networks hence channel utilisation is supposed to vary at every time. Therefore, channel utilisation of UW-ALOHA-QM (and other protocols mentioned above during learning processes), needs to be calculated by Equation (3-4).

End to End Delay

Delay is an important metric for time critical applications such as tsunami detection far off the sea. This thesis measures average end-to-end delay in section 4.8.4, which is the time between the generation of a data packet and the time that the packet successfully arrives at the sink node.

Network size (R), the number of nodes in a network (N), and frame size (S) significant impact on delay performance.

double [end to end delay] and initial value is 0; [packet generated time] is capsulated in a data packet; When a node receives an ACK [total end to end delay] = [total end to end delay] + [current time] – [packet generated time] – [propagation delay]; At the end of simulation [average end to end delay] = [total end to end delay] / [total number of packets successfully received at a sink node]

Convergence

UW-ALOHA-QM is not limited to be used for mobile networks but can be also used for applications which require reliable communication having fixed nodes on the seabed. Therefore, the ability to guarantee that all nodes can find their distinct slot and appropriate frame start time is the valid matric for the static networks. This thesis measures network convergence when absolute Q-values of all nodes are greater than 0.9 as discussed in section 4.7. Moreover, the duration from the start of learning and to the convergence can be considered as an important metric since it can be used to measure the learning speed in fixed networks. In sections 4.8.2 and 4.8.5, convergence speed is measured by the number of frames used to achieve the theoretical maximum channel utilisation of UW-ALOHA-Q in each scenario since the nodes interacts with the network by trial-and-error iterations during the learning process.

C. Pseudocode

Q-learning

Compared to Equation (3-2), Q-learning update is easy to understand and this pseudocode shows when data transmission is successfully delivered.

When I receive an ACK destined to my address before the guard time expires.
Number of received ACK ++; Number of successful data transmission ++; Number of consecutive collision = 0;

Number of consecutive success ++; Retransmission count = 0; /* Update Q-table */ Current slot number = transmission slot number; Q-value temp = Q-value in a Q-table [current slot number]; Q-value temp = Q-value temp + learning rate * (reward – Q-value temp); Q-value in a Q-table [current slot number] = Q-value temp;

Asynchronous operation

Nodes are assumed to start the frame at a uniform randomly distributed time within the range of zero to the length of one frame.

When all nodes are initiated
Asynchronous operation index = rand()%1000; Asynchronous operation delay = (double) asynchronous operation index / 1000 * frame duration Start the first frame at (current time + asynchronous operation delay)

7 - Uniform Random Back-off scheme

When a node does not receive an ACK until the guard times expires
The number of collision ++; Packet retransmission count ++; Number of consecutive collision ++; Number of consecutive success = 0; /* Update Q-table */ Current slot number = transmission slot number; Q-value temp = Q-value in a Q-table [current slot number]; Q-value temp = Q-value temp + learning rate * (punishment – Q-value temp); Q-value in a Q-table [current slot number] = Q-value temp; If the number of retransmission count > 6,

```
Destroy the delivered packet;
Packet retransmission count = 0;

If (Number of consecutive collision %7 == 0)
URB index = rand()%1000;
URB delay = (double) URB index / 1000 * frame duration;
Start the next frame at (next frame start time + URB delay);
else, start the next frame at the next frame time;
```

Glossary

ACK Acknowledgement

AI Artificial Intelligence

APD-TDMA Asymmetric Propagation Delay Aware TDMA

AUVs Autonomous Underwater Vehicles

BEB Binary Exponential Back-off

CDMA Code Division Multiple Access

CSMA Carrier Sensing Multiple Access

CSMA/CA Carrier Sensing Multiple Access / Collision Avoidance

CTS Clear To Send

DNN Deep Neural Networks

DOTS Delay aware Opportunistic Transmission Scheduling

DRL Deep Reinforcement Learning

EPSRC Engineering and Physical Sciences Research Council

FDMA Frequency Division Multiple Access

GPS Global Positioning System

IEEE Institute of Electrical and Electronics Engineers

IR Informed Receiving

ISO International Organisation for Standard

JSW Juggling-like Stop and Wait

LLC Logical Link Control

LOS Light Of Sight

LT-MAC Location based TDMA MAC

LTM-MAC Location based TDMA Mobile MAC

MAC Medium Access Control

MACA Multiple Access with Collision Avoidance

MCM Meandering Current Mobility

MDP Markov Decision Process

MIMO Multiple Input Multiple Output

MISO Multiple Input Single Output

ML Machine Learning

NLOS Non Line Of Sight

NODC National Oceanographic Data Centre

OSI Open System Interconnection

OFDMA Orthogonal Frequency Division Multiple Access

PDF Probability Density Function

PLAN-MAC Protocol for Long latency Access Networks – MAC

QL-MAC Q-Learning based – MAC

QoS Quality of Service

REMUS Remote Environment Monitoring UnitS

RFID Radio Frequency IDentificaiton

RL-MAC Reinforcement Learning based – MAC

RTS Request To Send

SARSA State Action Reward State Action

SDMA Space Division Multiple Access

SIMO Single Input Multiple Output

SSP Sound Speed Profile

S-MAC Sensor MAC

TCP/IP Transmission Control Protocol / Internet Protocol

TDA-MAC Transmit Delay Allocation – MAC

TDMA Time Division Multiple Access

T-MAC Timeout MAC

URB Uniform Random Back-off

UUVs Unmanned Underwater Vehicles

WLANs Wireless Local Area Networks

WPANs Wireless Personal Area Networks

WSNs Wireless Sensor Networks

7-URB 7-Uniform Radom Back-off

References

[1] "Goal 14: Conserve and sustainably use the oceans seas and marine resources." UN.org. https://www.un.org/sustainabledevelopment/oceans/ (accessed Sep. 18, 2020).

[2] “Ocean economy and innovation.” OECD.org. http:// http://www.oecd.org/ocean/topics/ocean-economy (accessed Sep. 18, 2020).

[3] 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*, Switzerland: Springer, 2015, pp. 109-115. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-11385-2_12

[4] A. C. Toz, B. Koseoglu, and C. Sakar, "Marine environment protection: New technologies on oil spill response industry,” in *Congress on Ship and Marine Technology*, Istanbul, Turkey, Dec. 8-9, 2016, pp. 1-16.

[5] P. Braca, R. Goldhahn, G. Ferri, and K. D. Lepage, "Distributed information fusion in multistatic sensor networks for underwater surveillance," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4003-4014, May. 2015.

[6] A. S. Iminova and E. G. Ivanova, "New technologies for ocean cleaning,” in *Achievements and Prospects of Innovations and Technologies*, Sevastopol, Russia, Apr. 18, 2018, pp. 335-340.

[7] M. Purcell, D. Gallo, G. Packard, M. Dennett, M. Rothenbeck, A. Sherrell *et al*., "Use of REMUS 6000 AUVs in the search for the Air France Flight 447," in *Oceans*, Waikoloa, HI, USA, Sep. 19-22, 2011, pp. 1-7.

[8] I. F. Akyildiz, D. Pompili, and T. Melodia, "Challenges for efficient communication in underwater acoustic sensor networks," *ACM Sigbed Review*, vol. 1, no. 2, pp. 3-8, Jul. 2004.

[9] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, Jul. 1948.

[10] 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, Oct. 2007.

[11] D. J. Wright, "Rumblings on the ocean floor: GIS supports deep-sea research," *Geo Info Systems*, vol. 6, no. 1, pp. 22-35, 1996.

[12] U. M. Qureshi, F. K. Shaikh, Z. Aziz, S. M. Shah, A. A. Sheikh, E. Felemban *et al.*, "RF path and absorption loss estimation for underwater wireless sensor networks in different water environments," *Sensors*, vol. 16, no. 6, pp. 890, Jun. 2016.

[13] T. S. Garrison, "Water," in *Essentials of oceanography*, 6th ed. Belmont, CA, USA: Cengage Learning, 2012, ch. 6, sec. 6.10, pp. 144.

[14] G. D. Ferguson, "Blue-green lasers for underwater applications," *Ocean Optics IV*, vol. 64, pp. 150-156, Nov. 1975.

[15] T. Wiener and S. Karp, "The role of blue/green laser systems in strategic submarine communications," *IEEE Transactions on Communications*, vol. 28, no. 9, pp. 1602-1607, Sep. 1980.

[16] M. Lanzagorta, "Underwater communication channels," in *Underwater communications*, San Rafael, CA, USA: Morgan & Claypool, 2012, ch. 3, sec. 3.5, pp. 32-40.

[17] W. D. Wilson, "Equation for the speed of sound in sea water," *The Journal of the Acoustical Society of America*, vol. 32, no. 10, pp. 1357-1357, Oct. 1960.

[18] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 84-89, Oct. 2009.

[19] J. Muth, "Building a 'deeper' understanding of underwater optical communications," *Laser Focus World*, vol. 53, no. 5, pp. 37-40, May. 2017.

[20] "35.100 Open Systems Interconnection (OSI)." ISO.org. https://www.iso.org/ics/35.100/x/ (accessed Sep. 18, 2020).

[21] J. Rice, B. Creber, C. Fletcher, P. Baxley, K. Rogers, K. McDonald et al., "Evolution of Seaweb underwater acoustic networking," in *Oceans*, Providence, RI, USA, Sep. 11-14, 2000, pp. 2007-2017.

[22] E. Voudouri-Maniati, "Multiuser robust CDMA detection for underwater acoustic communication channels," in *Oceans*, Biloxi, MI, USA, Oct. 29-31, 2002, pp. 612-618.

[23] M. Stojanovic and L. Freitag, "Wideband underwater acoustic CDMA: Adaptive multichannel receiver design," in *Oceans*, Washington, DC USA, Sep. 17-23, 2005, pp. 1508-1513.

[24] M. Stojanovic and L. Freitag, "Multichannel detection for wideband underwater acoustic CDMA communications," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 3, pp. 685-695, Jul. 2006.

[25] H. Tan and W. K. Seah, "Distributed CDMA-based MAC protocol for underwater sensor networks," in *Conference on Local Computer Networks*, Dublin, Ireland, Oct. 15-18, 2007, pp. 26-36.

[26] J. Kim, J. Lee, Y. Jang, K. Son, and H. Cho, "A CDMA-based MAC protocol in tree-topology for underwater acoustic sensor networks," in *Conference on Advanced Information Networking and Applications*, Bradford, UK, May. 26-29. 2009, pp. 1166-1171.

[27] D. Pompili, T. Melodia, and I. F. Akyildiz, "A CDMA-based medium access control for underwater acoustic sensor networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1899-1909, May. 2009.

[28] N. Chirdchoo, W. Soh, and K. C. Chua, "MU-Sync: a time synchronization protocol for underwater mobile networks," in *Workshop on Underwater Networks*, San Francisco, CA, USA, Sep. 15, 2008, pp. 35-42.

[29] F. Lu, D. Mirza, and C. Schurgers, "D-sync: Doppler-based time synchronization for mobile underwater sensor networks," in *Workshop on Underwater Networks*, Woods Hole, MA, USA, Sep. 30, Oct. 1, 2010 pp. 1-8.

[30] J. Liu, Z. Zhou, Z. Peng, J. Cui, M. Zuba, and L. Fiondella, "Mobi-Sync: efficient time synchronization for mobile underwater sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 2, pp. 406-416, Feb. 2013.

[31] Z. Li, Z. Guo, H. Qu, F. Hong, P. Chen, and M. Yang, "UD-TDMA: A distributed TDMA protocol for underwater acoustic sensor network," in *Conference on Mobile Adhoc and Sensor Systems*, Macau, China, Oct. 12-15, 2009, pp. 918-923.

[32] Y. Chen, C. Lien, S. Chuang, and K. Shih, "DSSS: a TDMA-based MAC protocol with dynamic slot scheduling strategy for underwater acoustic sensor networks," in *Oceans*, Santander, Spain, Jun. 6-9, 2011, pp. 1-6.

[33] R. Diamant, P. Casari, and M. Zorzi, "A TDMA-based MAC protocol exploiting the near-far effect in underwater acoustic networks," in *Oceans*, Shanghai, China, Apr. 10-13, 2016, pp. 1-5.

[34] N. Abramson, "Development of the ALOHANET," *IEEE transactions on Information Theory*, vol. 31, no. 2, pp. 119-123, Mar. 1985.

[35] N. Abramson, "The ALOHA system: another alternative for computer communications," in *Fall Joint Computer Conference*, Houston, TX, USA, Nov. 17-19, 1970, pp. 281-285.

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

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

[38] 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, Dec. 1975.

[39] P. Karn, "MACA-a new channel access method for packet radio," in *ARRL/CRRL Amateur radio computer networking conference*, London, Ontario. Canada. Sep. 22, 1990, pp. 134-140.

[40] B. Peleato and M. Stojanovic, "Distance aware collision avoidance protocol for ad-hoc underwater acoustic sensor networks," *IEEE Communications Letters*, vol. 11, no. 12, pp. 1025-1027. Dec. 2007.

[41] X. Guo, M. R. Frater, and M. J. Ryan, "A propagation-delay-tolerant collision avoidance protocol for underwater acoustic sensor networks," in *Oceans*, Singapore, Singapore, May. 16-19, 2006, pp. 1-6.

[42] H. Ng, W. Soh, and M. Motani, "MACA-U: A media access protocol for underwater acoustic networks," in *Globecom*, New Orleans, LO, USA, Nov. 30, 4 Dec, 2008, pp. 1-5.

[43] 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 *Oceans*, Shanghai, China, Apr. 10-13, 2016, pp. 1-5.

[44] Y. Noh, U. Lee, S. Han, P. Wang, D. Torres, J. Kim et al., "DOTS: A propagation delay-aware opportunistic MAC protocol for mobile underwater networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 4, pp. 766-782, Jan. 2014.

[45] J. Lee, M. Riess, S. Moser, and F. Slomka, "An Adaptive MAC Protocol for Underwater Mobile Ad-Hoc Networks," in *Oceans*, Kobe, Japan, May. 28-31, 2018, pp. 1-5.

[46] Y. Zhang, H. Chen, and W. Xu, "A Load-adaptive CSMA/CA MAC Protocol for Mobile Underwater Acoustic Sensor Networks," in *Conference on Wireless Communications and Signal Processing*, Hangzhou, China, Oct. 18-20, 2018, pp. 1-7.

[47] M. Gao, W. Li, and J. Li, "Performance analysis of a JSW-based MAC protocol for mobile underwater acoustic networks," in *Conference on Signal Processing and Communication Systems*, Cairns, QLD, Australia, Dec. 14-16, 2015, pp. 1-6.

[48] A. Cho, C. Yun, Y. Lim, and Y. Choi, "Asymmetric propagation delay-aware TDMA MAC protocol for mobile underwater acoustic sensor networks," *Applied Sciences*, vol. 8, no. 6, pp. 962, Jun. 2018.

[49] 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 *Conference on Cyber Technology in Automation, Control, and Intelligent Systems*, Shenyang, China, Jun. 8-12, 2015, pp. 1275-1280.

[50] 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, Nov. 2017.

[51] B. Marr. "27 Incredible Examples of AI and Machine Learning in Practice." Forbes.com. https://www.forbes.com/sites/bernardmarr/2018/04/30/27-incredible-examples-of-ai-and-machine-learning-in-practice (accessed Sep. 18, 2020).

[52] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*, Cambridge, UK: University of Cambridge Engineering Department, 1994.

[53] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's college, London, UK, 1989.

[54] J. N. Tsitsiklis, "Asynchronous stochastic approximation and Q-learning," *Machine Learning*, vol. 16, no. 3, pp. 185-202, Sep. 1994.

[55] N. B. Karayiannis and A. N. Venetsanopoulos, "Fast learning algorithms for neural networks," in *Artificial Neural Networks*, Switzerland: Springer, 1993, pp. 141-193. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4757-4547-4_4

[56] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, Jul. 2006.

[57] S. Galzarano, A. Liotta, and G. Fortino, "QL-MAC: a Q-learning based MAC for wireless sensor networks," in *Conference on Algorithms and Architectures for Parallel Processing*, Vietri sul Mare, Italy, Dec. 18-20, 2013, pp. 267-275.

[58] T. V. Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *Conference on Embedded networked sensor systems*, Los Angeles, CA USA, Nov. 5-7, 2003, pp. 171-180.

[59] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in *Conference of the Computer and Communications Societies*, New York, NY, USA, Jun. 23-27, 2002, pp. 1567-1576.

[60] Z. Liu and I. Elhanany, "RL-MAC: A QoS-aware reinforcement learning based MAC protocol for wireless sensor networks," in *Conference on Networking, Sensing and Control*, Ft. Lauderdale, FL, USA, Apr. 23-25, 2006, pp. 768-773.

[61] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems, in *Conference on Artificial Intelligence / Innovative Application of Artificial Intelligence*, Burgress Drive Menlo Park, CA, USA, Jul. 1998, pp. 746-752.

[62] Y. Tang, D. Grace, T. Clarke, and J. Wei, "Multichannel non-persistent CSMA MAC schemes with reinforcement learning for cognitive radio networks," in *Symposium on Communications and Information Technologies*, Hangzhou, China, Oct. 12-14, 2011, pp. 502-506.

[63] S. H. Park, P. D. Mitchell, and D. Grace, "Performance of the ALOHA-Q MAC Protocol for Underwater Acoustic Networks," in *Conference on Computing, Electronics & Communications Engineering*, Southend, UK, Aug. 16-17, 2018, pp. 189-194.

[64] N. Javaid, O. A. Karim, A. Sher, M. Imran, A. U. Yasar, and M. Guizani, "Q-Learning for energy balancing and avoiding the void hole routing protocol in underwater sensor networks," in *Wireless Communications & Mobile Computing Conference*, Limassol, Cyprus, Jun. 25-29, 2018, pp. 702-706.

[65] V. D. 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 Journal on Selected Area in Communication*, vol. 7, no. 11, pp. 2634-2647, Nov. 2019.

[66] Z. Jin, Q. Zhao, and Y. Su, "RCAR: A reinforcement-learning-based routing protocol for congestion-avoided underwater acoustic sensor networks," *IEEE Sensors Journal*, vol. 19, no. 22, pp. 10881-10891, Nov. 2019.

[67] X. Li, X. Hu, W. Li, and H. Hu, "A multi-agent reinforcement learning routing protocol for underwater optical sensor networks," in *Conference on Communications*, Shanghai, China, May. 20-14, 2019, pp. 1-7.

[68] 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, Jun. 2019.

[69] "RFC 1122 – Requirements for Internet Hosts: Communication Layers." Tools.iets.org. https://tools.ietf.org/html/rfc1122 (accessed Sep. 18, 2020).

[70] 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, May. 2013.

[71] S. Russell and P. Norvig, *Artificial intelligence: a modern approach*, 3rd ed. Essex, UK: Pearson, 2016.

[72] L. Wang, C. Lin, K. Chen, and Y. Zhang, "A learning-based ALOHA protocol for underwater acoustic sensor networks," in *Conference on Underwater Networks & Systems*, Shenzhen, China, Dec. 3-5, 2018, pp. 1-2.

[73] X. Geng and Y. R. Zheng, "MAC protocol for underwater acoustic networks based on deep reinforcement learning," in *Conference on Underwater Networks & Systems*, Atlanta, GA, USA, Oct. 23-25, 2019, pp. 1-5.

[74] X. Ye and L. Fu, "Deep reinforcement learning based MAC protocol for underwater acoustic networks," in *Conference on Underwater Networks & Systems*, Atlanta, GA, USA, Oct. 23-25, 2019, pp. 1-5.

[75] Y. Yu, T. Wang, and S. C. Liew, "Deep-reinforcement learning multiple access for heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1277-1290, Jun. 2019.

[76] Y. Chu, P. D. Mitchell, and D. Grace, "ALOHA and Q-learning based medium access control for wireless sensor networks," in *Symposium on Wireless Communication Systems*, Paris, France, Aug. 28-31, 2012, pp. 511-515.

[77] Y. Chu, P. Mitchell, and D. Grace, "Reinforcement learning based ALOHA for multi-hop wireless sensor networks with informed receiving," in *Conference on Wireless Sensor Systems*, London, UK, Jun. 18-19, 2012.

[78] Y. Yan, P. Mitchell, T. Clarke, and D. Grace, "Distributed frame size selection for a Q learning based Slotted ALOHA protocol," in *Symposium on Wireless Communication Systems*, Ilmenau, Germany, Aug. 27-30, 2013, pp. 1-5.

[79] S. Kosunalp, P. D. Mitchell, D. Grace, and T. Clarke, "Practical implementation issues of reinforcement learning based ALOHA for wireless sensor networks," in *Symposium on Wireless Communication Systems*, Ilmenau, Germany, Aug. 27-30, 2013, pp. 1-5.

[80] "HS Communication and Positioning Devices." Evologics.de. https://evologics.de/acoustic-modem/hs (accessed Sep. 18, 2020).

[81] 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, Nov. 2019.

[82] 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, Mar. 2000.

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

[84] S. Kosunalp, P. D. Mitchell, D. Grace, and T. Clarke, "Practical implementation and stability analysis of ALOHA-Q for wireless sensor networks," *ETRI Journal*, vol. 38, no. 5, pp. 911-921, Oct. 2016.

[85] "USMART – Smart dust for large scale underwater wireless sensing." EPSRC.org. https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/P017975/1 (accessed Sep. 18, 2020).

[86] 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.

[87] J Cui, J. K, M. Geria, and S. Zhou, H, "The challenges of building mobile underwater wireless sensor networks for aquatic applications," *IEEE Network*, vol. 20, no. 3, pp. 12-18, May. Jun. 2005.

[88] P. Smith, D. Hutchison, J. P. Sterbenz, M. Scholler, A. Fessi, M. Karaliopoulos et al., "Network resilience: a systematic approach," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 88-97, Jul. 2011.

[89] K. 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.

[90] "18/34 Communication and Positioning Devices." Evologics.de. https://evologics.de/acoustic-modem/18-34 (accessed Sep. 18, 2020).

[91] M. Lewis, S. Neill, P. Robins, M. Hashemi, and S. Ward, "Characteristics of the velocity profile at tidal-stream energy sites," *Renewable Energy*, vol. 114, pp. 258-272, Dec. 2017.

[92] A. Caruso, F. Paparella, L. F. Vieira, M. Erol, and M. Gerla, "The meandering current mobility model and its impact on underwater mobile sensor networks," in *Conference on Computer Communications*, Phoenix, AZ, USA, Apr. 13-18, 2008, pp. 221-225.

[93] J. Watson and O. Zielinski, "Underwater hyperspectral imagery to create biogeochemical maps of seafloor properties," in *Subsea optics and Imaging*, Cambridge, UK: Woodhead Publishing, 2013, ch. 20, sec. 20.3.2, pp. 516.

[94] G. Qiao, S. Gan, S. Liu, and Q. Song, "Self-interference channel estimation algorithm based on maximum-likelihood estimator in in-band full-duplex underwater acoustic communication system," *IEEE Access*, vol. 6, pp. 62324-62334, Oct. 2018.

[95] L. Shen, B. Henson, Y. Zakharov, and P. Mitchell, "Digital self-interference cancellation for full-duplex underwater acoustic systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 1, pp. 192-196, Jan. 2020.

[96] R. Wang, A. Yadav, E. A. Makled, O. A. Dobre, R. Zhao, and P. K. Varshney, "Optimal power allocation for full-duplex underwater relay networks with energy harvesting: A reinforcement learning approach," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 223-227, Feb. 2019.

[97] L. Shen, B. Henson, Y. Zakharov, and P. D. Mitchell, "Adaptive nonlinear equalizer for full-duplex underwater acoustic systems," *IEEE Access*, vol. 8, pp. 108169-108178, Jun. 2020.

[98] C. Li, Y. Xu, Q. Wang, B. Diao, Z. An, Z. Chen et al., "FDCA: A full-duplex collision avoidance MAC protocol for underwater acoustic networks," *IEEE sensors journal*, vol. 16, no. 11, pp. 4638-4647, Mar. 2016.

[99] F. Qu, H. Yang, G. Yu, and L. Yang, "In-band full-duplex communications for underwater acoustic networks," *IEEE Network*, vol. 31, no. 5, pp. 59-65, Sep. 2017.

[100] Y. Yan, P. Mitchell, T. Clarke, and D. Grace, "Adaptation of the ALOHA-Q protocol to Multi-hop Wireless Sensor Networks," in *European Wireless Conference*, Barcelona, Spain, May. 14-16, 2014, pp. 1-6.

[101] "The Acoustics Toolbox is Distributed under the GNU Public License." Oalib.hlsresearch.com. http://oalib.hlsresearch.com/AcousticsToolbox (accessed Sep. 18, 2020).

[102] R. Zhao, M. Li, and W. Bai, "Underwater acoustic networks environment simulation with combination of BELLHOP and OPNET modeler," in *Oceans*, Aberdeen, UK, Jun. 19-22, 2017, pp. 1-4.

[103] J. Llor, M. P. Malumbres, and P. Garrido, "Performance evaluation of underwater wireless sensor networks with OPNET," in *Conference on Simulation Tools and Techniques*, Barcelona, Spain, Mar. 21-25, 2011, pp. 19-26.

[104] N. Mastronarde, J. Modares, C. Wu and J. Chakareski, "Reinforcement learning for energy-e delay-sensitive CSMA/CA scheduling", in *Globecom*, Washington, DC, USA, Dec. 4-6, 2016, pp. 1-7.

[105] V. D. Valerio, C. Petrioli, L. Pescosolido and M. V. Shaar, "A reinforcement learning-based data-link protocol for underwater acoustic communications," in *Conference on Underwater Networks & Systems*, Arlington, VA, USA, Oct. 22-24, 2015, pp. 1-5.

[106] "IEEE Access." ieeeaccess.ieee.org https://ieeeaccess.ieee.org/ (accessed Mar. 3, 2020).

[107] L. Kleinrock, *Queueing systems volume 1: Theory*, New York, NY, USA: Wiley 1975.

[108] "Riverbed Modeler.", support.riverbed.com

https://support.riverbed.com/bin/support/static/5aloa5c31m2bji06sh8hdcta1a/html/gefmuvj461ej1ajt0tn0j3tbd4/modeler/wwhelp/wwhimpl/js/html/wwhelp.htm (accessed Sep. 24, 2020).

[109] C.L. Fullmer and J.J. Garcia-Luna-Aceves, "Floor Acquisition Multiple Access (FAMA) for Packet-Radio Networks," Proc. ACM SIGCOMM, 1995.

[110] C.L. Fullmer and J.J. Garcia-Luna-Aceves, “Analysis of Aloha Protocols for Underwater Acoustic Sensor Networks,” Proc. Workshop UnderWater Networks (WUWNet), 2006