PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
MOHAMED KHIDER UNIVERSITY OF BISKRA

Faculty of Exact Sciences, Natural and Life Sciences
Computer Science Department.

THESIS

Presented by

Mohamed SAHRAOUI

To obtain the degree of

PHD OF SCIENCE

Specialty : COMPUTER SCIENCE

Option : COMPUTER SYSTEMS ENGINEERING

Theme

Multi-channel Communication in Wireless Networks

Defended on Date 14/02/2022

Jury:

President:	Foudil CHERIF	Prof	University of Biskra
Supervisor:	Azeddine BILAMI	Prof.	University of Batna2
Examinators:	Abdelhamid DJEFFAL	Prof	University of Biskra
	Rachid SEGHIR	Prof	University of Batna2

Multi-channel Communication in Wireless Networks

Abstract

Multi-channel communication has been developed to overcome some limitations related to the throughput and delivery rate which become necessary for many applications that require sufficient bandwidth to transmit a large amount of data in Wireless Networks (WNs) such as multimedia communication. However, the requirement of frequent negotiation for the channels assignment process incurs extra-large communication overhead and collisions, which results in the reduction of both communication quality and network lifetime. This effect can play an important role in the performance deterioration of certain WNs types, especially the Wireless Sensor Networks (WSNs) since they are characterized by their limited resources. This work addresses the improvement of communication in multi-channel WSNs. Consequently, four protocols are proposed. The first one is the Multi-Channel Scheduling Protocol (MCSP) for wireless personal networks IEEE802.15.4, which focuses on overcoming the collisions problem through a multi-channel scheduling scheme. The second protocol is the Energy-efficient Reinforcement Learning (RL) Multi-channel MAC (ERL MMAC) for WSNs, which bases on the enhancement of the energy consumption in WSNs by reducing collisions and balancing the remaining energy between the nodes using a single-agent RL. The third work is the proposition of a new heuristically accelerated RL protocol named Heuristically Accelerated Reinforcement Learning approach for Channel Assignment (HARL CA) for WSNs to reduce the number of learning iterations in an energy-efficient way taking into account the bandwidth aspect in the scheduling process. Finally, the fourth contribution represents a proposition of a new cooperative multi-agent RL approach for Channel Assignment (CRLCA) in WSNs, which improves cooperative learning using an accelerated learning model, and overcomes the extra communication overhead problem of the cooperative RL using a new method for self-scheduling and energy balancing. The proposed approach is performed through two algorithms SCRLCA and DCRLCA for Static and Dynamic performance respectively. The proposed protocols and techniques have been successfully evaluated and show outperformed results in different cases through several experiments.

Keywords: Multi-channel; Wireless networks; Wireless sensor networks; IEEE802.15.4; Reinforcement learning; Heuristically accelerated reinforcement learning; Cooperative reinforcement learning.

الاتصال متعدد القنوات في الشبكات اللاسلكية

ملخص

تم تطوير الاتصال متعدد القنوات للتغلب على بعض القيود المتعلقة بمعدل الإنتاجية والتسليم والتي أصبحت ضرورية للعديد من التطبيقات التي تتطلب نطاقًا ترددیًا كافيًا لنقل كمية كبيرة من البيانات في الشبكات اللاسلكية مثل اتصالات الوسائل الاعلامية المتعددة.

ومع ذلك ، فإن شرط التفاوض المتكرر لتخصيص القنوات ينطوي على نفقات اتصال كبيرة جدًا وتصادمات ، مما يؤدي إلى تقليل جودة الاتصال وعمر الشبكة. يمكن أن يلعب هذا التأثير دورًا مهمًا في تدهور أداء بعض أنواع الشبكات اللاسلكية وخاصة شبكات الاستشعار اللاسلكية نظرًا لأنها تتميز بقيودها.

يتناول هذا العمل تحسين الاتصال في شبكات الاستشعار اللاسلكية متعددة القنوات. ومن ثم ، تم اقتراح اربعة بروتوكولات: أولاً ، بروتوكول الجدولة متعدد القنوات للشبكات الشخصية اللاسلكية IEEE802.15.4 ، والذي يركز على التغلب على مشكلة التصادمات عن طريق مخطط الجدولة متعدد القنوات. البروتوكول الثاني هو التعلم التعزيزي متعدد القنوات (ERL MMAC) الموفر للطاقة لشبكات الاستشعار اللاسلكية، والذي يعتمد على تعزيز استهلاك الطاقة في شبكات الاستشعار اللاسلكية من خلال تقليل الاصطدامات وموازنة الطاقة المتبقية بين العقد في استخدام نهج وكيل واحد للتعلم التعزيزي. العمل الثالث هو اقتراح بروتوكول للتعلم التعزيزي جديد مُسرَّع ارشاديا المسمى منهج التعلم التعزيزي المُسَرَّع ارشاديا لتعيين القنوات في شبكات WSN (HARL CA) لتقليل عدد التكرارات التعليمية بطريقة موفرة للطاقة مع مراعاة جانب النطاق الترددي في الجدولة. أخيرًا، تمثل المساهمة الرابعة اقتراحًا لنهج جديد للتعلم التعزيزي التعاوني متعدد الوكلاء لتعيين القنوات (CRLCA) في شبكات الاستشعار اللاسلكية، مما يحسن التعلم التعاوني باستخدام نموذج للتعلم السريع، ويتغلب على مشكلة الاتصالات الإضافية في التعلم التعزيزي التعاوني باستخدام طريقة جديدة للجدولة الذاتية وموازنة الطاقة. تم إجراء هذا المنهج من خلال خوارزميتين SCRLCA و DCRLCA للأداء الثابت والديناميكي على التوالي. تم تقييم البروتوكولات والتقنيات المقترحة بنجاح وظهرت نتائج متفوقة في حالات مختلفة من خلال عدة تجارب.

الكلمات الرئيسية:

متعدد القنوات؛ الشبكات اللاسلكية؛ شبكات الاستشعار اللاسلكية؛ IEEE802.15.4؛ تعزيز التعلم؛ التعلم التعزيزي المتسارع إرشاديًا؛ التعلم التعزيزي التعاوني.

La Communication Multi-canal dans les Réseaux Sans Fils

Résumé

La communication multicanal a été développée pour surmonter certaines limitations liées au débit et au taux de livraison qui sont nécessaires pour de nombreuses applications nécessitant une bande passante suffisante pour transmettre une grande quantité de données dans les Réseaux Sans Fils (RSF) telle que la communication multimédia. Cependant, l'exigence d'une négociation fréquente pour l'attribution des canaux entraîne une surcharge de communication et des collisions très importantes, ce qui entraîne une réduction à la fois de la qualité de la communication et de la durée de vie du réseau. Cet effet peut jouer un rôle important dans la détérioration des performances de certains types de RSF, en particulier les Réseaux de Capteurs Sans Fils (RCSF) car ils se caractérisent par leurs ressources limités. Ce travail porte sur l'amélioration de la communication dans les RCSF multi-canal. Par conséquent, quatre protocoles sont proposés : Premièrement, le protocole d'ordonnancement multicanal pour les réseaux personnels sans fil IEEE802.15.4 (MCSP), qui se concentre sur la résolution du problème des collisions par une méthode d'ordonnancement multicanal. Le deuxième travail est la proposition d'un protocole multi-canal basé sur l'Apprentissage par Renforcement (AR) écoénergitique (ERL MMAC) pour les RCSF, qui repose sur l'amélioration de la consommation d'énergie dans les RCSF, en réduisant les collisions et en équilibrant l'énergie restante entre les nœuds en utilisant l'approach AR basée agent unique. Le troisième travail représente la proposotion d'un nouveau protocole AR heuristiquement accéléré nommé Approche d'apprentissage par renforcement heuristiquement accéléré pour l'affectation de canaux dans les RCSF (HARL CA), dont le but est de réduire le nombre d'itérations dans la période d'apprentissage de manière économique en énergie en tenant compte de l'aspect bande passante dans le processus d'ordonnancement. Finalement, le quatrième contribution est une proposition d'une nouvelle approche d'AR multiagent coopérative pour l'attribution de canaux dans les RCSF (CRLCA), qui améliore l'apprentissage coopératif en utilisant un modèle d'apprentissage accéléré, et surmonte le problème de surcharge de communication supplémentaire du AR coopératif en utilisant une nouvelle méthode d'auto-ordonnancement et d'équilibrage d'énergie. L'approache proposée est réalisée à travers deux algorithmes SCRLCA et DCRLCA pour les performances statique et dynamique respectivement. Les protocoles et techniques proposés ont été évalués avec succès et montrent des résultats efficaces dans différents cas grâce à plusieurs expériences.

Mots-clés: Multi-canal; Réseaux sans fil; Réseaux de capteurs sans fil ; IEEE802.15.4; Apprentissage par renforcement; Apprentissage par renforcement heuristiquement accéléré; Apprentissage par renforcement coopératif.

Acknowledgements

I would like to express my sincere gratitude to those who gave me assistance and support during my Ph.D. study.

My special gratitude goes to my supervisor Pr. Azeddine BILAMI, who has spent dedicated time and efforts to train my research skills. This thesis would not have been possible without his encouragement, motivation, inspiration and guidance.

I am grateful to Pr. Foudil CHERIF, Pr. Abdelhamid DJEFFAL, and Pr. Rachid SEGHIR, for agreeing to be members of the exam committee for this thesis, and spending time in reading and reviewing. Their suggestions will be taken into account and will surely significantly improve the final version.

I am grateful for Pr.Abdelmalik TALEB-AHMED specifically and for IEMEN DOAE Laboratory of Hauts de France Polytechnic University in general for their welcome and support.

Last but not least, I wish to thank deeply my family for their love, encouragement and support.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Introduction

With its ubiquitous availability, versatility, and high reliability, wireless communication has revolutionized the world we live in. Mobile telephone services, television and radio broadcast, internet access and a large number of applications have changed the way people all over the globe interact. Nonetheless, these wide ranges of applications call for continuing increase in traffic demands and requirements. To cope for these needs, wireless communication technology must be continuously evolved towards greater capacity, higher flexibility, and better services. Hence, multichannel communication is used to improve the performance and overcome the limitations of traditional single-channel (Uni-channel) protocols commonly used in Wireless Networks (WNs) communication. The ultimate objective of multi-channel communication is to provide an alternative solution for many recent applications that required high throughput and delivery rate to exchange a huge amount of data in a short time such as multimedia applications. An example is given in Figure 1.1 which shows the improvement in throughput and delivery rate between the five nodes by the multiple using only two channels instead of one.

1.2 What is Multi-channel Communication?

Multichannel communication is the use of several non-overlapping channels rather than just one in simultaneous data exchange in order to increase throughput and delivery rate [1]. The non-overlapping channels are offered in different bands for each standard [2, 3]. Figure 1.2 shows an example of non-overlapping channels in two standards at 2.4GHz band: IEEE802.11 used for Wireless Local Area Network (WLAN) with 22 MHz for each channel, and IEEE802.15.4 used for Wireless Personal Area Networks (WPAN) with 2MHz for each channel.

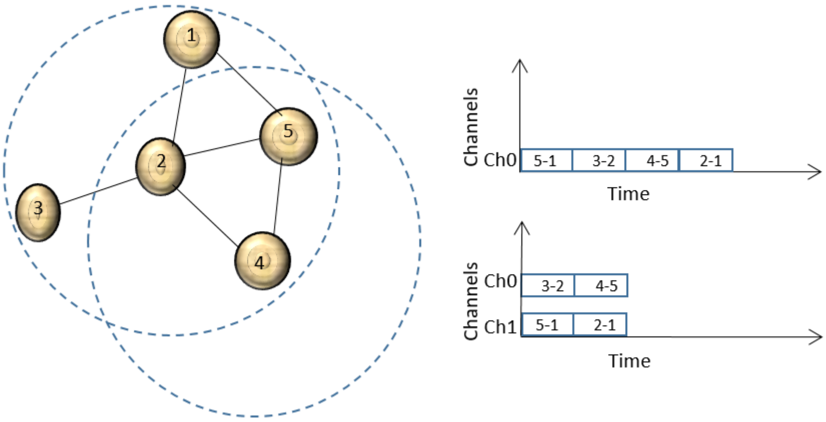


FIGURE 1.1: Uni-channel Vs Multi-channel communications.

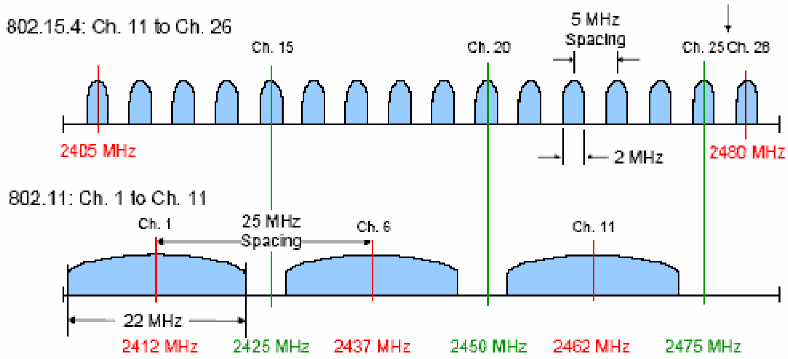


FIGURE 1.2: IEEE802.11 and IEEE802.15.4 non-overlapping channels at 2.4 GHz band.

As indicated previously, multichannel communication can be used in different types of wireless networks according to different strategies which depend on the specific characteristics of each network. Many WN types exist such as Wireless Sensor Networks (WSNs), WLAN, Worldwide Interoperability for Microwave Access (WiMAX) and Wireless Mesh Networks (WMNs). They are standardized by the different wireless standards respectively: IEEE802.15.4 [3], IEEE802.11 [2] and IEEE802.16 [4] except the WMNs that can work with the different wireless technologies [5]. Figure 1.3 shows the cited WN types.

Several multi-channel protocols have been proposed for each WN type in the literature [6]. Each protocol performs through three steps:

1. The network initialization step in which the different nodes exchange their basic information using the dedicated channel.

2. Channels assignment and time scheduling processes to schedule and manage the different offered channels among the different nodes of the network.

3. The data exchange between the network nodes using the established channels assignment and time schedule.

In the channels assignment process, the protocol follows one of the three possible strategies: static or fixed channel allocation, dynamic which changes the channel for each communication, or hybrid channel assignment strategy [7].

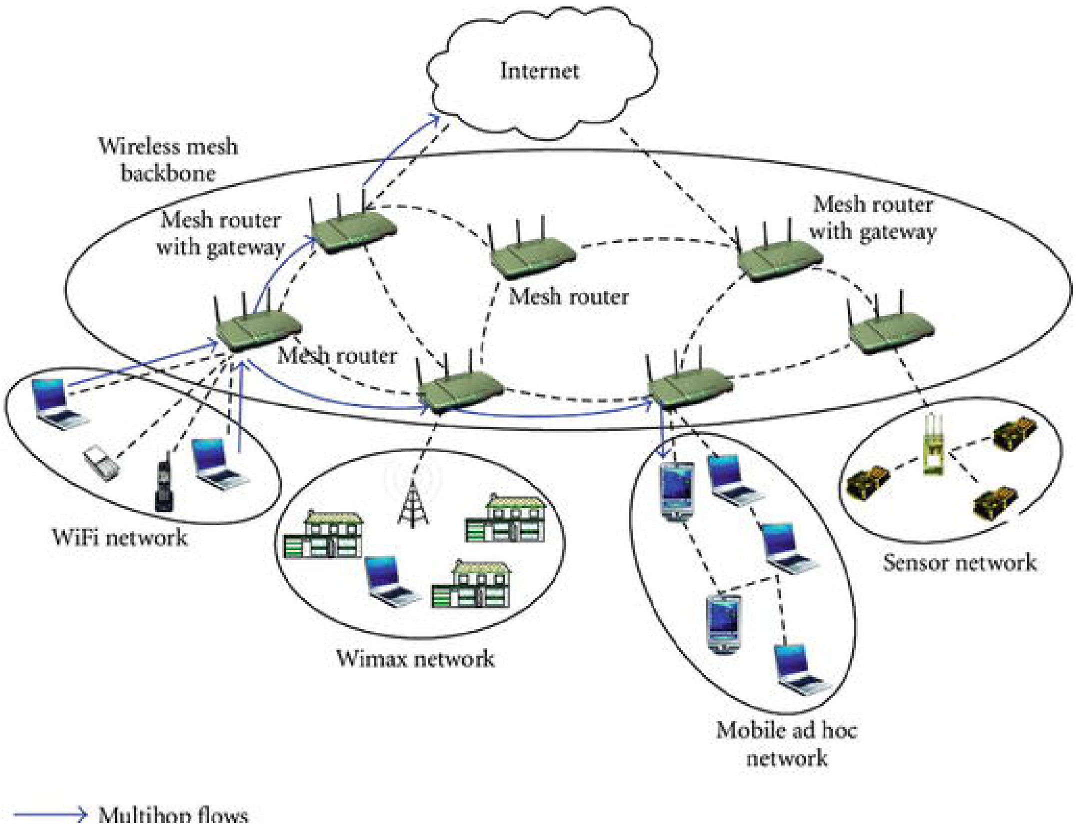


FIGURE 1.3: Some WN types.

The time scheduling process focuses on three strategies. It can be either synchronous, asynchronous or semi-synchronous [8].

☞ The synchronous strategy bases on the use of a shared clock to synchronize communication between the neighboring nodes.

☞ The asynchronous strategy bases on the use of Carrier Sense Multiple Access with Collision Avoidance (CSMA CA) method.

☞ The semi-synchronous, in which the different nodes negotiate the channels to be used using a shared one.

1.3 Motivation

In order to enhance the multi-channel communication in wireless networks, several works are proposed. They address all two problems: channels assignment and/or time scheduling processes in one of the WN types taking into account its characteristics.

In this thesis, we have addressed the two problems based on the following analysis.

1.3.1 An Analysis of Channels Assignment

The static strategy can give an optimal solution in collision-free multichannel communication, but it can not fully exploit the network resources since nodes must share information dynamically and be involved in the channel assignment and scheduling process. On the other side, the dynamic strategy is considered to be the best for optimally exploiting the network resources as the nodes can share information dynamically and manage their resources well, however, it requires a high amount of negotiations and thus communication overhead that can take most of the energy consumption which plays the first role in the decrease of the network lifetime, as it is shown in the Figure 1.4 that presents an experiment for different energy consumptions in a short wireless network carried out by the author of [9], which results in the highest level of the communication overhead energy consumption.

The hybrid strategy benefits from the advantages of the two previous strategies in such a way that each node selects a fixed channel but the assignment can be changed during the communication. However, the requirement of frequent negotiation between the neighboring nodes involves a significant amount of communication overhead and collisions, which leads to reduced communication quality and network lifetime. This effect can play an important role in the deterioration of the network performance especially in the WSNs since they are characterized by their limited resources and sensitive constraints.

This work focuses on the enhancement of the multichannel communication in both static and hybrid channels assignment for WSNs as an area of study since wireless networks present a set of specific wireless networks as explained before that we cannot address them simultaneously. The choice of WSN has come for two reasons:

- ☞ Firstly, there are few works that target this type of network in multi-channel communication compared to the other types.
- ☞ Secondly, for its importance in recent fields such as the Internet of thing (IoT) in which it has been widely used in such a way that the WSN represents the interface between the IoT and the real world.

1.3.2 An Analysis of Time Scheduling Process

The asynchronous strategy can only be used in static channels assignment since one transceiver can not listen to more than one channel at the same time. Thus, it keeps suffering from the same inconvenience related to poor management of resources. The semi-synchronous strategy is used to efficiently coordinate the multichannel operation and enhance channel utilization with avoidance of collisions and minimal channel coordination overhead, however, the synchronous strategy remains the best in energy consumption and time exploitation which present the most important factors in WSNs as a specific type of WNs [8]. Furthermore, the communicating nodes must ensure that the transmitter and receiver are on the same channel at the time of communication which requires time synchronization to avoid

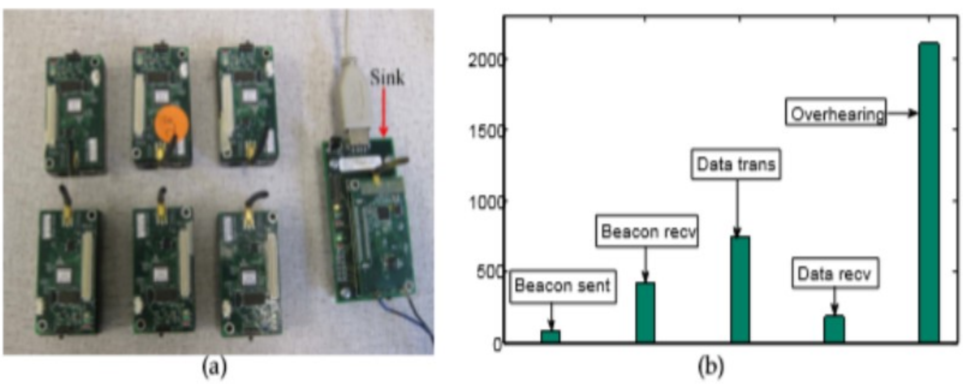


FIGURE 1.4: Different multi-channel energy consumption.

deafness problem that could arise because an intended receiver may currently be in communication on another channel of a third node [10].

Consequently, the schedule of both channels and links for wireless networks has been proved that be an NP-hard problem [11, 12]. Hence, some intelligent protocols have been proposed in the last years to overcome this challenge and accommodate parallel transmissions using multiple channels for the required increase in throughput and data delivery rate by ensuring communication reliability and better use of network resources with avoidance of collisions and minimal channel coordination overhead. These protocols are distinguished from the conventional multi-channel ones by the fact of using the information acquired from the external environment in the channel selection process to obtain approximately the best selection. They can be divided into two subcategories: game theory-based approach and reinforcement learning-based one. [13]. The game theory-based protocols are not desirable for WSNs since they need a high level of energy consumption and wasted time in their performance due to their complexity. However, the reinforcement learning approach takes the dominant place in distributed WSNs because of their fully distributed nature and their implementation which can take advantage of the routing information with less channel coordination overhead. Implementing such protocols can improve gradually the performances of WSNs based on the actions and their feedback analyses principle [14].

Nevertheless, reinforcement learning solutions require a number of iterations to obtain the best solution without using the collision avoidance mechanism, which in turn creates communication overhead and time-wasting.

In order to address the channels assignment and time scheduling problems and improve multi-channel communication in WSNs by focusing on intelligent methods to propose solutions that are efficient in energy consumption, we have based on the use of Reinforcement Learning (RL) approach in decentralized tree-based topology to perform an energy-efficient synchronous hybrid channel assignment strategies that are characterized by their advantages compared to the other strategies as it is mentioned before.

1.4 Goals

The major objective of this thesis is to investigate/improve multi-channel communication in wireless networks by focusing on wireless sensor networks as an area of work. The main challenge facing multi-channel wireless networks in general, and multi-channel wireless sensor networks in particular, is to provide effective solutions for channels assignment and time schedule in order to ensure communication reliability and better use of the network resources, however, those solutions should be based on improving energy consumption in multi-channel WSNs as the primary consideration to take into account since WSN nodes are constrained by their limited batteries. To achieve these primary objectives and guide this research, we have created a set of those objectives, which can be summarized as follows:

- ☞ To provide a brief overview of the multi-channel communication in wireless networks and bring out its importance as an alternative field to single-channel communication.
- ☞ To choose an area of work among the wireless network types presented by the multi-channel WSNs.
- ☞ To investigate the different proposed multi-channel solutions in this area focusing on intelligent ones since the problem of scheduling of both channels and links for wireless networks has been proved as an NP-hard problem.
- ☞ To propose improvement solutions with regard to the previously investigated solutions.

- ☞ To propose a new intelligent solution based on the Artificial Intelligent (AI) literature which improves the throughput and delivery rate through good managing of the network resources with enhancement in the energy consumption since the area of work is WSNs.

1.5 Major Contributions

In this thesis, we focus on intelligent solutions that have been solved certain multi-channel communication problems in WSNs and propose some other intelligent solutions in order to improve the communication in such type of network. Therefore, this thesis makes the following major contributions:

The first contribution is an overview presented by a survey of collection and investigation of the different intelligent solutions proposed in the literature for multi-channel WSNs. This work is focused on the intelligent solutions for the multi-channel WSNs by collecting, categorizing and comparing between these solutions which can be divided into two groups: GT-based and RL-based solutions as mentioned before.

Second, a new beacon collision avoidance algorithm named Multi-Channel Scheduling Protocol (MCSP) for Low Rate Wireless Personal Networks (LR WPANs) IEEE802.15.4 is proposed. The aim is the use the opportunity of multiple channels offered by the standard IEEE802.15.4 to assign them in a static manner in order to avoid the direct and indirect collisions of the beacon frames that play a crucial role in network synchronization and management.

Third, we developed an Energy-efficient method for Reinforcement Learning based Multi-channel MAC (ERL MMAC) that performs a hybrid channel assignment based on TDMA method for decentralized tree-based WSNs. The proposal focuses on the reduction of both communication overhead and collisions by using the least chosen default channel allocation in two hops rather than one hop in order to reduce as much as possible the conflict links in one side, and use of parent selection strategy rather than parent channel selection strategy to avoid the redundant data messages in the other side.

Forth, we proposed the Heuristically Accelerated Reinforcement Learning approach for Channel Assignment (HARL CA) algorithm to reduce the number of learning iterations in an energy-efficient way. The proposal considers the channel chosen by the other neighboring sender nodes as external information and uses it to accelerate the learning process and avoid collisions. Also, it takes into account the bandwidth of the used channel as an important factor in such a way that we developed a Bandwidth-based Scheduling scheme that assigns slot-times depending on the bandwidth of the used channel since the channels have different bandwidths.

Fifth, we proposed the Schedule-Based Cooperative Multi-agent Reinforcement Learning for Multi-channel Communication in Wireless Sensor Networks algorithm to reduce the number of learning iterations in an energy-efficient way. The proposal improves the cooperative RL by using an accelerated learning model and overcomes the extra communication overhead problem of the cooperative RL by using a new method for self-scheduling and energy balancing. the proposed approach is performed through two algorithms SCRLCA and DCRLCA for Static and Dynamic performance respectively.

The studies conducted during this work have produced conference papers and publications as follows:

- ☞ A conference paper entitled "Decentralized Intelligent Schemes for Channels Assignment In Multi-channel Wireless Sensor Networks: a Survey and Future works" at the $1_s t$ International Conference on Digitization and its Applications [13].

- ☞ A conference paper entitled "A new Reinforcement Learning based for Energy-efficient Multi-channel Data Gathering in Wireless Sensor Networks" at the IEEE 4th International Symposium on Informatics and its Applications (ISIA) [15].
- ☞ A published paper entitled "Multi-Channel Scheduling Protocol for Wireless Personal Networks IEEE802.15.4" at the International Journal of Soft Computing and Software Engineering [16].
- ☞ A published paper entitled "Heuristically Accelerated Reinforcement Learning for Channel Assignment in Wireless Sensor Networks" at the International Journal of Sensor Networks [17].
- ☞ A published paper entitled "Schedule-Based Cooperative Multi-agent Reinforcement Learning for Multi-channel Communication in Wireless Sensor Networks" at the Wireless Personal Communications journal [18].

There are still papers under review for conferences and journals

1.6 Thesis Structure

The list below shows the organization of the chapters that make up this thesis. In addition, a brief description of the topics dealt with in each chapter is given.

Chapter 1 shows a general overview of the communication in multi-channel wireless networks, the motivation, objectives, contributions, and structure of this thesis. In addition, a list of the published manuscripts that have been written during the course of the thesis is presented in this chapter.

Chapter 2 covers the necessary aspects of the WNs with a global view through defining the main related concepts. It then focuses on the WSNs as the interested WN type of this thesis.

Chapter 3 covers the necessary theoretical background of the multi-channel communication in WSNs through definitions and categorization of the different algorithms that have been proposed in the literature. It then focuses on the RL approach as our state of the art, with a global view through a discussed comparison between the RL-based protocols to give to the reader an overview of their similarities and differences.

Chapter 4 introduces our contributions presented by the four following protocols: first, the Multi-Channel Scheduling Protocol (MCSP) for wireless personal networks IEEE802.15.4, second, the Energy-efficient method for Reinforcement Learning based Multi-channel MAC (ERL MMAC), third, the Heuristically Accelerated Reinforcement Learning for Channel Assignment in wireless sensor networks (HARL CA), and finally, the schedule-Based Cooperative multi-agent Reinforcement Learning for Channel Assignment (CRLCA) in WSNs. The description of each protocol is performed by defining its different concepts, steps, techniques and algorithms. Also, the performance against other algorithms in the literature with analyzing of the different results is presented.

Chapter 5 summarizes the overall findings of this thesis and provides insights into our future research directions that emanate from it.

1.7 Conclusion

By improving multi-channel communication in WSNs, we give more and more opportunities to recent WSN-based applications that required a huge data transfer to keep forward.

In this chapter, we have briefly described the multi-channel communication in wireless networks, the motivations, the research goals, then outline the major contributions and thesis structure.

Chapter 2

Overview of Wireless Networks

2.1 Introduction

The wireless communication revolution is bringing fundamental changes to data networking telecommunication and is making integrated networks a reality. By freeing the user from the cord in its communications, wireless networks harbor the promise of fully distributed mobile computing and communications any time, anywhere. Numerous wireless services are also maturing and are poised to change the way and scope of communication, which documents the fast-growing areas of interest. This chapter covers and briefly describes the different aspects of this field and gives special consideration for WSNs as the working area of this thesis.

2.2 Definition

Wireless networks are networks that use wireless communication between the network nodes [19]. They involve the transmission of information over a distance using a wireless signal through electromagnetic waves based on wireless communication technologies and devices without the help of wires, cables or any other forms of electrical conductors. The evolution of wireless technology has given birth to many wireless devices and services through different wireless networks as shown in figure 2.1, which has led to many advancements in communications that become anywhere from a few meters (for example, a television's remote control) to thousands of kilometers (for example, radio broadcasting communication).

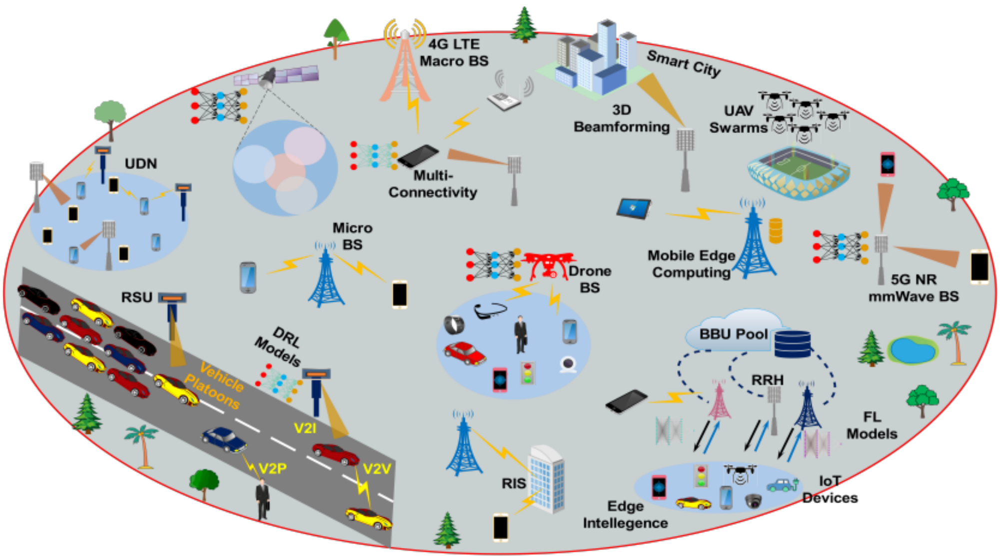


FIGURE 2.1: Illustration of the main wireless networks.

The first professional wireless network was developed under the brand ALOHAnet in 1969 at the University of Hawaii and became operational in June 1971. The first commercial wireless network was the WaveLAN product family, developed by the American National Cash Register (NCR) in 1986 [20]. Since the 80's decade, wireless networking and devices have seen rapid growth that has played an important role in replacing complex wired systems. Therefore, various wireless applications and devices are continued to be offered such as security systems, television remote control, Wi-Fi, Cell phones, wireless power transfer, computer interface devices, etc.

2.3 Wireless Networks Classification

In the literature, two main classifications of WNs are used, connectivity distance-based and simplex/duplex-based classifications. Consequently, in the first classification, the WNs can be classified as follows [19]:

- ☞ Wireless Personal Area Networks (WPANs): The devices in this type of WNs are connected within a relatively small area, that is generally within a person's reach. For example, Bluetooth, Wireless Body Area Networks (WBAN) and Zigbee that base on the IEEE 802.15.1, IEEE 802.15.6 and IEEE 802.15.4 standards respectively.

- ☞ Wireless Local Area Networks (WLANs) link two or more devices over a short distance that can be up to a few kilometers. As WPANs, the WLANs use two wireless distribution methods, either providing a connection through an Access Point (AP) or without it (Ad hoc mode). An example is the Wireless Fidelity (Wi-Fi) networks which are based on the standard IEEE 802.11.

- ☞ Wireless Metropolitan Area Networks (WMANs) are wireless network type that connects several wireless LANs. As an example, WiMAX (an acronym for Worldwide Interoperability for Microwave Access), that is described by the IEEE 802.16 standard.

- ☞ Wireless Metropolitan Area Networks (WMANs) are wireless networks that typically cover large areas, such as between neighboring towns and cities. The main examples are: Radio broadcasting, cellular (GSM, 4G and 5G) and satellite (Global Positioning System (GPS) and Television broadcasting) networks.

While in the second classification, the WNs can be classified as follows [21]:

- ☞ Simplex WNs that are one-way communication. As examples are Radio and television broadcast system.
- ☞ Half-duplex WNs which are two-way communication but not simultaneous. An example is walkie – talkie.
- ☞ Full-duplex WNs that are two-way communication but in a simultaneous manner. The best example of full-duplex is cellular systems.

2.4 Properties of Wireless Communication

There are numerous advantages of wireless communication compared to wired one, which can be summarized as follows [21, 22]:

- ☞ Cost: The cost of installing wires, cables and other infrastructure is eliminated in wireless communication and hence lowering the overall cost of the system compared to the wired communication systems that need to install a wired network and and run those wires through streets and buildings, which is a very difficult, costly and time-consuming job.
- ☞ Mobility: Mobility is the main advantage of the wireless communication systems. It offers the freedom to move around while still connected to the network.
- ☞ Ese of installation: The setup and installation of wireless communication network equipment and infrastructure is very easy as we need not worry about the hassle of cables. Also, the time required to set up a wireless system like a Wi-Fi network for example, is very less when compared to setting up a full cabled network.
- ☞ Reliability: Since there are no cables and wires involved in wireless communication, there is no chance of communication failure due to damage of these cables, which may be caused by environmental conditions, cable splice and natural diminution of metallic conductors.
- ☞ Disaster Recovery: In case of accidents due to fire, floods or other disasters, the loss of communication infrastructure in wireless communication systems can be minimal.

On the other side, wireless communication have some disadvantages that can be described as follows:

- ☞ Interference: Wireless Communication systems use open space as the medium for transmitting signals. As a result, there is a huge chance that radio signals from one wireless communication system or network might interfere with other signals. The best example is Bluetooth and Wi-Fi. Both these technologies use the 2.4GHz frequency for communication and when both of these devices are active at the same time, there is a chance of interference.
- ☞ Absorption and reflection: Some materials cause absorption of electromagnetic waves, preventing them from reaching the receiver, in other cases, particularly with metallic or conductive materials reflection occurs. This can cause dead zones where no reception is available. Aluminum foiled thermal isolation in modern homes can easily reduce indoor mobile signals by 10 dB frequently leading to complaints about the bad reception of long-distance rural cell signals.

- ☞ Multi-path fading: In multi-path fading, two or more different routes taken by the signal, due to reflections, can cause the signal to cancel out each other at certain locations, and to be stronger in other places.
- ☞ Hidden node problem: In a hidden node problem Station A can communicate with Station B. Station C can also communicate with Station B. However, Stations A and C cannot communicate with each other, but their signals can interfere at B. The hidden node problem occurs in some types of networks when a node is visible from an AP, but not from other nodes communicating with that AP. This leads to the problem of collisions in the medium access process.
- ☞ Security: One of the main concerns of wireless communication is the security of the data. Since the signals are transmitted in open space, it is possible that an intruder can intercept the signals and copy sensitive information.
- ☞ Health Concerns: Continuous exposure to any type of radiation can be hazardous. Even though the levels of RF energy that can cause the damage are not accurately established, it is advised to avoid RF radiation to the maximum.

2.5 Basic Elements of Wireless Communication System

A typical Wireless Communication System can be divided into three elements: the Transmitter, the Channel and the Receiver. Figure 2.2 shows the block diagram of wireless communication system [22].

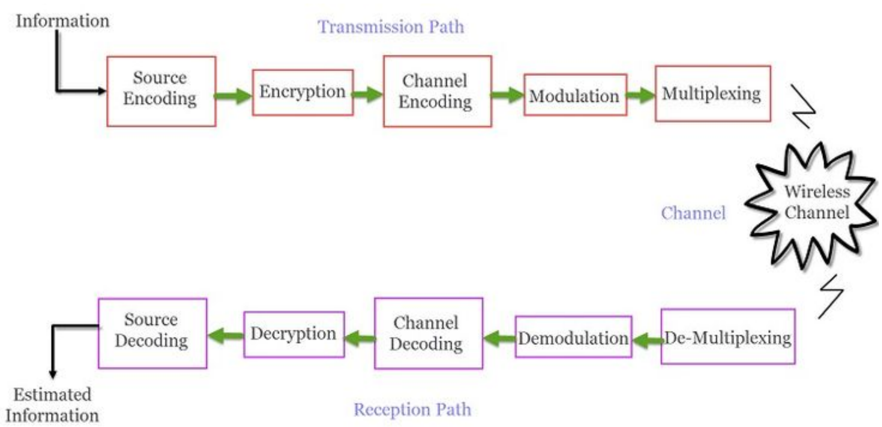


FIGURE 2.2: Elements of wireless communication system.

2.5.1 The Transmission Path

A typical transmission path of a Wireless Communication System consists of Encoder, Encryption, Modulation and Multiplexing. The signal from the source is passed through a Source Encoder, which converts the signal into a suitable form for applying signal processing techniques.

The redundant information from the signal is removed in this process in order to maximize the utilization of resources. This signal is then encrypted using an Encryption Standard so that the signal and the information are secured and don't allow any unauthorized access.

Channel Encoding is a technique that is applied to the signal to reduce the impairments like noise, interference, etc. During this process, a small amount of redundancy is introduced to the signal so that it becomes robust against noise. Then the signal is modulated using a suitable Modulation Technique (like PSK, FSK and QPSK, etc.) , so that the signal can be easily transmitted using the antenna.

The modulated signal is then multiplexed with other signals using different Multiplexing Techniques like Time Division Multiplexing (TDM) or Frequency Division Multiplexing (FDM) to share the valuable bandwidth.

2.5.2 The Channel

The channel in Wireless Communication indicates the medium of transmission of the signal. A wireless channel is unpredictable and also highly variable and random in nature. As mentioned before, a channel may be subject to interference, distortion, noise, scattering, etc. and the result is that the received signal may be filled with errors.

2.5.3 The Reception Path

The job of the Receiver is to collect the signal from the channel and reproduce it as the source signal. The reception path of a Wireless Communication System comprises Demultiplexing, Demodulation, Channel Decoding, Decryption and Source Decoding. From the components of the reception path, it is clear that the task of the receiver is just the inverse to that of the transmitter.

The signal from the channel is received by the Demultiplexer and is separated from other signals. The individual signals are demodulated using appropriate Demodulation Techniques and the original message signal is recovered. The redundant bits from the message are removed using the Channel Decoder.

Since the message is encrypted, the Decryption of the signal removes the security and turns it into a simple sequence of bits. Finally, this signal is given to the Source Decoder to get back the original transmitted message or signal.

2.6 Channel Access techniques

In wireless communication, radio frequency is crucial since it is the means for transferring information between two or more connections. Radio frequency usage varies from one application to another. For each wireless application, a continuous block of radio frequency, called a frequency spectrum or a frequency band is required. It is the International Telecommunication Union (ITU) that coordinates the shared global use of the available frequency spectrum [23].

The channel access process is performed through two main operations, the use of the hole frequency spectrum such as the Code division multiple-access (CDMA) method where multiple codes are used by the communicating nodes, the nodes having same code can communicate with each other, or the use of multiple channels by the fact that the frequency spectrum is partitioned into a set of channels N, all with the same frequency bandwidth Δ. These channels are indexed consecutively as $F = f_1, ..., f_N$ where N is calculated using 2.1.

$$N = (f_{max} - f_{min})/\Delta \tag{2.1}$$

Hence two main operations are needed, channel assignment operation and/or time schedule one.

In the channels assignment operation, the protocol follows one of the three possible strategies: static, dynamic or hybrid channel assignment strategy [7].

☞ In the static strategy, the nodes of the same collision domain are divided via generally a central unit into fixed and separate sets according to the number of channels.

☞ The dynamic strategy requires that the communicating nodes start by negotiating the channel to be used before each data transmission at all times.

☞ The hybrid strategy performs in such a way that each node selects a fixed channel but the assignment can be changed during the communication. There are two methods for doing this: sender-based or receiver-based methods.

  - In the sender-based method, each node has its own fixed channel used in the data sending operation and the receiving nodes must change their channels to that of the sender node.
  - In the receiver-based method, each node has its own fixed channel used in the data reception operation and the sending nodes must change the channel to that of the receiver node.

The time scheduling operation focuses on three strategies. It can be either synchronous, asynchronous or semi-synchronous [8].

☞ The synchronous strategy bases on using a shared clock to synchronize communication between the neighboring nodes. Hence, two methods are used: The beacon-based method and the TDMA (Time Division Multiple Access) one.

  - The beacon-based method, usually used by the centralized protocols, is based on the use of beacon messages delivered periodically by a central unit on a dedicated channel to synchronize the communication between the nodes.
  - The TDMA method divides the time into consecutive frame periods which are in turn divided into a fixed number of slots with the same duration enough at least to send and receive one data message on a coordinated channel.

☞ The asynchronous strategy addresses the coordination problem between the nodes using CSMA CA (Carrier Sense Multiple Access with Collision Avoidance) method.

☞ The semi-synchronous strategy establishes a dynamic interval within which all sender-receiver nodes can perform their negotiation using the control channel then data transmission on the coordinated data channel, by the end of the interval, they will return to the control channel and that interval is referred as rendezvous interval.

2.7 Overview of Wireless Sensor Networks

2.7.1 Definition

The Wireless Sensor Network (WSN) is formed by a large number of sensors deployed on an area to monitor some events and sense the environment such as temperature, sound, and pressure. The obtained data are, then, relayed through the network to a central location or a Sink which is generally more powerful than the other nodes, and where this information can be analyzed or transmitted via the Internet [24]. The development of Micro-Electro-Mechanical Systems (MEMS) brings various capabilities to sensor nodes. Therefore, WSNs have been used in a wide range of applications such as in military environments, medical treatments, industry, security and environmental monitoring.

Typically, a sensor node is a miniature device that includes four main components: a sensing unit for data acquisition, a micro-controller for local data processing and memorization, a communication unit to allow the transmission/reception of data and, finally, a power source which is usually a small battery as illustrated in figure 2.3 where an example of a WSN and the typical architecture of sensor node are shown. .

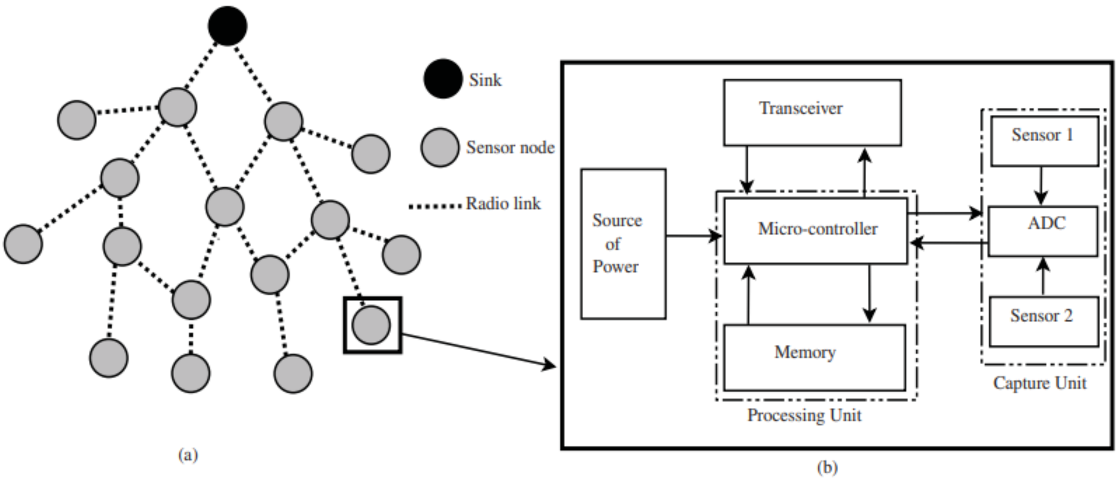


FIGURE 2.3: (a) An example of a Wireless Sensor Network and (b) The typical architecture of sensor node.

2.7.2 Constraints

Considering the size and cost of a sensor node, the processor, storage and data rate are all limited. For example, Mica motes have only 128 kB flash, 4 kB data storage and 20 kbps data rate [25]. Furthermore, sensor nodes are supplied by limited batteries, and thus limited power sources, which implies low power consumption for the various wireless sensor operations as shown in table 2.1 in which the energy consumption of the main operations of Fleck3 and XBee^{TM} wireless sensors is made [26]. In many cases, these batteries can not be replaceable or rechargeable according to the application goals, such as the deployment of sensor nodes in animal bodies, contaminated regions or battlefields which represent unreachable areas. Therefore the lifetime of the network will mainly rely on energy efficiency in communication between the sensor nodes. The energy consumed by radio transceivers is mostly happened from idle listening, transmitting and receiving, depending on the useful throughput. The reduction of energy consumption represents the main challenge that faces WSN and

TABLE 2.1: Energy consumption of the main operations for two examples of wireless sensor nodes

Parameters	**Fleck3 sensor**	**XBee^{TM} sensor**
Sleep mode (μA)	80	10
Transmit mode (μA)	36.8	45
Receive mode (μA)	18.4	50
Duty cycle (μA)	0.27	0.51
Operating voltage(V)	3.3	2.8
Batteries required	3	2
Battery life (Days)	440	230

plays an important role in designing energy-efficient communication protocols. The figure 2.4 represents an overview of energy-efficient schemes for uni-channel WSNs made by the authors of [27].

As depicted in figure 2.4, the energy-efficient schemes can be classified into three big categories: Duty cycling, Data-driven and Mobility-based categories. The duty cycling category focuses on the fact that there is no need for all nodes to have their radio on all the time. Nodes will turn their radio off when there is no network activity and wake up depending on the chosen schema. The Data-driven category tackles the issue of when to take a measurement and how to send fewer data to the network. The Mobility-based category tries to handle the power consumption when the node is in movement following predictable or totally random patterns. Hence, the shared aim of the different schemes is to achieve at least one of the following aspects:

- ☞ The effective reduction of the wake-up time of the nodes to reduce idle listening and overhearing.
- ☞ The avoidance of both types of collisions: direct and indirect (Hidden problem).
- ☞ The avoidance of redundant packets (Deafness problem)
- ☞ The reduction of control packet overhead.

Furthermore, the sensor nodes are designed so the failure of a single node will not affect the task of the network since the WSN is characterized by its self-organizing manner, however, the death of some specific nodes that are placed in critical relaying places may burden the communication in the network by paralyzing of the depending parts to these dead relaying nodes. This can happen specifically in WSNs characterized by their static topology. To deal with this problem, energy balancing techniques have been widely used. Their aim is to balance the energy consumption between the WSN nodes in an equal manner as much as possible.

2.7.3 System Architecture

The protocol stack used by the sink and all sensor nodes is given in figure 2.5 [24]. By following the layers of the WSN protocol stack depicted, we can see that the number of layers in a WSN has been reduced to five compared to the seven layers of the Open Systems Interconnection (OSI) model by leaving aside Session and Presentation layers. In this figure, we notice that WSNs have five basic architectural layers: physical layer, data link layer, network layer, transport layer and application layer. The physical layer provides signal transmission and reception, modulation and coding. Medium Access Control (MAC) protocols in the data link layer manage channel sharing between neighboring nodes. The network layer forwards data through chosen directions (known as routes) to the sink nodes. The transport layer is in charge of connecting WSNs to other networks when required for specific applications. The application layer determines the relevant data sensed according to the application tasks and communicates with the lower layers for these interests. In addition, the power, mobility, and task management planes monitor the power, movement, and task distribution among the sensor nodes. These planes help the sensor nodes to coordinate the sensing task and lower the overall power consumption. The mobility management plane detects and registers the movement of sensor nodes, so the sensor nodes can keep track of who their neighbor nodes are at all times.

The main benefit of using this system architecture in WSNs is that every node involves simply in less-distance, low-power transmissions to the sink through the neighboring nodes

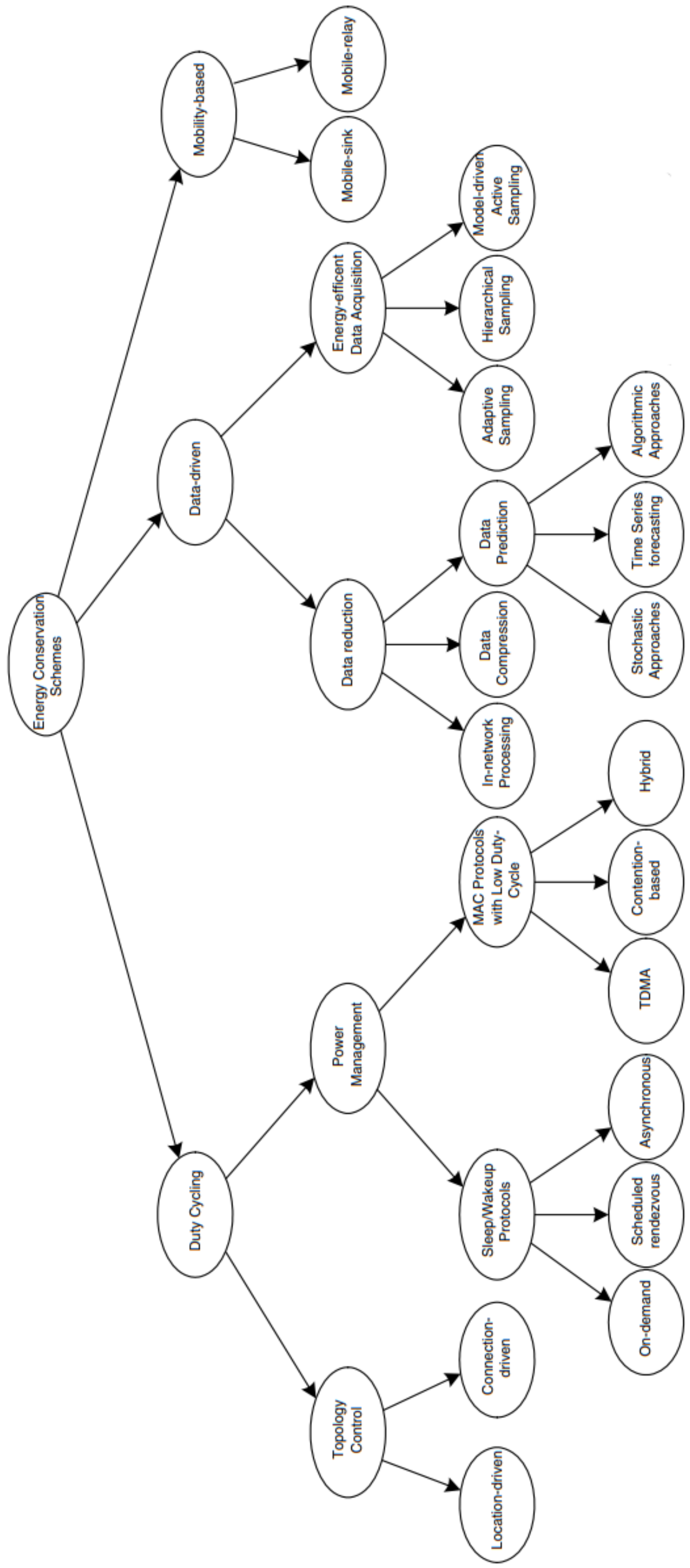


FIGURE 2.4: Overview of energy-efficient schemes for uni-channel WSNs.

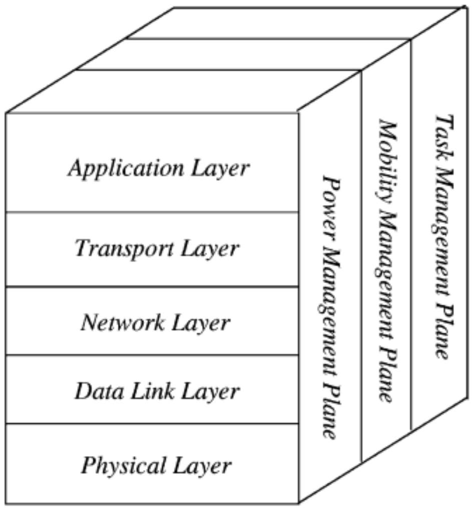


FIGURE 2.5: Comparison between OSI and WSNs layer stacks.

which makes this kind of network scalable as well as includes a high fault tolerance. Depending on this architecture, several working groups such as IEEE and IETF, have standardized several protocols in order to specify standards for Low-power wireless networks that target low data rate, low power consumption and low-cost wireless networking such as Low Rate Wireless Personal Area Networks (LR-WPANs) as specified by IEEE802.15.4 [[16],[28]] or Low-power 802.11(Low-power WIFI) [29]. A collection of the standards for Low-Power wireless networks is made by the authors of [30] as shown in the figure2.6. The main lay-

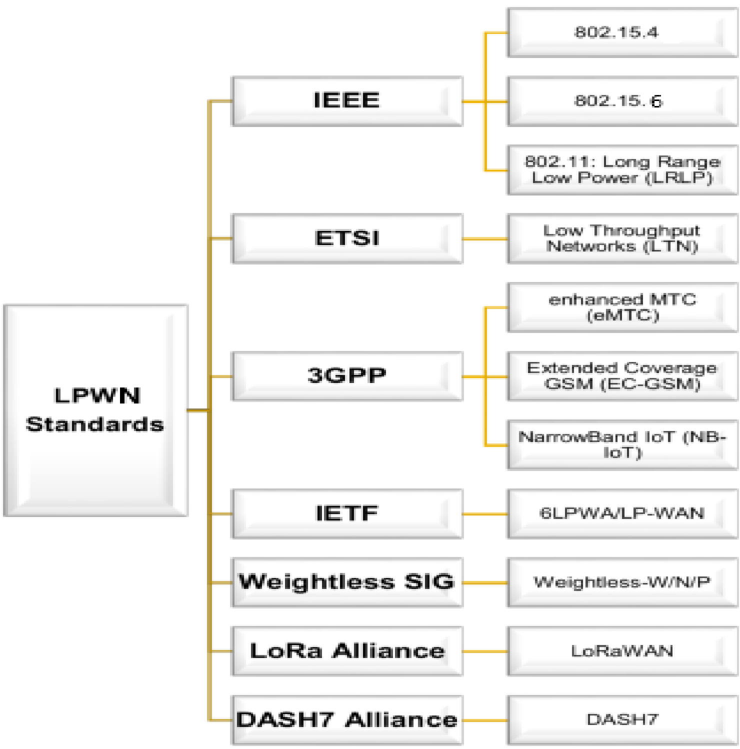


FIGURE 2.6: Low-Power wireless networks standards.

ers that these standards are concerned with are physical and data link layers due to their importance in the design of low-powered wireless network protocols.

The research work in this thesis focuses on WSN MAC protocols at the data link layer which depends on multichannel communication which is highly dependent on the offer of multiple channels in the physical layer. Furthermore, these two layers play an important role in designing low-power WSNs protocols as explained before. Hence, special consideration is given to these two lowest layers in order to more clearly highlight the multi-channel communication.

2.7.3.1 Physical layer

In WSN, the physical layer provides an edge for transferring data as a stream of bits presented by an electromagnetic signal through the radio chip over a specific channel (s). This layer is responsible for the selection of frequency, generation of a carrier frequency, signal detection, modulation and data encryption. WSNs have been widely used in recent fields specifically in the Internet of Things (IoT) field in such a way that they represent the interface between the IoT and the real world. Hence, several protocol standards have been proposed by many working groups as shown in the figure 2.6 previously. IEEE-based standards have gained dominant interest by various researchers due to their great use in the real world. They are divided into three main standards: Low-Power IEEE 801.11, IEEE 802.15.4 and IEEE 802.15.6 standards. They have been proposed to deal with low-powered devices which have increasingly been demanded by the IoT in real life because the IEEE technologies are commonly used in large areas of companies and homes. Consequently, IEEE 802.11af [31], IEEE 802.11ah [32] and IEEE 802.11ba [33] are created respectively to be as alternatives standards in managing low powered devices. Note that the IEEE 802.11ah and .11ba are known also as Ultra Low-Power IEEE802.11 standards. The table 2.2 shows a comparison based on the offer of multiple channels among the latest three standards.

TABLE 2.2: Comparison between Low-Power WIFI protocols

Parameters	**802.11af**	**802.11ah**	**802.11ba**
Date	2014	2017	2019
Range	Up to 3 km	Up to 1 km	Up to 1 km
Frequency band	54MHz to 590MHz	900Mhz to 1GHz	900Mhz to 1GHz
Data Rate	1.8 to 568.9 Mb/s	0.6 to 8.6 Mb/s	up to 230 Mb/s
Channel band	6 to 8 MHz	1 to 8 MHz	4 to 20 MHz

Although the Low-Power IEEE 802.11 standards provide connectivity to a large number of devices by offering a wide range of channels, they are designed to perform upon star topology that depends on the existence of Access Points (APs) which are responsible to manage the channels. The IEEE 802.15.6 designed for Body Area Networks (BAN) is also focused on star topology only with multiple channels over many frequency bands as follow [34]:

- ☞ 402 to 405 MHz (10 channels).
- ☞ 420 to 450 MHz (12 channels).
- ☞ 863 to 870 MHz (14 channels).
- ☞ 902 to 928 MHz (60 channels).
- ☞ 950 to 958 MHz (16 channels)

☞ 2360 to 2400 MHz (39 channels)

☞ 2400 to 2483.5 MHz (79 channels)

On the other side the standard IEEE 802.15.4 can perform upon per to peer, star or tree topologies without specifying especial schemes for channels management which allow it to be used for a variety of different higher-layer standards such as Zigbee [35], Wireless HART [36] and 6LoWPAN [37]. Hence, several protocols are based on this latest standard since is suggested as typical for for low rate wireless networks with low cost, power consumption, density and range of communication to improve the battery life. The orthogonal communication channels offered by IEEE 802.15.4 can be resumed in the table 2.3. Other frequencies

TABLE 2.3: IEEE 802.15.4 channel details.

Frequent band	Channels available	Data rate	Region use allowable
868-868.6 MHz	1	20 Kb/s	Europe
902-928 MHz	10 (2003), 30 (2006 Revision)	30 Kb/s	USA
2.4 GHz	16	250 Kb/s	Global

and bands are being considered. These include: 314-316 MHz, 430-434 MHz, and 779-787 MHz frequency bands in China and the 950-956 MHz band in Japan. Other frequencies are also being considered for UWB variants of IEEE 802.15.4.

2.7.3.2 Data Link layer

The data link layer is responsible for multiplexing data frame detection, data streams, Medium Access Control (MAC) and error control, as well as ensuring the reliability of a point-to-point or point-to-multipoint connection.

The wireless medium is a shared resource and nodes in the same communication range may interfere whenever they try to access the medium at the same time. Thus, the Medium Access Control (MAC) protocol of the link layer is in charge of regulating the access to the medium. It decides which node can use the medium, when and on which channel.

Many WSN MAC protocols have been proposed in the literature as explained in the previous section 2.7.2. These protocols use a set of medium access mechanisms. Therefore, each protocol belongs to one of the following kinds:

☞ Time-based relying on Time Division Multiple Access (TDMA) mechanism.

☞ Frequency-based using Frequency Division Multiple Access (FDMA) approach.

☞ Code-based following the Code-Division Multiple Access (CDMA) mechanism.

☞ Random-access based using the Carrier Sense Multiple Access (CSMA) approach.

☞ Hybrid combining some of the previous medium access schemes.

As an example, The IEEE 802.15.4 standard is widely used as a MAC protocol in WSNs. It allows two types of medium access mechanisms: beacon-enabled and non-beacon-enabled modes. The latter case uses unslotted CSMA with Collision Avoidance (CSMA/CA), whereas the former uses a slotted CSMA/CA algorithm with superframe structure for time schedule, which focuses on combining CSMA/CA and TDMA mechanisms. By integrating the multi-channel communication issue in the IEEE 802.15.4g version proposed in 2012 [38], it uses only beacon-enabled slotted time with superframe structure that bases on TDMA mechanism since it is the best in energy consumption and time exploitation which present the most important interests in WSNs [8].

2.8 Conclusion

In this chapter, we reviewed the main concepts of WNs starting with their definition then classification and properties with the main elements of the wireless communication, and finally, the techniques used in the channel access process which represent the key element that differs the wireless communication from the wired one. In the second part, we reviewed the main concepts of WSNs by giving special consideration to them since they represent the area of work of this thesis.

Chapter 3

Multi-channel Communication in Wireless Sensor Networks

3.1 Introduction

In this chapter, we focus on the multi-channel communication in WSNs and the RL approach. Then, we describe for each RL type the most proposed protocols in the literature and highlight the methodologies they follow. Furthermore, a comparison between these protocols is made and discussed to emphasize important observations that have played important role in the precision of the directions of our work.

3.2 Definition

With the tremendous development in radio and MEMS technologies, sensor nodes capable of tuning their frequency over different channels are becoming more and more progressive. This reality has given birth to the multichannel communication paradigm in wireless networks generally and in WSNs specifically. Therefore, a tremendous attraction has happened towards this paradigm that is considered as an alternative to single-channel communication in terms of increased throughput and delivery ratio which can be performed using parallel communication through multiple used channels. In single-channel MAC protocol, network performance gets deterioration quickly due to higher contention and collision that are sources of sensor exhaustion, and the network becomes more exhibit to faster performance degradation as the number of nodes increases. Consequently, multichannel communication overcomes the single-channel contention/collision problems by exploiting the available multiple channels offered by different standards such as illustrated in the previous section, in parallel communication, which makes the benefit beyond avoiding contention/collision problems to the increase in throughput and delivery ratio because there will not be any limitation for bandwidth and the probability for collisions is less compared to single-channel protocol. To enhance the multi-channel communication in WSNs, the proposed methods must be favored to the nature of the sensor nodes that have reduced resources (processing, storage, energy) and their energy consumption cost should remain small to ensure an improved network lifetime as much as possible. Thus, most envisioned sensor network applications, whether single- or multi-channel, face specific challenges that are explained in the previous section 2.7.2.

3.3 Issues in Multichannel Communications

Although the advantages offered by multichannel communication in WSNs, it raises new problems that can be illustrated as follow [39]:

1. *Multichannel deaf node*: A transmitter wrongly considers a destination node as unreachable because it does not get any response to its requests. This occurs when the destination node is tuned to another channel while the transmitter is trying to communicate with it.

2. *Multichannel hidden node*: In a single channel condition, a hidden node problem may occur in a configuration with at least three nodes, where at least two nodes are out of each other radio range. In a multichannel environment, the hidden node problem occurs when the node misses an RTS/CTS exchanged on one channel while listening on another, causing the hidden terminal problem despite the use of RTS/CTS signaling [40].

3. *Internal interference:* Due to the broadcast nature of the wireless medium, the performance of a multichannel wireless network is drastically limited by interference du to concurrent transmissions on the same or adjacent channels in the same network [41]. Within a WSN, we can distinguish two types of internal interferences:

   - Inter-channel interferences: Where there exists at least one node in the WSN where two transmissions on two different (but partially overlapping) channels interfere with each other.
   - Intra-channel interference: Where there exists at least one node in the WSN where two transmissions on the same channel interfere with each other. Notice that such interference also exist in single channel communications.

4. *External interference:* IEEE 802.15.4 is not the only wireless network operating in the unlicensed band 2.4 GHz. There are WiFi, Bluetooth to name a few. Furthermore, external electromagnetic sources may create perturbations in this frequency band, like for instance electric appliance causing microwave radiation, radar polluting the IEEE 802.15.4. As example shown in the figure 1.2, only four channels of IEEE 802.15.4 do not overlap with WiFi (IEEE 802.11) channels at 2.4 GHz band.

5. *Channel switching:* Whatever the channel assignment method, channel switching is performed by on of the three cases:

   - By the receiver if the channel is assigned to the sender (Sender based policy),
   - By the sender if the channel is assigned to the receiver (Receiver based policy),
   - By both if the channel is assigned to a link and also in case of frequency hopping (Hybrid policy).

Since channel switching is not instantaneous but takes some time (around 200 µs for CC2420 radio), it can lead to packet losses within multihop flows [42]. If each node has its own channel distinct from its neighbors, the end-to-end transmission of a message from a sensor node to the sink requires as many channel switching as the number of visited nodes. This may result in prohibitive delivery delays.

6. *Stability of links:* Radio links in low-power WSNs are often unpredictable. Indeed, their quality fluctuates over time and space. Therefore, selecting high quality links is primordial for data delivery. It enhances throughput by minimizing packet losses, maximizes the network lifetime by limiting retransmissions and avoids path re-selection triggered by links failure. While the problem of links stability exists in single channel WSNs, it is more challenging in multichannel WSNs.

7. *Multicast/broadcast support:* Many multi-channel assignment protocols focus on unicast transmissions enabling a sender and a receiver to tune to the same channel. However, in wireless ad hoc networks and in WSNs, broadcast communications are used to advertise a service, a gateway or more generally some information whose value has a regional scope (in between the immediate neighborhood that is restricted to one-hop neighbors and the whole network). The open question is how to support broadcast in multichannel communication?

8. *QoS support:* In a WSN, all messages have neither the same importance degree nor the same requirements from the application point of view. For instance, an alarm must be delivered to the sink in the shortest delays, whereas the application tolerates the loss of some sampled data as long as the number of successive losses is below a threshold. Service differentiation is required. Furthermore, the network should avoid congestion and support bursty traffic if needed.

9. *Auto-adaptivity:* The multichannel protocol should be environment aware. Indeed, it should be able to take into account the radio environment in which the WSN operates. Channels that encounter perturbations such as jamming, external interference or noise caused by external sources or other coexisting wireless networks (e.g. a coexisting WiFi network) must be avoided. Usually, such channels are blacklisted for a given period. Their status should be periodically reevaluated to use them again if the perturbations have disappeared. In addition, the multichannel protocol should be able to adapt to the application requirements in order to provide the service required by the application and not more. This would result in network resources optimization.

3.4 Multichannel Protocols for WSNs

Several multi-channel protocols have been proposed for WSNs to accommodate parallel transmissions using multiple channels. This interest has become increasingly widespread since the researchers have focused on how to schedule and improve the reliability performance in WSN multi-channel communication. Many classifications exist in the literature based on different overlapping criteria, such as topology criteria which results in centralized, decentralized and clustered multi-channel protocol classes, channels assignment methods resulting in static, dynamic and hybrid groups or using the intelligence factor conducted to intelligent and non-intelligent categories. Although the most used is the one that focused on channel assignment methods, we focus on the one that is based on the intelligence factor to bring out the intelligent reinforcement learning approach since it represents our state of the art.

Hence, many non-intelligent protocols have been proposed in the literature in such a way that they represent the major part. They address the multichannel communication in the different types of the WSN: centralized, distributed and clustered WSNs. As examples, the protocols proposed in [43, 44, 45] which are non-intelligent protocols for centralized, clustered and decentralized WSNs. In [43], the neighboring nodes are splitting into different groups and the time slots in the frame period are divided between these groups by the sink. The receiving nodes then, select the channel-slot pairs in the group sub-period that are not chosen in tow-hops neighboring nodes. In [44], the network is divided into a set of disjoint sub-networks according to the number of used channels. Each sub-network uses its proper channel to communicate with the sink through child-parent links. In [45], a decentralized loopy belief propagation (BP) approach using a factor graph is proposed. The algorithm imprints space, time, frequency and radio hardware constraints into a loopy factor graph and performs iterative message-passing loopy belief propagation with randomized initial priors. For more information about non-intelligent multi-channel WSN protocols, the reader can refer to[46, 42].

On the other hand, scheduling of both channels and links problem has been proved as an NP-hard problem [11, 12], which gave birth to the proposition of intelligent protocols. The main objective of such protocols is maximizing/minimizing some utility/cost functions while assigning different channels. However, since centralized WSNs are scale limited, the multi-channel communication can be well managed by the sink using only deterministic methods such as graph coloring one, whereas the problem becomes more difficult in distributed multi-channel WSNs. Hence, intelligent protocols have been proposed in the last years to overcome this challenge and accommodate parallel transmissions using multiple channels for the required increase in throughput and data delivery rate by ensuring communication reliability and better use of network resources. These protocols can be categorized into three broad classes: heuristic-based, optimization-based and game-based approaches [47]. The heuristic-based algorithms are used to provide approximate and sub-optimal solutions without any performance guarantees in cases the solution of the formulated optimization problem is quite complex or intractable, such as particle swarm optimization (PSO) [48]. The optimization-based algorithms are generally used for the non-convex problems to find either the global optimum solution or sub-optimal solution, such as the proposed algorithm termed Multi-Channel-based Data Gathering with Minimum Latency (MCDGML) in [11]. The algorithms of the two previous approaches are extremely computationally-extensive and typically executed in a central unit with full and real-time information about the network state. On the other side, game-based algorithms are used to deal with the distributed wireless networks with partial information about the state of the network. Such algorithms have shown efficient results, and they are widely used as tools to model complex wireless optimization problems. They can be classified into two subcategories: Game Theory-based (GT-based) and Reinforcement Learning-based (RL-based) approaches. In GT-based algorithms, the channel allocation problem is formulated as a cooperative or non-cooperative game/optimization problem between the network nodes (called players) to achieve the best channels allocation solution, known as the Nash Equilibrium (NE), according to the network goal. Examples of non-cooperative and cooperative games are in[49] and [50] respectively. The RL-based algorithms focus on enabling the network nodes (called agents) to learn in an interactive environment by trial and error using feedback from their own actions and experiences in single, cooperative or deep manners. The Deep Reinforcement Learning (DRL) approach bases on a combination of the deep learning and reinforcement learning approaches. Hence, different DL architectures are used, such as Deep Neural Networks (DNN) [51] and Convolution Neural Networks (CNN) [52], to model a high level of channels management in large spaces characterized by its broad dynamism. As result, the game theory-based and DRL algorithms are computationally expensive and incur considerable communication overhead which leads to extra energy consumption which is not favorable to the nature of the WSNs. Furthermore, the computational and overhead complexity of these approaches proportionally increases with the increase of the number of nodes in the network, making them delayed in time and unscalable. A detailed study of the RL-based approach is presented in section 3.6.

3.5 The Optimization Problem

The authors of [11] have proved that scheduling of both channels and links is an NP-hard problem in a simple manner as follow:

As it has been proved that the Minimum Information Gathering Problem (MIGTP) problem in [53]] is NP-hard, a reduction process is used to prove the NP-hardness of the Multi-channel MIGTP (MMIGTP)problem. Note that the MIGTP is defined as a schedule that can find a path to the sink node from each node in the network and assigns a time slot to each link in the path. To be a proper schedule, no two links can occupy the same time slot if they are within a conflicting range. Two edges are within a conflicting range if there is an intersection between them. Hence, the MIGTP problem is reduced to an MMIGTP problem from the aspect that a single channel is adopted in the MIGTP problem, which is a special case of MMIGTP that multiple channels are allocated to links, especially, both are similar when the number of channels C satisfies 3.1.

$$|C| = 1 \tag{3.1}$$

Hence, the MIGTP problem is a special case of the MMIGTP problem. As the special-case problem is already NP-hard, the general case presented by the MMIGTP problem is NP-hard that cannot be solved in a polynomial time.

3.6 Reinforcement Learning Approach

3.6.1 Definition

The idea of Reinforcement Learning (RL) first appeared in 1979 when machine learning and neuron-like elements for artificial adaptive intelligence were developed. The potential of RL was rapidly explored in psychology, control theory, neuroscience, optimal control and dynamic programming [54]. RL has been developed from an understanding of human interactions with the environment, and how decisions will be made in response to the environment. It records historical experience generated by rewards and punishments of the learning behaviors and determines future actions through trial-and-error interactions with a dynamic environment [55]. Hence, RL is a biologically inspired machine learning technique, in which an agent acquires its knowledge through trial-and-error interactions, called strategies, with its environment including the other agents. It is a very successful Artificial Intelligence approach that has been employed for increasingly difficult tasks.

RL usually performs by quantizing the environment to a set of different states, and agents learn their behaviors through actions and rewards in each state. Each state depends on a subset of actions, and the union of these subsets represents the set of actions. The agents perform actions in each state then receive rewards after each action. The received rewards usually depend on the goal of the RL process and affect only the action and state pair for which they belong. Each action and state pair has an associated weight presented usually by a probability measurement which is updated by received rewards to indicate which actions will be preferred by the agents in the selection process of each state. The reward can be positive or negative which affects the weight of the corresponding state-action to push the RL process to the predefined goal. The basic framework for RL is shown in figure 3.1.

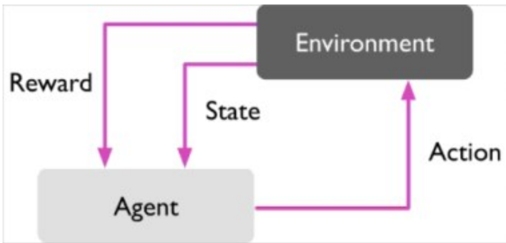


FIGURE 3.1: Reinforcement Learning operation.

Figure 3.2 shows an example given in [55] which presents how RL can be used in a simple square maze. An agent stands at the point A and aims to reach the destination presented by the point B. The squares are considered as the set of states, and the set of actions is considered by the four possible directions of the agent moving to the four neighbor squares (up, down, left and right). Each action receives a negative reward if it exceeds the maze limits, otherwise, it receives a positive reward. If the agent reaches the destination B, it receives a big reward (+100 for example). In each state the agent selects the action with the highest weight, if several actions have the same highest weight, it selects a random one. After many trials, the RL process produces an optimum route (one of the two solid routes as examples), but the dashed route will not be selected by the RL process (although it is an optimum route), because it is close to the limit and the actions selection process have a higher probability of receiving negative rewards.

The trade-off between exploration and exploitation is an important challenge in RL. To achieve higher weights and rewards, agents usually select actions based on their experience history, so the actions which receive good rewards will always be preferred. However, it is possible that the preferred action is a sub-optimal solution. Hence, the agent needs to explore other actions to emphasize the existence of an optimum solution rather than a sub-optimal one. However, an optimum solution may not be found by exploration despite the incurred costs. Consequently, addressing a balance between exploration and exploitation has a significant impact on the efficiency of RL. To do that, the ε-greedy algorithm [56] is proposed to define the trade-off between the exploration and the exploitation in such a way that the RL performs at each selection process a preferred exploration based on random selection for a probability ($p = \varepsilon$), while it performs exploitation for a probability ($p = 1 - \varepsilon$). Generally, the value of ε is taken 0.1.

Figure 3.3 shows an example of exploration and exploitation for a similar environment to the previous example except for the gray area. Once the agent moves into the gray area,

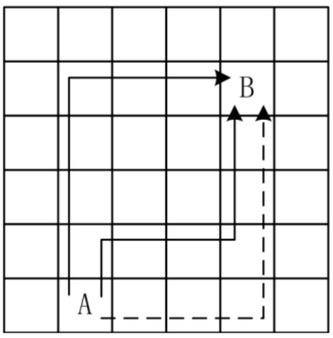


FIGURE 3.2: Reinforcement Learning example.

a negative reward will be returned, the same as exceeding the limit of the maze. For the RL algorithm without exploration, the solid route will be selected, but it is a sub-optimal solution with 9 steps to the destination. The dashed route will not be selected because the actions in these states have a large probability of receiving a negative reward. Through exploration, the optimum dashed route can be found, which is 7 steps to the destination.

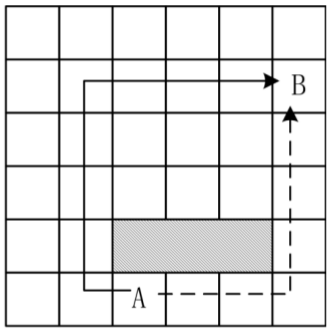


FIGURE 3.3: Exploration and Exploitation example.

3.6.2 Reinforcement Learning for Multi-channel WSNs

RL has already been used to propose solutions to several problems in WSNs, such as neighborhood management [57], task scheduling [58], Duty cycling [59], etc. However, these solutions are limited to single-channel operation and a fixed routing topology. It has been also used to deal with the channel assignment problem. Therefore, the RL solutions for channel assignment in WSNs can be divided into three categories: single-agent, cooperative multi-agent RL and deep RL.

In addition to the protocols that are proposed to WSNs in general, some other protocols are proposed specifically for Cognitive Radio Sensor Networks (CRSNs) that represent specified WSN characterized by more power and capabilities for its sensors in such a way that the CR sensors can sense a large number of band and use more than one wireless network. With these capabilities, CR sensors can operate in licensed bands as well as in unlicensed bands. The unlicensed bands presented by the Industrial, Scientific and Medical (ISM) bands such as the 2.4GHz band, are used by the WSNs in general. However, the licensed bands are reserved for proprietary communications and represent the battlefield of the CRSNs. Therefore, the main focus of the CRSN proposed protocols is how to enable the CR sensors considered as Secondary Users (SUs) to use the licensed channels without disturbing the communication of the licensed users called Primary Users (PUs) that have the priority to access the channel. Figure 3.4 shows a CRSN where PBS denotes the Primary Base station and the CBS denotes the Cognitive Base Station.

3.6.2.1 Single-agent Reinforcement Learning

In single-agent RL, The agent performs an action at each step independently from the other agents and gets feedback which it uses to optimize its behavior in the future as shown in figure 3.5.

Referring to [60], single-agent RL can be formulated as a Markov Decision Process (MDP) named also Q-learning, that can be expressed mathematically as MDP=$\{S, A, T, R_{1...N}, \gamma\}$, where;

- ☞ S is the set of states,

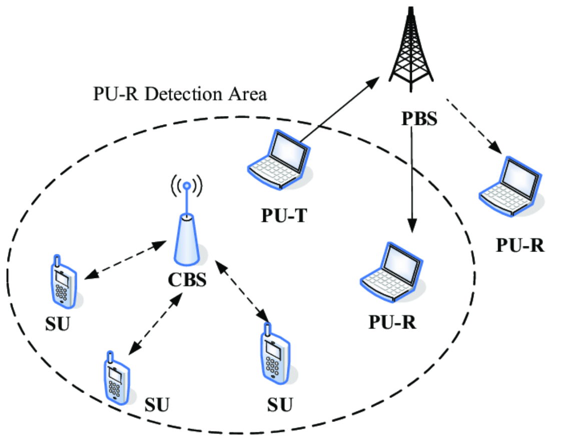


FIGURE 3.4: Cognitive Radio Sensor Network.

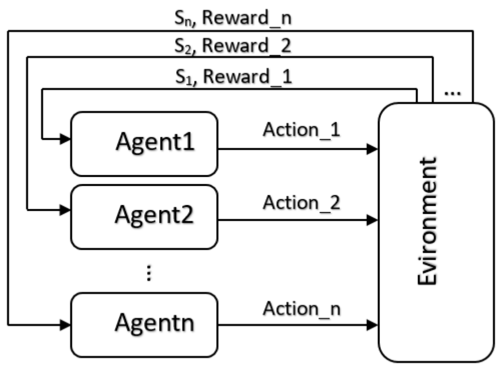


FIGURE 3.5: Single-agent RL.

- A is the set of actions,
- T:$SAS \longrightarrow [0,1]$, is the state transition function,
- R_i:$SAS \longrightarrow \mathbf{R}$ is the reward function of agent i.
- $\gamma \in [0,1]$, is the discount factor, which represents the relative importance of future and present rewards.

Therefore, T is defined by 3.2 and 3.3 as follow:.

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha(R(s_t, a_t) \\ + \gamma V_t(s_{t+1}) - Q_t(s_t, a_t)) \tag{3.2}$$

where, $Q_{t+1}(s_t, a_t)$ means the update of Q value at time t + 1, after executing at time step t, the action a_t in the state s_t that belongs to the state set S. $R(s_t, a_t)$ means the immediate reward after executing the action a_t at time t. γ and α are the discount factor and the learning rate parameter respectively which can be set to a value in [0,1]. The higher value of γ pushes the agent to rely on the future reward than the immediate one. Also, when α is set close to 0 the agent gives more priority to the previously learned value. V_t and V_{t+1} are the values of function 3.3 (or 3.4) at times t and $t+1$ respectively. As explained before, based on the ε-greedy algorithm, $V_{t+1}(s_t)$ is calculated using 3.3. However, for certain cases, the exploration doesn't matter such as the case in some channel assignment methods, where the focus is the successful channel affection without the need to explore the other channels to obtain the best one. Consequently, a Greedy algorithm, that can be considered as the ε-greedy with $\varepsilon = 0$, is applied rather than the ε-greedy one and $V_{t+1}(s_t)$ is calculated using 3.4.

$$V_{t+1}(s_t) = \begin{cases} Max_{a \in A} Q_{t+1}(s_t, a) & \text{for } \varepsilon = 0.9 \\ Random & \text{for } \varepsilon = 0.1 \end{cases} \tag{3.3}$$

$Max_{a \in A} Q_{t+1}(s_t, a)$ means the maximum Q value after performing an action from the action set A. ε is the parameter that defines the exploration/exploitation trade-off.

$$V_{t+1}(s_t) = Max_{a \in A} Q_{t+1}(s_t, a) \tag{3.4}$$

For each state, the Q-table of every single agent is constructed with one dimension according to the set of actions as shown in table **??**.

TABLE 3.1: Q-values vector for single RL agent

	Actions			
	a_1	a_2	**...**	a_j
Q-values	q_1	q_2	...	q_j

The algorithm of single-agent RL is presented in Algorithm 2.1.

Algorithm 2.1: Single-agent RL

1: Initialization: $Q(s,a) = \delta, \forall s \in S, \forall a \in A, \alpha, \gamma / \delta, \alpha, \gamma \in [0,1]$

2: For t=1,2,3,... **do**

3: Select action in this state s_t based on the exploitation/exploration strategy using the Eq3.3

4:Perform the selected action and obtain reward r

5:Calculate the Q_{t+1} value of the performed action using the Eq3.2

6:Update the Q value of the performed action

7: End for

A Q-Learning based bidding algorithm for spectrum auction has been proposed by the authors of [61] for the secondary users (SUs) in Cognitive Radio (CR) sensor networks. The proposal aims to maximize profits in dynamic channel access opportunity based on the histories of visited states and the utility of selected bidding actions performed via sending at different time periods for spectrum auction, which result in owning bidding credits for each time slot. Then, the secondary users perform the proposed algorithm based on bidding credits to learn from their competitors and select better bids for available frequency bands. The results show that the proposed algorithm performs efficiently.

In [62], the schedule-based multi-channel communication protocol that performs upon Reinforcement Learning Multi-channel MAC (RL MMAC) algorithm is proposed. It represents an extension of the work proposed in [63]. In RL-MMAC, a decentralized tree-based topology is used, then the nodes learn to perform their transmission schedule on their parent's channels in a distributed manner to avoid supplement overhead and to obtain a traffic-adaptive schedule using "*win stay, lost shift*" strategy. After the learning process, the best action in each slot is chosen, based on the success probabilities of the available set of actions. The new nodes perform the algorithm on the sleeping slots of their neighboring nodes to learn the best action to do in these specific slots. The results show the best performance of the proposed algorithm in simulated and real environments.

The authors of [64] have proposed a Normal Equation based Channel quality prediction (NEC) algorithm that starts by performing Channel Rank Measurement (CRM) based on the Received Signal Strength Indicator (RSSI) and the average of the link quality indicator for each channel. The set of accepted channels is then used for training a learning operation between the neighboring nodes. this approach is applied through two algorithms, The Normal Equation based Weighted Moving Average Channel quality prediction (NEWMAC) algorithm and Normal Equation based Aggregate Maturity Criteria with Beta Tracking based Channel weight prediction (NEAMCBTC) algorithm, that can perform channel quality estimation based on both current and past values of channel rank estimation. The results show that the two algorithms succeed in minimizing interference and energy consumption, while the algorithm NEAMCBTC outperforms the compared algorithms in terms of channel quality and stability assessment.

In [65], the Channel-aware Reinforcement learning-based Multi-path Adaptive (CARMA) routing algorithm is proposed for forwarding data with adapting to varying channel conditions in an underwater environment in which the wireless radio signals are strongly attenuated. The proposed solution enables nodes to learn optimal routes based on the number of transmissions and the channel quality by considering the energy consumption in the utility function of the selection process to minimize energy consumption and perform penalties against the dropped packets. The results show that the delivery ratio is enhanced by 40% with fast and energy-saving performance compared to other protocols.

The spectrum management in Industrial IoT (IIoT) is studied in [66], where two RL algorithms are proposed for dynamic spectrum hand-off access based on IEEE802.15.4 protocol. The first algorithm focuses on the learning process to predict the channels occupancy time by the gateways based on the historical use of the channels, whereas the second is used to enable the gateways to learn the quality of the channels based on the mean Received Signal Strength Indication (RSSI) model in order to rank them in candidate channel list as shown in figure 3.6, to enable efficient channels management for the IIoT devices.

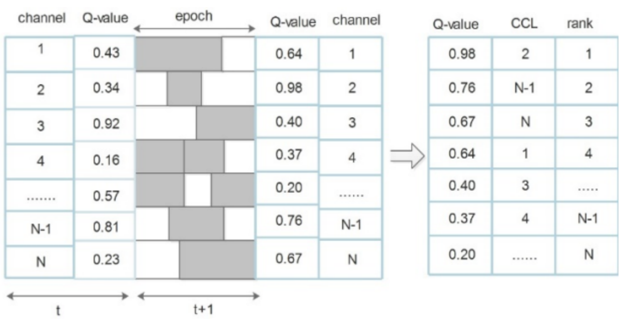


FIGURE 3.6: Learning process to rank accepted channels.

In [67], a trade-off between an online non-cooperative game and offline schedule-based single-agent learning methods for channel assignment is proposed. The first method is used if the network can satisfy the energy requirements especially at the beginning of the network performance, however, the second one is used dynamically through a decentralized tree-based topology to ensure less energy consumption by the fact that the nodes communicate only with their parent nodes rather then all the neighboring nodes.

Some works have used the Multi-Armed Bandit (MAB) learning that is considered as a sub-field of reinforcement learning where the learning process focuses on action space and received rewards for only one state (one-state Markov decision process) for multiple targets (arms) by sampling one target at a time, with the objective is to maximize the sum of the collected rewards to choose the most optimal target by the bandit. Thus, the MAB learning algorithm represents the simplest form of the Q-learning algorithm by the fact that it uses only a set of actions to perform and focuses on the mean obtained rewards as a utility function which represents a Q-learning algorithm with $\alpha = \frac{1}{t}$ and $\gamma = 0$, as shown by the following derivation:

$$Q_{t+1}(a_t) = \frac{1}{t}\sum_{i=1}^{t} R_i(a_t) \tag{3.5}$$
$$= \frac{1}{t}(R_t(a_t) + \sum_{i=1}^{t-1} R_i(a_t)) \tag{3.6}$$
$$= \frac{1}{t}(R_t(a_t) + (t-1)Q_t(a_t)) \tag{3.7}$$
$$= (1-\frac{1}{t})Q_t(a_t) + \frac{1}{t}R_t(a_t) \tag{3.8}$$

Where $R_i(a_t)$ is the reward of the action a_t at time i.

Hence, In [68], a Multi-Armed Bandit (MAB) algorithm for access to multiple channels with the unknown environment information for cognitive users is proposed. Therefore, the channels are affected into several groups according to the water-filling principle based on the learning algorithm UCB-K index which bases on the fact that the channel with the largest index value will be included in the channel group with the smallest sum of index values. After that, the algorithm allocates fairly channel group for each cognitive user by assigning a rank of priority access for each one to be used in choosing the corresponding channel.

In [69], two stages of learning are used in CR-based WSNs to assign channels for SUs and predict the time in which the channel will remain free. The first learning focuses on picking the idle channels against the PUs traffic, through a MAB formulation, while the second one is used to select the appropriate strategy for sensing these channels by using a Bayesian method to predict the free interval duration of the sensed channel. The results show that the SUs spend lesser energy in sensing, while they achieve higher throughput.

In [70], TOW-Dynamic for Channel Assignment (TOWD CA) algorithm is proposed, wherein the channel assignment is modeled as a multi-armed bandit problem and a machine-learning-based channel assignment algorithm utilizing Tug-Of-War (TOW) dynamics method is proposed. Based on the histories of visited states for each channel, the nodes select the best channel then they perform the non-beacon-enabled mode of IEEE802.15.4 protocol for access to the medium.

A proactive RL for multi-channel communication in cognitive IoT with multiple interfaces is proposed in [71]. The IoT devices start by performing a MAB algorithm on different available channels to calculate their idle probabilities and predict the time for which the channels remain idle. Then, the channels are organized in descending order according to their estimated idle probabilities to permit the IoT devices to select optimal multiple channels and use them simultaneously during the predicted idle times.

The security issue is tackled in [72] by using an online MAB algorithm to enhance the channel assignment and energy consumption of the sensors in cyber-physical systems. Hence, each sensor learns to jointly select the best channel by avoiding the ones infected by DoS jamming attacks on one hand, and select the best transmission power by balancing between the packet delivery ratio and the consumed energy on the other hand.

3.6.2.2 Cooperative Multi-agent Reinforcement Learning

The cooperative multi-agent RL approach is used to build a cooperative machine learning system between more than one agent as shown in figure 3.7, which allows multiple agents to learn together utilizing one another's strength for decreasing individual learn weaknesses and enabling learning to be accelerated. Nevertheless, the use of this approach opens up new necessary issues to be tackled that can be resumed in the response of the three following questions: Why, How and when the multiple agents can learn together? To answer these

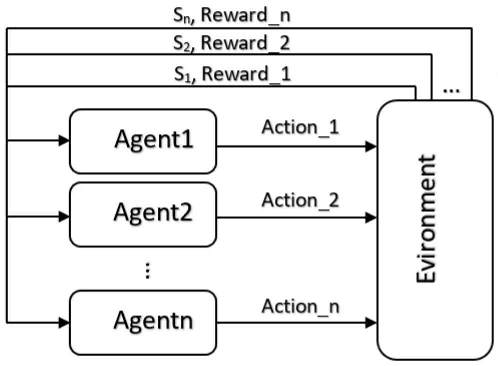


FIGURE 3.7: Cooperative multi-agent RL.

questions, several solutions have been proposed in the Artificial Intelligence multi-agent RL sub-field for Transfer Learning (TL) depending on the cases for which these solutions are proposed. Hence, the cooperation between the RL agents can be in many kinds such as coordination, competition and advising [60]. Nevertheless, these solutions can not be applied directly to the WSNs without considering their limited resource constraints. For this reason, some other solutions have been proposed to deal with the TL between the RL agents in a decentralized WSN focusing on energy efficiency factor such as in [73] and [74] for task scheduling. However, these solutions are based on a single channel.

Referring to [60], the cooperative multi-agent RL can be formulated as a Partial Observation Markov Decision Process (POMDP) that can be expressed mathematically as POMDP=$\{n, S, A, T, R_{1...N}, \gamma\}$, where;

- ☞ n is the number of agents,
- ☞ S is the set of states of all agents,
- ☞ A is the joint action space composed of local actions for all the agents,
- ☞ $T{:}SAS \longrightarrow [0,1]$, is the state transition function,
- ☞ $R_i{:}SAS \longrightarrow \mathbf{R}$ is the reward function of agent i.
- ☞ $\gamma \in [0,1)$, is the discount factor, which represents the relative importance of future and present rewards.

Thus, each agent performs the equation 3.2 based on modified versions of the two equations 3.3 and 3.4 by integrating the cooperation aspect in the process of action selection by the fact that each agent must collect the Q-values of other agents before performing 3.9or 3.10according to the presence of the ε-greedy or Greedy method as explained before.

$$V^i_{t+1}(s_t) = \begin{cases} Max_{a\in A} \sum_{i=1}^{p} Q^i_{t+1}(s_t, a) & \text{for } \varepsilon = 0.9 \\ Random & \text{for } \varepsilon = 0.1 \end{cases} \tag{3.9}$$

$$V^i_{t+1}(s_t) = Max_{a\in A} \sum_{i=1}^{p} Q^i_{t+1}(s_t, a) \tag{3.10}$$

Were $V^i_{t+1}(s_t)$ is the action selection function for the agent i at time $t+1$ and p represents the set of agents with which the agent i cooperates ($p \subseteq n$). Consequently, the Q-table of each agent for each state is constructed with two dimensions according to the set of actions as shown in table **??**.

TABLE 3.2: Q-values vector for cooperative RL agent

	Actions			
	a_1	a_2	**...**	a_J
Q-values of $agent_1$	q^1_1	q^1_2	...	q^1_J
Q-values of $agent_2$	q^2_1	q^2_2	...	q^2_J
⋮	⋮	⋮	...	⋮
Q-values of $agent_p$	q^p_1	q^p_2	...	q^p_J

The algorithm of cooperative multi-agent RL is presented in Algorithm 2.2.

Algorithm 2.2: Cooperative multi-agent RL

1: Initialization:$Q_i(s,a) = \delta, \forall s \in S, \forall a \in A, \forall i \in p \subseteq n, \alpha, \gamma / \delta, \alpha, \gamma \in [0,1]$

2: For t=1,2,3,... **do**

3: Select action in this state s_t based on the exploitation/exploration strategy using the Eq3.9

4:Perform the selected action and obtain reward r

5:Calculate the Q_{t+1} value of the performed action using the Eq3.2

6:Send the pair (Action, Q value) to the cooperated agents

7:Receive the pairs (Action, Q value) from the cooperated agents

8:Update the Q values of the own performed action and the cooperated agents actions

9: End for

In [75], the Reinforcement Learning based clustered Cooperative Channel Sensing (RL CCS) algorithm for CR-based WSNs is proposed. In which, the cluster heads are constructed using eligibility probabilities based on the nodes' residual energy, then the SUs cooperate in coordination to learn an optimal policy that determines the optimal sensed channel to be used in communication with the cluster heads. After taking the channel sensing facts as actions and the channels as states, each node learns to maximize its utility based on four metrics: local decision accuracy, channel availability, channel bandwidth and energy cost for sensing the channel using the Softmax strategy by making local decisions then final decisions after receiving the local decisions of the neighboring cluster members about the channel occupancy. After the learning period, each SU determines the optimal sensed channel adequate for the required bandwidth. The results indicate the efficiency of the proposed algorithm in reducing the sensing energy, PUs detection and sensor performance.

In [76], central coordination and distributed execution approach between the SUs in CR sensor networks is proposed to choose the best vacant channel in a dynamic environment with less communication overhead. A belief function is used by each SU with its SU partners which are updated at each time using the soft-max method. According to partner levels of reliability that are considered as weights, the belief of each channel by the SU is taken as the set of states and the sensing of each channel is considered as the set of actions. Each SU has an actor and critic components, the actor is used for online local learning during the execution, while the critic component is used for the central cooperative online training. The results demonstrated that the proposed algorithm enhances the sensing accuracy compared to non-cooperative or centralized algorithms.

An online learning for gigabit WSN using multiple energy-aware MAB algorithm is proposed in [77]. Unlike the previous works that consider the channels as arms, the proposed solution considers the sensors as arms and the energy aspect is integrated into the arm selection process by the bandit (presented by the fusion center) at each round by the fact that the nodes with more throughput and remaining energy have more chance to be optimal compared to nodes with the same throughput and distance from the center node, but with less remaining energies. This selection function is used by two algorithms in two ways. The first is the Randomized UCB (RUCB) algorithm in which a variable random selection is used at each round, then the optimal armed is selected, whereas the second is Perturbated History Exploration (PHE) in which the history of the arm is used. The proposed algorithms showed a more significant performance compared to competing algorithms.

In[78], the Collaborative Multi-agent Ant-jamming Algorithm (CMAA) is proposed for avoiding jamming channels in CR-based WSNs by using a trade-off between coordination and competition strategies in order to avoid the Jammed channels and compete between the nodes in using of the secured channels. Thus, the nodes take as actions the sending and sensing operations, and the joint (action, channel) as states, then they learn in cooperative fashion using a common channel to coordinate the jammed and the allocated channels in each slot according to the ε greedy method by considering the miss detection and false alarm probabilities. Since the channel sensing process is performed after packet sending, the nodes use the “decision-feedback-adjustment" method after receiving the neighboring decisions to compete in exploiting the secured and vacant channels.

The joint coordination and cooperation is also used in [79] by using enforcement cooperation based on consensus between the SUs nodes to avoid PUs allocated channels in CR WSNs. A variant set of actions is used composed of sensing, sending or switching to one of the available channels, whereas the states are considered as a pair of channel and its state that can be idle, busy or unknown. Based on a modified ε Greedy method that performs by allocating different ε values for the SUs to explore different sets of channels, each SU learns by transiting between the different states, as illustrated in figure 3.8, in a cooperative manner with its neighboring nodes by using a distributed stochastic gradient method to perform the consensus aspect. The obtained results demonstrated the efficiency of the used consensus method which is close to the central-based algorithm in performance but with scalability and robustness.

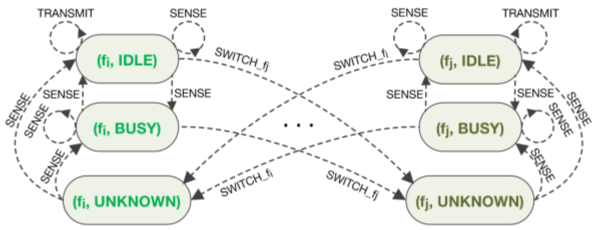


FIGURE 3.8: Learning process presented as MDP.

3.6.2.3 Deep Reinforcement Learning

Deep Learning (DL) approach is a sub-class of machine learning methods attempting to model with a high level of data abstraction through articulated architectures of different non-linear transformations. The concept of deep learning emerged in the 2010s, with the convergence of four factors [47]:

- ☞ Multilayer artificial neural networks.
- ☞ Discriminant and learner analysis algorithms.
- ☞ Machines whose processing power can process massive data.
- ☞ Huge databases capable of training large systems.

Machine learning techniques have played an important role in dealing with complex situations such as resource allocation, queue management, congestion control, etc. On the other hand, as more desires are brought to wireless communication due to advances in many fields such as higher computing capacity and bigger data sets, modern wireless communications are becoming more and more complex. To accommodate these urges, deep learning applications have been proposed which made significant impacts on improving communication performances in many areas, such as communication signal characteristics, routing delay, queuing state, path congestion situation, etc [80].

As mentioned before, different DL architectures exist such as DNN, CNN, and DRL. They have already shown spectacular capabilities in dealing with several fields of application, such as pattern recognition, computer vision, image classification and network resources management.

DL uses cascaded layers of the nonlinear processing unit to extract features from the input data and ultimately make a decision as output data. Each layer has a specific number of neurons and between the input and the output layers, there are multiple hidden layers as shown in figure 3.9. Each neuron takes as input the output of the dependents neurons belonging to the previous layer. Through the different layers, DL moves from low-level parameters to higher-level parameters, where the different levels correspond to different levels of data abstraction. The application of the neural network is performed in two steps: training and verification phases. A set of complex propositional computations based on weights and thresholds must be made to perform a number of transformations on the data between the input and the output layers. Therefore, two main computational operations exists: forward feature abstraction and backward error feedback. The training phase usually needs both elements, while the verification phase implements the first only.

Forward feature abstraction: Used to compute the output of each neuron as shown in figure 3.9. Hence, this output is obtained through two steps. First, node n_{ij}, which is the $j_t h$ neuron n of the $i_t h$ layer, computes a weighted sum of all its inputs. Then the result is sent to an activation function $f()$ to obtain the output z_{ij} of node n_{ij} as shown in 3.11 [52].

$$z_j^i = f(\sum_{k=1}^{L_{i-1}} w_k^i x_k^{i-1} + \xi_{ij}) \tag{3.11}$$

where w_k^i is the weight from node $n_{i-1,k}$ to node n_{ij}, L_{i-1} is the number of nodes for layer $i-1$ and ξ_{ij} is the threshold coefficient of the node n_{ij}. Therefore, the output of l_{th} layer is given by 3.12.

$$z^l = f(W^l.z^{l-1} + \xi^l) \tag{3.12}$$

The activation functions commonly used are: the Rectified Linear Unit (ReLU) $f(z) = max(0,z)$, the hyperbolic tangent function $f(z) = [exp(z) - exp(-z)]/[exp(z) + exp(-z)]$, and the logistic function $f(z) = 1/[1 + exp(-z)]$.

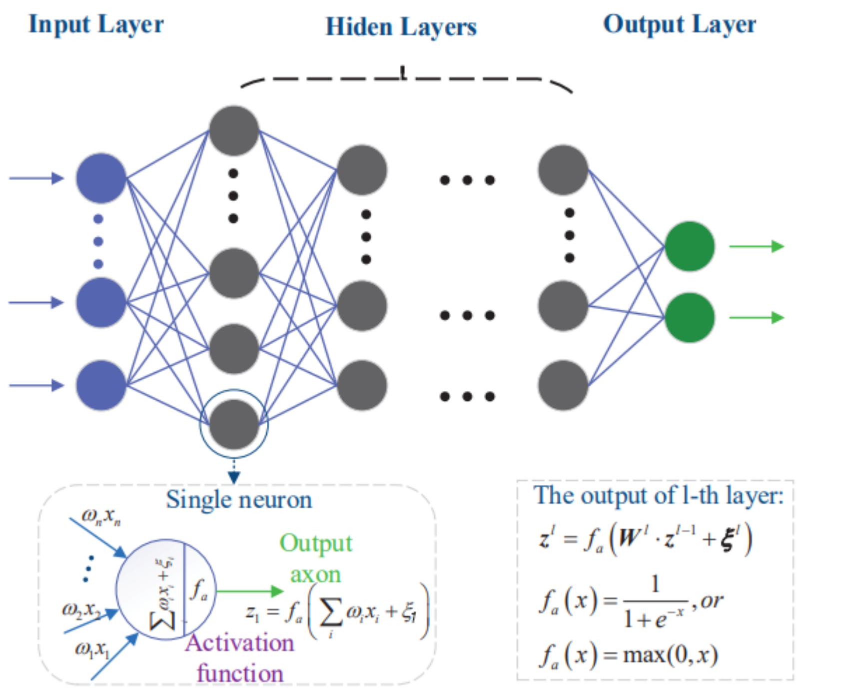


FIGURE 3.9: Deep learning schematic.

Backward error feedback: This operation aims to adjust the different parameters to improve the accuracy of the final output of the learning system. The initial values are random or empirical values, then, the DL trains the parameters W and ξ by minimizing the cost function by the gradient descent method that is usually used. The cost function is defined by 3.13 as follow [81]:

$$W', \xi' = argmin_{W,\xi}(E) \tag{3.13}$$

Where E is the loss function that usually uses a mean squared error function or a cross entropy function defined by3.14 and **??**espectively as follow [Ref]:

$$E_{mean-square} =\| y - t \|^2 \tag{3.14}$$

$$E_{cross-entropy} == y^T.log(t) \tag{3.15}$$

Where y and t are the generated output and the correct output, and $(.)^T$ denote the transpose of the matrix. As shown in 3.10, the classification accuracy is feedbacked, according to which the connection weights are modified. For a node of the deepest layer, say, node n_{Nj}, the error derivative is $y_{Nj} - t_{Nj}$, then the error derivative of lower layer connection is defined by 3.16 as follow:

$$\frac{\partial E}{\partial z_{Nj}} = \frac{\partial E}{\partial y_{Nj}} \frac{\partial y_{Nj}}{\partial z_{Nj}} \tag{3.16}$$

where $\partial E / \partial y_{Nj} = y_{Nj} - t_{Nj}$ and $j = 1, 2, ..., L_N$. For the j-th node of layer i ($i = 1, 2, ..., N-1$), first a weighted sum of the error derivatives of all the inputs (from deeper layer) to the node is computed, denoted as $\partial E / \partial y_{Nj}$. Then the error derivative of the upper layer connection is defined by 3.17 as follow:

$$\frac{\partial E}{\partial z_{ij}} = \frac{\partial E}{\partial y_{ij}} \frac{\partial y_{ij}}{\partial z_{ij}} \tag{3.17}$$

where $\dfrac{\partial E}{\partial y_{ij}} = \sum_{k=1}^{L_i+1} w_{jk}^{i+1} \dfrac{\partial E}{\partial z_{i+1,k}}$, $i = 1, 2, ..., N-1$ and $j = 1, 2, ..., L_i$.

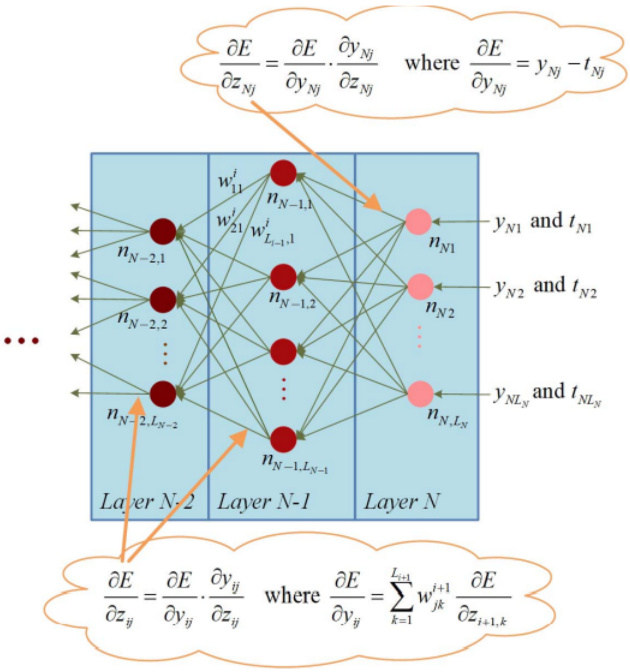


FIGURE 3.10: Deep learning schematic.

In multi-channel WSN communication, the use of reinforcement learning algorithms can suffer from a low learning speed in large states that can not be all observable (partially observable) and/or continuous action spaces, which leads to an instability problem. The generalization and function approximation capabilities of the deep learning approach can overcome this limitation. Hence, a combination between reinforcement learning and deep learning techniques has been introduced as DRL and has demonstrated good performance in wireless multi-channel communication and thus WSNs.

Therefore, two novel strategies are proposed: experienced replay and iterative update. Thus, the agent interacts with the environment in different states through a sequence of actions, in order to maximize the cumulative reward defined by 3.18, where the γ is the discount factor.

$$E_{(s,a,r,s')} = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + ...|s_t = s, a_t = a, \pi \tag{3.18}$$

The reward is represented as $Q_t(s,a,\theta_i)$, where θ_i is the weight of the Q-network at iteration i. To replay the experience, for each moment t, the agent's experience $e_t = (s_t, a_t, r_t, s_{t+1})$ is stored in the experience pool $U(D)$. Q-network updates its weights θ that is fixed between individual updates, every c constant steps to minimize the loss function defined by 3.19.

$$L_i(\theta_i) = E_{(s,a,r,s')\in U(D)}((y_i - Q(s,a,\theta_i))^2) \tag{3.19}$$

where y_i is derived from the same Q-network with old weights θ_{i-1} and new state s' after taking action a from state s. It is defined by 3.20 as follow:

$$y_i = E((r + \gamma max_{a\in A} Q(s_{t+1}, a, \theta_{i-1})) \tag{3.20}$$

The parameterization of reward Q for each action is achieved by a neural network yielding the predicted Q value for a specific pair (state,action), where each possible pair is provided a separate output unit. The DRL algorithm is shown in Algorithm 2.3.

Algorithm 2.3: DRL

1: Initialization: A,S,$Q_i(s,a) = \delta$, W,C

2: For t=1,2,3,... **do**

3 : if t $\leqslant$ W **then**

4: Select an action a_t randomly

5: else

6: imput $\phi^t = (s^{t-W}, s^{t-(W+1)}, ..., s^t)$ to the DRN with weight θ and obtain output

7: choose a_t via ε-greedy

8: endIf

4:Perform the selected action and obtain reward r

5: Store experience (s_t, a_t, r_t, s_{t+1})

3 : if C **then**

6: Randomly choose a tuple (s_t, a_t, r_t, s_{t+1})

7:Estimate reward y of (s_t, a_t, r_t, s_{t+1}) using 3.20

8:Compute gradient with respect to network weight using 3.19

9:Update θ

8: endIf

9: End for

Different multi-channel DRL-based protocols have been proposed for WSNs in the last years. Since DRL necessitates complex calculation, most of them train the DRL in a centralized manner and few works devise the load of calculation between the network devices.

In [82], the problem of dynamic spectrum sensing and aggregation is investigated in a CRSN containing N correlated channels. At each time slot, a single cognitive user with a certain bandwidth requirement either stays idle or selects a segment comprising C (C < N) continuous channels to sense. Then, the vacant channels in the selected segment will be aggregated for satisfying the user requirement. The user receives a binary feedback signal indicating whether the transmission is successful or not. To find a policy that can maximize the number of successful transmissions without interrupting the primary users (PUs), DRL is used to address the challenge of unknown system dynamics and computational expenses. The results show that DRL can achieve near-optimal performance among different system scenarios.

The authors of [83] are proposed a DNN algorithm for multimedia IoT tasks divided between the cloud server and the edges servers in such a way that the DNN trains first at the cloud server then the learning network is divided into two parts, the first part which includes the lower layers near the input is deployed into edge servers and the second part which includes the higher layers is deployed into the cloud server for offloading processing. Thus, The edge servers collect multimedia data from the IoT devices to load the intermediate data from the lower layers and then transfer the resulting data to the cloud server as input data for the higher layers as shown in figure 3.11. In order to schedule both data and bandwidth

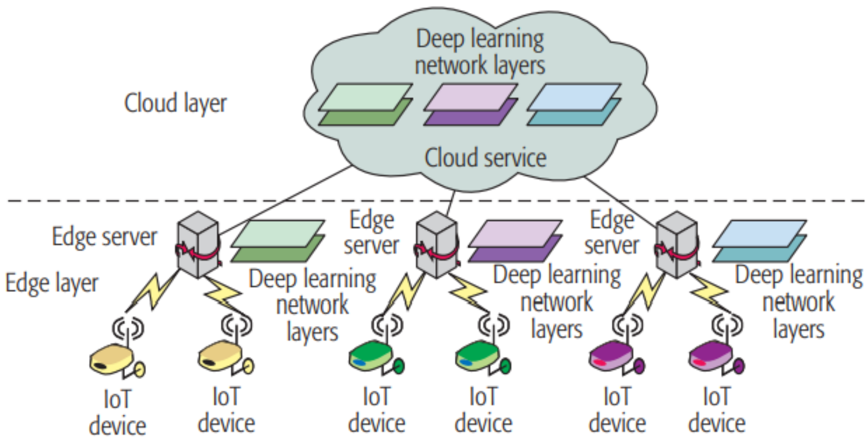


FIGURE 3.11: System architecture in [Ref].

in transferring processes, two methods are proposed, online and offline methods. The online method schedules the communication based on the learning historical tasks and channel bandwidth between the IoT devices and the edges servers, while the offline method schedule the traffic according to data sizes and the corresponding channel bandwidths between the edges servers and the cloud server. Results show that the proposed algorithm outperforms other DL optimization solutions.

In [84] a DRL approach that can predict spectrum usage of unknown neighboring networks in the near future to increase the throughput rate by optimizing the spectrum sharing using online supervised learning in a multi-agent setting. The proposed approach bases on five main operations: the spectrum monitor, preprocessing unit, predictor unit, probability matrix and the overall scheduler. Hence, each node starts by capturing the energy on the overall spectrum, then this monitor information is forwarded to the next component, the preprocessing unit. The preprocessing unit of the node creates the correct observation values. For every slot in the superframe, the predictor unit predicts the upcoming spectrum usage as highly used or free, then the predicted values form a probability matrix that the scheduler uses conjointly with the correct observation values to schedule the communication in every slot in the superframe on any channel to avoid collisions with neighboring networks or other electromagnetic sources. The DRL is trained in an online way. The results show a reduction in the number of collisions with the surrounding networks of 30%, and an increase in throughput rate by 10% in a medium-sized network.

The multi-channel scheduling problem is addressed in [85]. The aim is to estimate all of the N processes using the M (M<<N) offered channels. The problem is formulated as Markov decision process (MDP) then this MDP is solved using a DRL by central gateway after receiving different information from the distributed sensors to estimate the states of multiple remote, dynamic processes. The proposed algorithm is compared to other scheduling algorithms such as round-robin and reduced-waiting-time and is shown a significant outperforming over these algorithms in many example scenarios.

In [86] a DRL-based Relay Selection Scheme, named DQ-RSS is proposed to realize path transmission sharing, and thus improving the throughput in relay-based WSNs. In DQ-RSS, a deep-Q-network (DQN) is trained according to the outage likelihood probability and channel efficiency information, and the optimal relay is selected from multiple other candidate relay nodes without the need for a network model or prior data. Therefore, a DQN is used to process high-dimensional state spaces and accelerate the learning rate. The results indicate that DQ-RSS can achieve better performance and save the convergence time compared to existing schemes.

A summary state of comparison between the different RL-based algorithms is presented in table 2.6.

3.6.3 Discussion

Based on the previous comparison, we can observe the following observations: that a few works are focused on energy balance techniques compared to the other protocols that are not, despite its importance in the energy-conserving aspect that plays an important role in improving the lifetime of the network. the following results which lead to future works:

☞ Despite the importance of energy balancing techniques in the energy-conserving aspect that plays an important role in improving the lifetime of the network, few works focus on them compared to the other protocols, which represent only 23.8% from the overall studied protocols.

☞ Most of the cooperative RL are focusing on CRSNs that have more capabilities than the WSNs as explained before and few cooperative multi-channel algorithms that take into account the resource constraints of WSNs.

☞ Finally, most of the cited algorithms are more complex in their performance by the fact that they periodically collect control packets, especially in cooperatives RL ones. They have based on reducing the number of the exchanged learning messages by communicating only with the direct neighboring nodes, without tackling the scheduling issue in the transfer learning process, which can result in a high cost of collisions leading to more energy consumption and negative transfers.

☞ Most of the CRSN protocols focus only on the licensed bands despite their competence to use the unlicensed bands.

3.7 Conclusion

In this chapter, the multi-channel communication in WSNs and RL approach are covered, and most protocols and schemes that are proposed in different types of RL based multi-channel wireless sensor networks are explained, which can be divided into three categories: Single-agent RL, cooperative multi-agent RL and deep RL, for addressing optimal solutions for multi-channel communication. Finally, we have discussed the characteristics of those proposals and brought out the different possible observations.

TABLE 3.3: Comparison between the different RL-based algorithms

Reference	Algorithm Type	Network topology	Interfrence-free	Band used	Energy balance	Medium access
[61]	Single Q-learning	CRSN	✓	Both licensed and unlicensed		TSMA
[62]	Single Q-learning	Tree based WSN	✓	Both licensed and unlicensed		TDMA
[64]	Single Q-learning	Distributed WSN	✓	Both licensed and unlicensed		CSMA/CA
[65]	Single Q-learning	Distributed WSN	✓	Only unlicensed	✓	TDMA
[66]	Single Q-learning	IIoT	✓	Only unlicensed		IEEE802.15.4
[67]	Single Q-learning	Star WSN	✓	Only unlicensed	✓	CSMA/CA
[68]	Single MAB	CRSN	✓	Only licensed		TDMA
[69]	Single MAB	CRSN	✓	Only licensed		TDMA
[70]	Single MAB TOW	Distributed WSN	✓	Only unlicensed		IEEE 802.15.4
[71]	Single MAB	CRSN	✓	Both licensed and unlicensed		TDMA
[72]	Single MAB	Distributed WSN	✓	Both licensed and unlicensed	✓	TDMA
[75]	Cooperative Q-learning	CRSN	✓	Both licensed and unlicensed	✓	TDMA
[76]	Cooperative Q-learning	CRSN	✓	Only licensed		TDMA
[77]	Cooperative MAB	CRSN	✓	Only licensed	✓	TDMA
[78]	Cooperative Q-learning	CRSN	✓	Both licensed and unlicensed		TDMA
[79]	Consensus Q-learning	CRSN	✓	Only licensed		TDMA
[82]	DRL	CRSN	✓	Both licensed and unlicensed		TDMA
[83]	DRL	Multimedia IoT	✓	Both licensed and unlicensed		TDMA
[84]	DRL	Distributed WSN	✓	Both licensed and unlicensed		TDMA
[85]	DRL	Star WSN	✓	Only unlicensed		TDMA
[86]	DRL	Relay-based WSN	✓	Both licensed and unlicensed		TDMA

Chapter 4

Contributions

4.1 Introduction

This chapter presents the description of our contributions presented by the proposition of four multi-channel WSN protocols. The first protocol, named Multi-Channel Scheduling Protocol

(MCSP) for Wireless Personal Area Networks (WPAN) IEEE802.15.4, focuses on overcoming the collisions problems that represent one of the main energy-draining sources in WPANs IEEE802.15.4, by multi-channel scheduling scheme. The second protocol, called Energy-efficient Reinforcement Learning Multi-channel MAC (ERL MMAC) for WSNs, bases on the enhancement of the energy consumption that is considered as the main challenge for the WSNs, by reducing as much as possible the collisions problem on one hand, and by balancing the energy consumption between the nodes on the other hand, in using of single-agent RL. Whereas, the third work represents a proposition of a new heuristically accelerated RL protocol named Heuristically Accelerated Reinforcement Learning approach for Channel Assignment (HARL CA) algorithm, with the aim is to reduce the number of learning iterations in an energy-efficient way taking into account the bandwidth aspect in the scheduling process. Finally, a proposition of a new cooperative multi-agent RL approach named Cooperative multi-agent RL for Channel Assignment (CRLCA) in WSNs that have missed cooperative algorithms such as is observed in the previous chapter. Therefore, CRLCA improves the acceleration learning of the cooperative RL by using an accelerated learning model and overcomes the extra communication overhead problem of the cooperative RL by using a new self-scheduling method that ensures a balancing energy consumption between the neighboring nodes. Ultimately, we have addressed the use of our approach CRLCA in a Static environment by using the algorithm SCRLCA, and in a Dynamic environment by using the algorithm DCRLCA.

4.2 Multi-Channel Scheduling Protocol for Wireless Personal Area Networks IEEE802.15.4

4.2.1 IEEE 80.15.4 LR-WPAN Overview

IEEE802.15.4 is a standard that specifies the physical and medium access control for Low-Rate, low-cost and low-power Wireless Personal Area Networks (LR-WPANs), and is maintained by the IEEE 802.15 working group. As mentioned before, it represents the basis for several higher protocols such as Zigbee and 6LoWPAN, each of which extends the standard by developing higher layers that are not defined in IEEE 802.15.4. In the IEEE 802.15.4 framework, the networks can be built as either peer-to-peer or star networks with two types of nodes, Full-Function Device (FFD) and Reduced-Function Devices (RFDs). The difference between the FFD and the RFD is that the first is used for sensing, coordinating and relaying data, whereas the second is a simple device with modest resources, used for only sensing and direct communication. The framework conceives a communication in a range of 10-meter with a transfer rate of up to 250 kbit/s based on two communication modes in the medium access, beacon-enabled and beacon-less modes. The first mode is used for star or hybrid topologies where the nodes are scheduled using a broadcasting beacon frame to coordinate the transmission to the coordinator, however, the second mode is used for full peer-to-peer topology where the nodes communicate based on CSMA/CA algorithm. An example of hybrid topology is illustrated in figure 4.1. In the beacon-enabled, a slotted super frame named Beacon Interval (BI) is used to manage the communication between the nodes. It consists of two main parts, the active and the inactive parts as shown in figure 4.2. The active part, also named Super-frame Duration (SD), takes 16 slots started by beacon transmission in the first slot then ten slots for the portion named Contention Access Period (CAP), where network nodes use the slotted version of the CSMA/CA algorithm for the channel access, and finally five slots for the portion named Contention Free Period (CFP), where the access to the channel is guaranteed to each device for a period determined in a number of slots, called GTS (Guarantee Time Slot). In spite of being able to avoid data packets colliding with each other

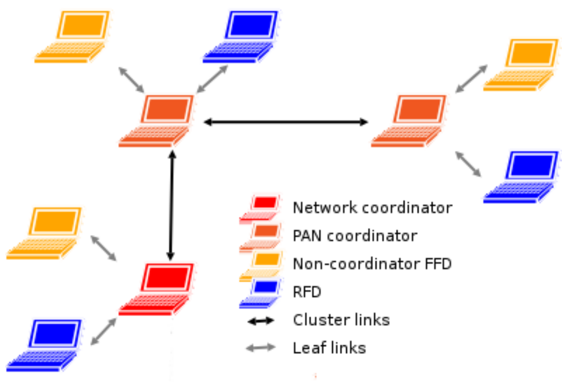


FIGURE 4.1: Example of LR-WPAN hybrid network.

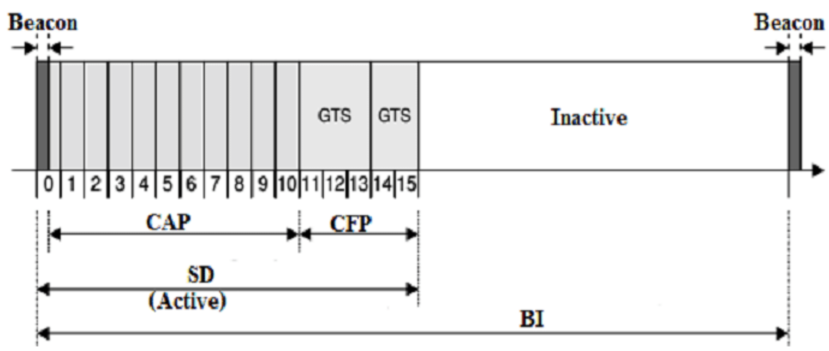


FIGURE 4.2: Beacon enabled super-frame structure.

by using either slotted CSMA/CA or GTSs, the risk of beacon collisions that could paralyze the network, still remains especially in the hybrid network topology where there exists more than one coordinator. In order to solve this problem, we have proposed a novel multi-channel scheme named Multi-Canal Scheduling Protocol (MCSP) for IEEE 802.15.4 LR-WPAN, by exploiting the available multiple channels offered by the IEEE802.15.4 as explained before in section 2.7.3.

4.2.2 Beacon Collisions problem in IEEE 80.15.4 LR-WPAN

As presented in the figure 4.2, the beacon frames are sued periodically by the coordinator nodes to synchronize and schedule the communication in the network. they play an important role in the successful communication of all the WPAN devices in the network. Due to the punctual importance of issuing the beacon frames, they have the highest priority in the network. Unlike the normal data or control packets which can be scheduled using slotted CSMA/CA or GTSs, the standard does not propose any scheduling for the beacon frames except the use of the back-off algorithm. However, it is possible that a number of coordinators in the same WPAN transmit different beacon frames in order to manage their proper clusters. In this case, collisions may occur between these beacon frames. Therefore, it might result in unexpected network panic. The cases in which the beacon collision problem is susceptible to occur can be divided into two categories in terms of the location of the coordinators: direct and indirect beacon frame collisions as shown in figure 4.3. Direct beacon frame collisions occur when two or more coordinators are in the transmission range of each other (direct neighbors or parent-to-child relation) and send their beacon frames at approximately

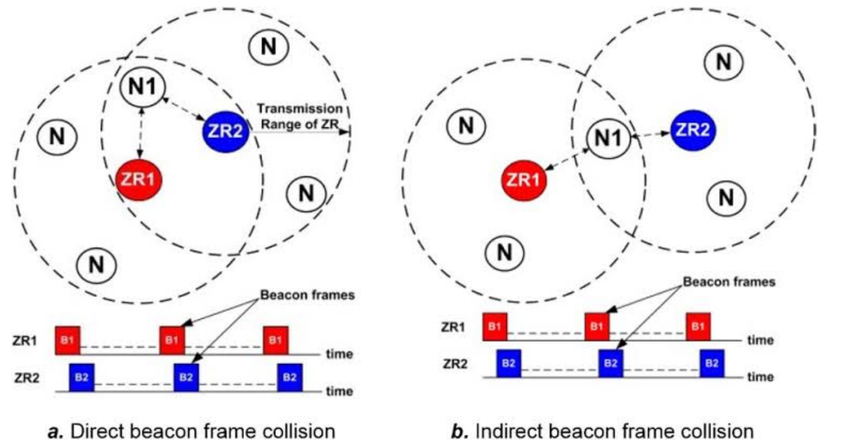


FIGURE 4.3: Beacon enabled super-frame structure.

the same time, as shown in figure 4.3 (a) where the node N1 is associated with ZR1 and ZR2 is another coordinator. If ZR1 and ZR2 transmit their beacon frames at approximately the same time, node N1 may lose the beacon information due to the collision of the two beacons. If the superframe duration (SO) of the two coordinators is the same, their beacons will be continuously in conflict with each other at the node N1 and the two coordinators will not be aware of the collisions. In contrast to direct beacon collisions, the indirect beacon frame collisions occur when two or more coordinators cannot hear each other (indirect neighbors), but have overlapped transmission ranges and transmit their beacon frames at approximately the same time, as shown in 4.3 (b) where the node N1 is located in the overlapped region of the transmission ranges of ZR1 and ZR2 coordinators, which leads to the fact that the node N1 will not be able to correctly receive the beacon frame from either coordinator, since the beacons will collide with each other.

4.2.3 Multi-Channel Scheduling Protocol

In this section, a new beacon collision avoidance algorithm named Multi-Channel Scheduling Protocol (MCSP) is introduced. The proposed algorithm utilizes the opportunity of multiple channels existing in the protocol IEEE802.15.4 to assign them in a static manner, supposing the existence of a multi-radios network coordinator. This assignment devises the whole network into multiple disjoint sub-networks in an optimal manner. Each sub-network uses an independent channel in order to avoid the inter sub-networks beacon frame collisions. Also, nods on the same channel can be scheduled by using the optimized Beacon-Only approach proposed in [87], in order to avoid the intra sub-network beacon frame collisions.

Even though [88] and [87] tried to efficiently avoid collisions in LR-WPAN IEEE802.15.4, their propositions remain focusing on the existence of only one channel which makes them suffer again of the beacon collisions especially when the number of nodes increases (collisions happen if more then 7 nodes exist, in [87]). On the other hand, multi-channel solutions depending on fully fixe channels assignment strategy succeed in the avoidance of inter sub-networks beacon frame collisions with using of a small number of channels such as Tree-based Multi-channel Communication Protocol (TMCP) proposed in [44] that is explained in the previous section 3.4, and seems the more adaptable for the protocol IEEE802.15.4. But it didn't give a solution about the intra sub-networks beacon collisions. Therefore, to overcome the limitations of these solutions, we have proposed our method MCSP that combines the advantages of the two proposed algorithms in [87] and [44].

Therefore, the new multi-channel method combines the fixed channel assignment and the optimized Beacon-only approach to maximize the avoidance of the beacon frame collisions in the protocol IEEE802.15.4. The whole network is divided into multiple disjoint sub-networks by the base station which allocates different channels to each sub-network, and then forward each flow only along its corresponding sub-network. The superiority of MCSP is two-fold. First, it can increase network throughput and eliminate inter and intra-sub networks collisions and exploit spatial reuses of parallel transmissions among sub-networks. Secondly, for practical concerns from the fact that MCSP requires much fewer channels and it does not need a sophisticated channel coordination scheme, which implies that MCSP supports the specificity of the IEEE802.15.4 standard by the fact that it can work without the need for time synchronization between the channels. Hence, as shown in figure 4.4, MCSP performs as follow:

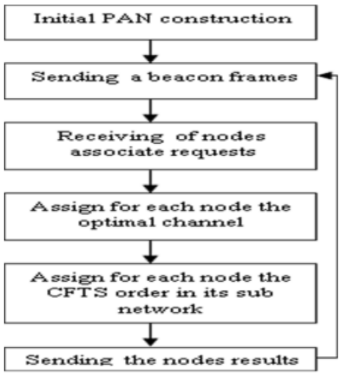


FIGURE 4.4: MCSP steps.

At the initial stage, the network coordinator performs an active scan to select a PAN identifier prior to starting a new network coordinator identifier. Then, it performs an ED scan to detect and save the set of available channels. Among all channels, it selects the non-adjacent channels according to the number of the existing radios. After that, the network coordinator starts the beacon sending through the different radios. After receiving the beacon frame, the nodes transmit association requests to the network coordinator in order to be allowed to join the network. The network coordinator executes the MCSP algorithm depicted in 4.4 to assign for each node the optimal channel (sub-network) with its node parent and the Contention Free Time Slot (CFTS) order of its parent in this sub-network. The CFTS represents the order of transmitting the beacon message in Beacon only Period (BP) using the same channel, and its assignment must respect the following rules:

1. The coordinator CFTS should be different from its neighbor CFTSs, thus its parent node.

2. given a CFTS set organized in an increasing order from index 0 to $n-1$, coordinator R_i CFTS index will not be higher than the one given to its parent parent.

Consequently, MCSP begins by the fact of researching for each node the sub-network that shares the least number of coordinator neighbors (FFD). This number must be superior to 0 in order to ensure connectivity. The first shared FFD node will be selected as the node parent. If more than one sub-network, MCSP selects those that have a parent with the least number of FFD neighbors in order to avoid as much as possible the beacon frame collisions. Furthermore, MCSP selects those that have the least size in order to construct balanced sub-networks as much as possible. If more than one sub-network responds to the previous criteria, the choice is random. The allocation process of CFTS order for each node within the sub-network is performed as follows. MCSP marks the variable CFTS with the value 0 if the node is not a coordinator (RFD). However, if the node is an FFD, it takes the value of its CFTS parent then it checks whether this value is different from the other neighboring nodes' CFTS values as well as the parent neighboring nodes' CFTS values in common with the sub-network. If the value exists, MCSP increments it and then resumed the audit until the obtaining of nonexistent value. In this manner, MCSP ensures optimal CFTS management by applying the rules mentioned above. In this way, the collisions among different sub-networks (inter sub-network collisions) can be eliminated by the above channel assignment, and the potential collisions among nodes within the same sub-network (Intra sub network collisions) are avoided using CFTSs.

The goal of this partitioning is the decrease potential beacon frame collisions as much as possible, increase the throughput and delivery rate and balance the energy consumption between the sub-networks. Finally, the network coordinator transmits the results that combine the channel, the parent, the CFTS order and the parent CFTS order (CFTSp) to the node. The node, in turn, switches to the new channel and begin to transmit its own beacon frames in the order received after it receives the patent one if it is a coordinator node (FFD) by calculation the required time using the Eq 4.1:

$$T_{sb} = (CFTS_u - CRTS_p)P_R \tag{4.1}$$

Where T_{sb} is the required time for sending the own $CFTS_u$ after receiving the one the the parent coordinator, $CRTS_p$ is the order index of the parent coordinator and P_R is the required time to receive a physical packet with maximum size according a certain delivery capacity db, and is calculated using 4.2.

$$P_R = \frac{aMaxPHYPacketSize \quad . \quad 8}{db} \tag{4.2}$$

Each node calculates the end of Beacon only period (BP) after the reception of the parent beacon frame, using Eq 4.3 as follows:

$$BP = ((CFTS_{Max} + 1) - CFTS_u)P_R \tag{4.3}$$

Each node picks the CFTSmax value within each beacon frame updated by the network coordinator. This period must always maintain the minimum period of CAP section. For this reason, the network coordinator must verify the validity of CFTSu before sending it to the concerned coordinator using 4.4.

$$CFTS \leqslant (\frac{(SD - aMinCAPLength)DS}{P_R}) - 1 \tag{4.4}$$

Where SD represents the Superframe Duration defined in the standard using 4.5, and DS represents the symbol duration [28].

$$SD = (aBaseSuperframeDuration2^{SO})[symbols] \tag{4.5}$$

The proposed algorithm is presented in Algorithm 3.1.

4.2.4 Performance Evaluation

To evaluate the performance of MCSP, the network illustrated in figure 4.5 is used, with 4 variant protocols:

Algorithm 3.1: MCSP

Input:sub network List SL, The network coordinator NC, For each node U its nature N and the set V_u of FFD neighbors within 1 hop.

Output For each node U: optimal channel C_u, the parent P_u, $CFTS_u$ and $CFTS_p$ order indexes.

For each channel i **do**

    Insert(SL_i, CP, null,i)

End for

Sort Global_Node_Liste in descending order by the number of neighbors,
then organize FFDs before RFDs according to the nature N.

For each node U in Global_Node_Liste **do**

    $C_u = null$, $P_u = null$, $CFTS_u = 0$

    Find optimal sub network SN(SL, U,C_u,P_u, $CFTS_p$)

    Find optimal CFTS(SL(C_u), P_u, $CFTS_p$,$CFTS_u$)

    Insert (SL(C_u),U,P_u,$CFTS_u$)

    Send (U,C_u, P_u, $CFTS_u$, $CFTS_p$)

End for

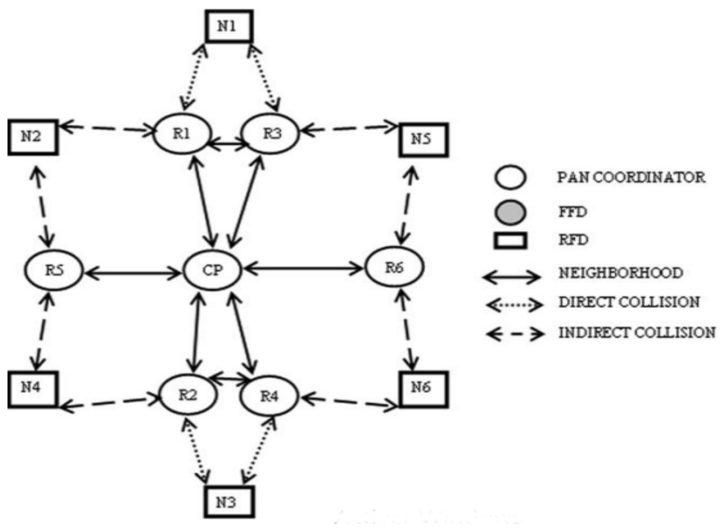


FIGURE 4.5: Experimental model.

1. The classical IEEE802.15.4 protocol.
2. Uni-channel MCSP with CFTS management.
3. MCSP without CFTS management using 2 channels.
4. Complete MCSP with using of 2 channels.

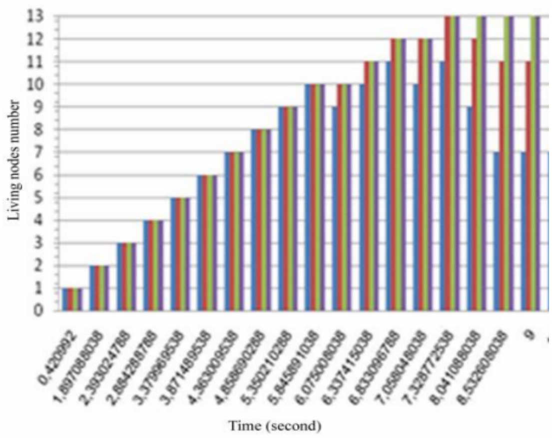


FIGURE 4.6: Living nodes with SO=3

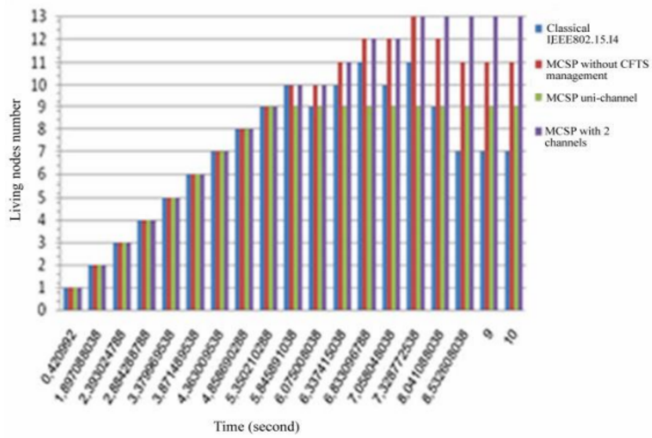


FIGURE 4.7: Living nodes with SO=1

Figures 4.6 and 4.7 depict the sum of living nodes with respect to time in the 4 protocols. The original IEEE 802.15.4 doesn't avoid the two types of collisions, the direct collision of N1 then N3 and the indirect collision of N2, N4, N5 and N6 respectively. However, MCSP without CFTS management succeeds in the avoidance of inter sub-network beacon collisions, but it doesn't succeed concerning the intra sub-network collisions within each sub-network represented by the node N2 between the two coordinators R1 and R5 in the first sub-network then the one of the node N5 between the two coordinators R3 and R6 in the second sub-network. The two other protocols are succeeded in the save process of all the nodes from the two kinds of collisions avoidance by using of 1 channel as well as two channels if the Superframe Order (SO) is equal to 3 which means that there is a sufficient beacon only period to schedule all the nodes in both cases. However, if SO equals 1, the beacon only period covered all the nodes only with using of two channels (from 8 nodes with uni channel to 12 nodes with two channels) by the fact of minimizing the scheduled node from 12, in MCSP with 1 channel, to 6 nodes in MCSP with 2 channels which means that the use of only two channels increases the accepted node in scheduling operation by 50%.

In figures 4.8 and 4.9, the energy consumption average versus the 200 first seconds is presented. In order to cover all the nodes in different MCSP protocols, the superframe order

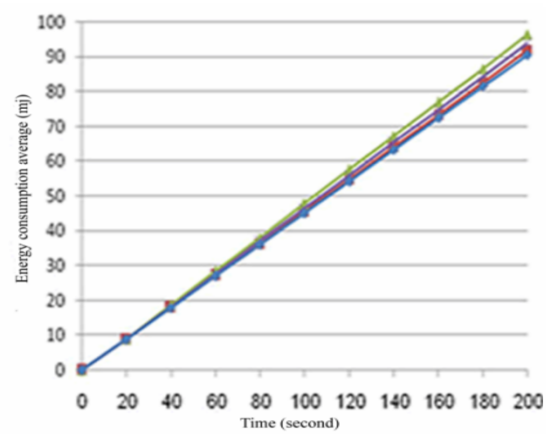


FIGURE 4.8: Energy consumption without PAN.

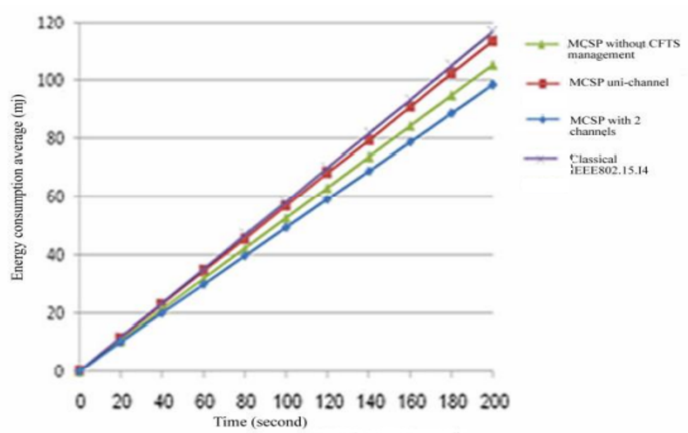


FIGURE 4.9: Energy consumption with PAN.

(SO) is fixed by the value 3. In this simulation, the different protocols consume less energy because of the less number of hearing packets after the channel assignment process made by the network coordinator (PAN) if we don't take into account the energy consumed by this latter. However, if this energy is taken into consideration, MCSP uni-channels consume more energy than MCSP multi-channels because of the beacon frames duplication operation in transmission made by the network coordinator in the different radios.

4.3 Energy-efficient Reinforcement Learning based Multi-channel Data Gathering for Wireless Sensor Networks

An Energy-efficient method for Reinforcement Learning based Multi-channel MAC (ERL MMAC) that performs a hybrid channel assignment using a decentralized tree for multi-channel data gathering in WSNs is proposed. The proposal focuses on the reduction of energy consumption by using the least chosen default channel allocation in two hops rather than one hop in order to reduce as much as possible the conflict links in one side, and use of parent selection strategy rather than parent default channel selection strategy in the learning phase to avoid the redundant data messages, in the other side.

4.3.1 System model

We define the WSN as a set of N nodes denoted as N=$\{1,...,N\}$, randomly dispersed throughout an area to sense data and convey it to the sink. Each node uses a predefined set of non-overlapping channels K=$\{f_1,...,f_k\}$. The time is divided into consecutive frames which in turn divided into a fixed number of slots. In order to reduce the energy consumption and the end to end delay as much as possible, we have adopted a decentralized tree-based topology with a hybrid multi-channel communication approach. Hence, the network performance is realized through three consecutive phases, initialization, learning then communication phases.

4.3.2 Initialization phase

At the beginning of the initialization phase, each node synchronizes with the sink via the Beacon messages delivered periodically by the sink and broadcasted by the other nodes using a dedicated channel. During the synchronization process, each node discovers its M neighbors located within their communication range. Before the sending of the Beacon messages, each node increments the lowest depth D received which is included in the Beacon message and represents the number of hops from the sink. Note that the D value starts by the value 0 at the sink. To discover the lowest D from the received Beacon messages, each node, except the sink neighboring nodes, must wait a threshold of time before the sending of its own Beacon message (in our case, we have used a much sufficient time considered by one frame). Therefore, each node keeps the lowest number D, and a list of its parent nodes that is selected from the reliable list of neighboring nodes. The parent nodes are the neighboring nodes that have a D smaller than that of the node. All the nodes can communicate with the sink node in a multi-hop fashion through the communication with their ($C \subseteq M$) parent neighboring nodes.

To adopt a hybrid channel assignment strategy, the nodes perform the default channel selection process. Unlike [**11**, 62, 63] which use one hop least chosen channel method that remains vulnerable to the indirect collisions even in the presence of a sufficient number of channels, we have used two hops least chosen channel method beginning from the sink to the leaf nodes in such a way that the nodes with the lowest depth are prioritized than the nodes with higher depth. If the neighboring nodes have the same depth, the node with the lowest identifier (ID) is prioritized. Each node selects its default channel only if it receives the default channels chosen by its prioritized neighboring nodes in two hops. The aim is to reduce the overhead communication and the interference costs as much as possible.

The figures 4.10 and 4.11 show the average results of the collided links obtained on 10 different experiments of 100 nodes dispersed in an area of 100100m. The Figures show the indirect and the global collided links. Each node uses a radio communication range of 15m and a variety of channels ranging from 2 to 10. In order to evaluate the same network by the two methods, each of the ten experiments is run by the two methods. The obtained results show that our method can optimize the avoidance of the global collided links by an average of 2.96% for all used channels and by 7.32 % in the case of two used channels.

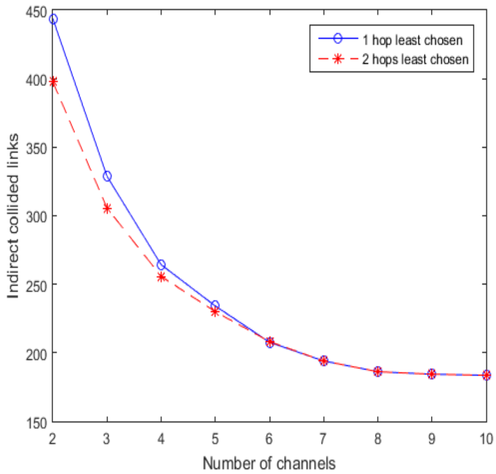


FIGURE 4.10: Indirect collided links over different channel.

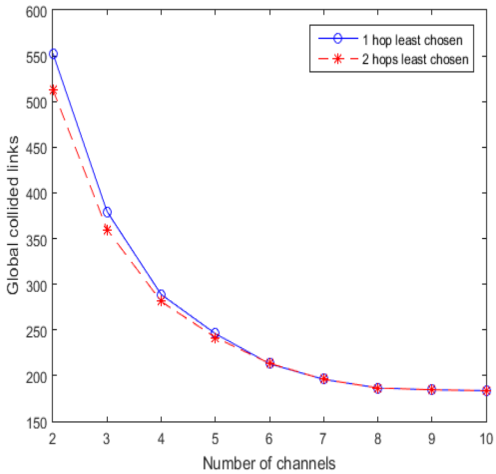


FIGURE 4.11: Direct and indirect collided links over different channel.

4.3.3 Learning phase

To obtain both channel and time (channel-time) schedule for each time slot, each node must use the RL approach to learn at each slot the best action to be followed after this period. As mentioned before, RL can be formulated as a Markov Game that can be expressed mathematically as MG=$\{S, A, T, R_{1...N}, \gamma\}$, where;

- S is the set of states of the agent;
- A is the action set of the agent ;
- *T:SAS*$\longrightarrow$*[0,1]*, is the state transition function,
- R_i:*SAS*$\longrightarrow$**R** is the reward function of the agent i.
- $\gamma \in$ *[0,1)*, is the discount factor, which represents the relative importance of future and present rewards.

Consequently, we consider each node as an agent with a stateless variant since we have a temporal difference problem without states. Each agent has a set of action textitA defined as follow: A=$\{a_1, ..., a_C, R\}$, where $a_i/1 \leq i \leq C$ means the sending to the parent i using the parent default channel, while R is receiving on the default channel. Referring to [**11**], we use the strategy *"win stay, lost shift"* that plays an important role in the acceleration of the learning process by the fact that if the agent fails in performing action after some successes, it does not lose time in the gradual return. Hence, the transition function is based on the reward that can be either successful (*"win stay"*) or failed communication (*"lost shift"*). Therefore, the discount factor γ is set to 0.

In the beginning, each node looks if it has a data message in its queue. If it is the case, it selects randomly an action among its set of actions then it performs this action and looks at the reward, otherwise it selects the receiving action since it has no data message to send. The reward depends on the nature of the selected action. if it is a sending action to one parent, the successful action is considered when the sender node receives an Ack message from this parent and the node keeps selecting this action in the same slot for the next frames, otherwise, the node must select another action in the same slot for the next frame. On the other side, if the action is receiving on the default channel, the successful action is considered when the node receives a data message proper to it and the node keeps selecting this action in the same slot for the next frames, otherwise, the node must select another action if it is a sender. Before each sending, the sender node must sense the channel for time W_t measured by 4.6as follow:

$$W_t = e^{\frac{1}{\beta+1}} \tag{4.6}$$

Where, β represents the number of success of the selected action in this slot. The aim is to give priority to the previous successful nodes than the new ones in the allocation of the channel. The different steps are presented in the flowchart of Figure 4.12.

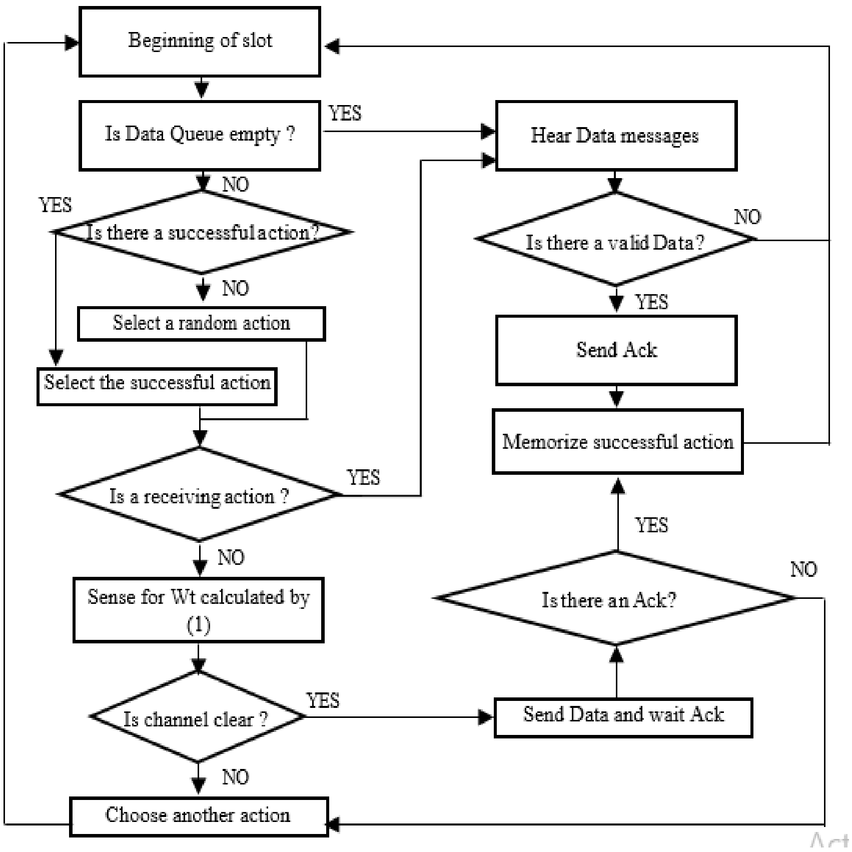


FIGURE 4.12: Different learning steps

4.3.4 Communication phase

After the learning period, each node looks for each slot at whether its sleeping probability (P_{SLEEP}) is greater than a threshold value Δ as it is presented in Algorithm 3.2. If it is the case, it takes the sleeping action during this slot as the slot action (lines 18, 19), otherwise, it takes the best-performed action in the learning period as the slot action (lines 21, 22). Furthermore, the new nodes can discover their parents through the beacon messages that are transmitted periodically, and start the learning process with the sleeping slots of their parents. In order to accelerate the learning process of the new nodes, each parent keeps a blacklist of channels (Bl_{Chs}) for each slot, created during the learning phase, which includes the different used channels by the neighboring nodes in the concerned slot. During the negotiation between the new node and its parents, each parent sends the list of sleeping data time slots as well as the Bl_{Chs} for each data time slot to be excluded from the available actions of the concerned slot. In addition, when the learned parent of the node doesn't respond three times in a specific slot, the node starts a new learning process with its parents that have a sleeping action in this specific slot.

Algorithm 3.2: ERL MMAC

1\. Initialize Q table,Δ,P_{Sleep}
LEARNING
2\. **while** not end of learning frames
3\. **for** each frame
4\. **for** each time slot
5\. do LEARNING scheme depicted in Figure (4.12)
6\. **if** there is an action to do **then**
7\. initialize P_{Sleep}
8\. observe Reward
9\. update Q-table
10\. update Data queue
11\. **else**
12\. Increment P_{Sleep}
13\. **end if**
14\. **end for**
15\. **end for**
16\. **end wile**
SELECTION
17\. **for** each data time slot in the frame
18\. **if** $P_{Sleep} > \Delta$ **then**
19\. Slot-action <– sleep
20\. **else**
21\. select a* with Q(a*) > Q(a) for all actions
22\. Slot-action <– a*
23\. **end if**
24\. **end for**

4.3.5 Performance Evaluation

Herein, we present the simulation results of our algorithm ERL MMAC in comparison with RL MMAC ([62]). The different results are averaged over 10 simulations. Each simulation experiment is run for the same topology over the two protocols. Therefore, the sink is placed in the middle of the area and the nodes are fixed in their random places to execute each of the two protocols in a sequential manner. The leaf nodes generate data messages every 5 seconds. Each frame is divided into 10 slots, and the length of the slot is 40 ms. The length of the data message is 64 bytes, while Ack message is 12 bytes. Referring to ([62]) for the length of the learning period that depends on the maximum path to the sink node, we have used 15 frames which is more than enough.

We have used two scenarios: the first one is used to evaluate the learning period, while the second one is used to evaluate the network lifetime.

In the first scenario, ERL MMAC optimizes the energy consumption ratio by an average of 43.02% compared to RL MMAC as it is presented in Figure4.13. Note that the energy consumption ratio means the average of the consumed energy of all the nodes over the 10 experiments.

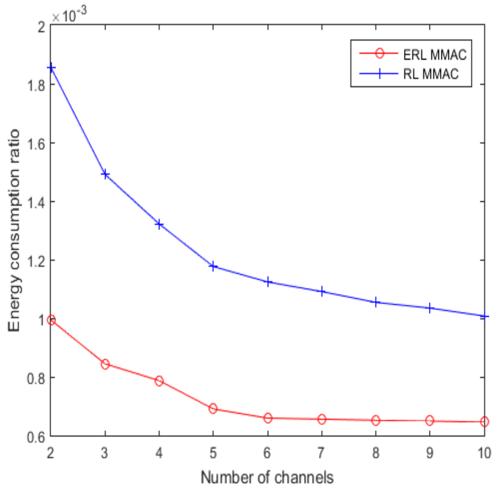


FIGURE 4.13: Comparison of energy consumptio in Learning period.

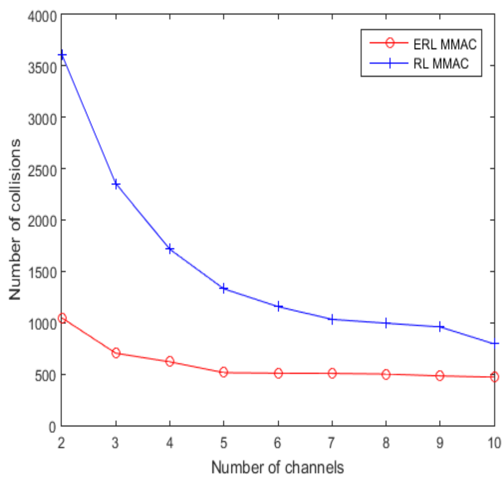


FIGURE 4.14: Comparison of collisions in Learning period.

This obtained result can be explained by the fact of decreasing collisions by 61.65% compared to RL MMAC as it is shown in Figure4.14. This is due to the 2 hops least chosen default channel used by ERL MMAC. Moreover, ERL MMAC has succeeded in the avoidance of duplicated data messages due to the parent selection method used by ERL MMAC since the parent channel selection method used by RL MMAC permit for all the parents that have received the data message to relay it as it is shown in Figure4.15.

In the second scenario, we simulate the two protocols for 120 minutes (two hours) in order to compare the network lifetime by measuring the average of the dead nodes in the 10 simulations as it is mentioned before. Hence, Figure4.16 shows that ERL can optimize the network lifetime by an average of 97.53% due to the energy preservation that is reached 41.03% as it is presented in Figure4.17.

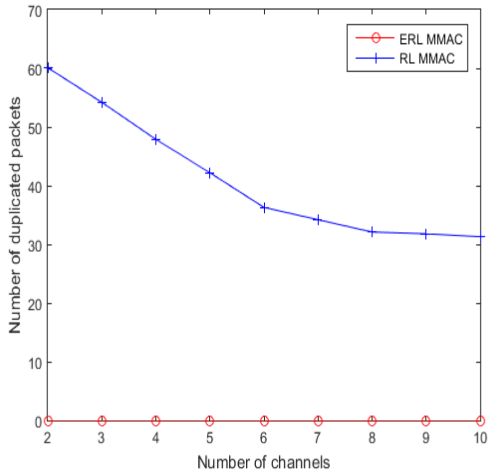


FIGURE 4.15: Comparison of duplicated data messages in Learning period.

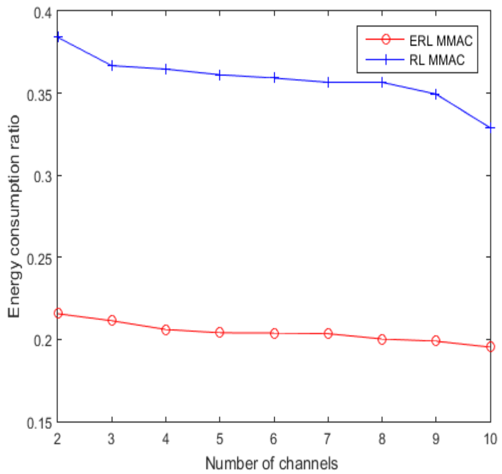


FIGURE 4.16: Comparison of dead nodes in tow hours.

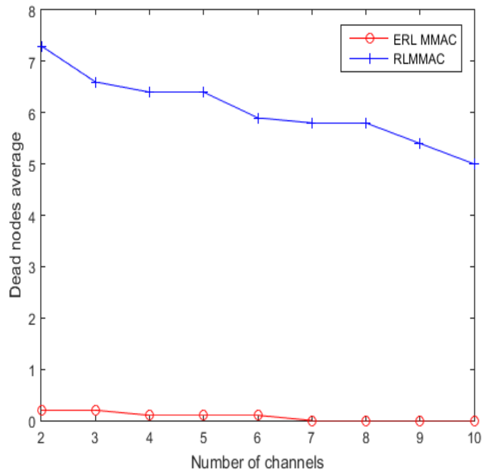


FIGURE 4.17: Comparison of energy consumption in two hours.

4.4 Heuristically Accelerated Reinforcement Learning for Channel Assignment in Wireless Sensor Networks

This section presents a description of our new protocol named Heuristically Accelerated Reinforcement Learning approach for Channel Assignment (HARL CA) in schedule-based distributed WSNs that is proposed to reduce the number of learning iterations in an energy-efficient way. The proposal considers the channel chosen by the other neighboring sender nodes as external information and uses it to accelerate the learning process and avoid collisions. Also, it takes into account the bandwidth of the used channel as an important factor in the scheduling process to increase the delivery rate.

4.4.1 Overview of Heuristically Accelerated Reinforcement Learning approach

The Heuristically Accelerated Reinforcement Learning (HARL) approach is a modification of the standard ε-Greedy used in RL by using a heuristic function H to influence the action choice during the learning process. As it is presented in [89], the action choice rule in HARL becomes as in (4.7).

$$V_{t+1}(s_t) = \begin{cases} argMax_{a \in A}[Q(s_t, a) + \xi H_t(s_t, a)] \\ \qquad\qquad\qquad \text{if q} \geqslant \varepsilon \\ a_{random} \qquad\qquad \text{otherwise} \end{cases} \tag{4.7}$$

Were; $H_t(s_t, a_t)$ is the heuristic function which influences the action choice. ξ is a real variable used to weigh the influence of the heuristic function. $H_t(s_t, a_t)$ is defined by (4.8) as follow:

$$H_t(s_t, a_t) = \begin{cases} max_{a \in A} Q(s_t, a) - Q(s_t, a_t) + \eta \\ \qquad\qquad\qquad \text{if } a_t{=}\pi^H(s_t) \\ 0 \qquad\qquad\qquad \text{otherwise} \end{cases} \tag{4.8}$$

Where η is a small real value and $\pi^H(s_t)$ is the preferred action.

4.4.2 System model

We define a WSN as a set of N nodes denoted as $N{=}\{1,...,N\}$, randomly dispersed throughout an area to convey regular traffic to the sink. Each node uses a predefined set of non-overlapping channels $K{=}\{f_1, ..., f_k\}$ in a range of communication R. The non-overlapping channels interval used is that including the channels proposed by the IEEE802.15.4 standard (16 channels in 2.4 GHz band and 10 channels in 915 MHz band) [3]. The network performance is realized through 3 consecutive phases, initialization, learning and communication phases.

In the initialization phase, each node discovers and synchronizes with its M neighbors located within their communication range through the Beacon message delivered by the sink and broadcasted by the other nodes after the increment of the depth D that is included and represents the number of hops from the sink. Note that the Beacon sending schedule process is out of the scope of this study. Thus, each node keeps the lowest number D, and a list of its parent nodes selected from the list of neighboring nodes. The parent nodes are the neighboring nodes that have a D smaller than that of the node. All the nodes can communicate with the sink node in multi-hop fashion through the communication with their ($C \subseteq M$) parent neighboring nodes. In addition, to adopt a hybrid channel assignment strategy, the nodes perform the default channel selection process. For that, we use the least chosen channel method to reduce, as much as possible, the communication overhead and interference. Hence, the method starts from the sink to the leaf nodes in a consecutive manner by prioritizing the nodes of the lower depth D to select the channels which have more bandwidth since they will be concerned by the communication more than the other nodes of increased depth.

In the learning phase, each node trains to get the best channel-time scheduling, while in the communication phase, each node follows the scheduling result from the learning phase.

The time is divided into consecutive frame periods which are in turn divided into a fixed number of slots started by the Beacon slot. Two slot types with the same duration are used, HARL and communication slots. The HARL slot is used in the learning phase, while the communication slot is used in the communication phase. Referring to [90], the HARL slot is divided into two sub-periods, the Scheduling (SP) and Transmission (TP) sub-periods as it is shown in Figure 4.18. The SP sub-period is used for the action choice process using a dedicated channel, and it is divided into a set of sub-lots. Each sub-slot is used for sensing the dedicated channel using a Backoff algorithm and exchanging RTS (Ready To Send) and CTS (Clear To Send) control messages. However, the TP one is used for data transmission and acknowledgment (Ack) messages using the selected channel in the SP sub-period.

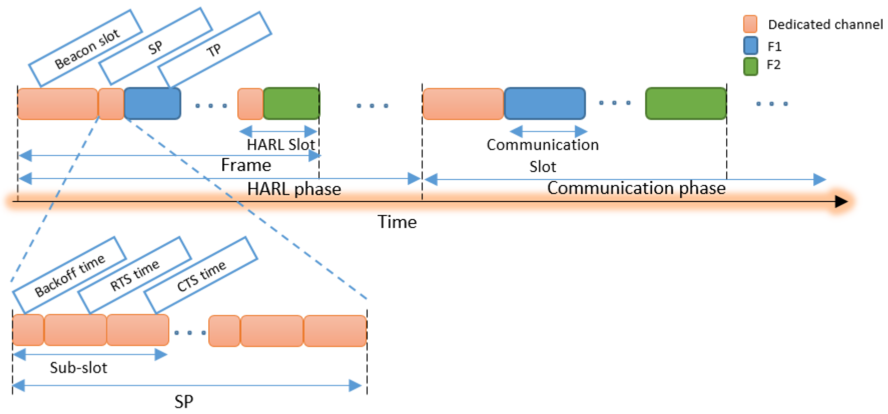


FIGURE 4.18: Time division in HARL CA

The communication slot is considered as learned slot and it reserves the totality of the slot for sending/receiving of the data message. Hence, the length of the data message is increased according to the length of the slot.

4.4.3 Problem formulation

To perform both channel and time (channel-time) schedule for each time slot, we consider each node as an agent that follows the MDP formulation cited previously. Hence, the different parameters are defined as follow:

1. The set of the environment states S is the set of the different couples $\{sl_i, a_i\}$, where sl_i represents the slot $i \in \{1,2,..., sl_{Max}\}$ and a_i is the action i from the available set of actions A.

2. The available actions A=$\{S_{p1}, ..., S_{pC}, R\}$, where S_{pi} means the sending to the parent pi, while R is receiving on the default channel. Note that each node starts the learning process by assigning the same probability to all the actions.

3. The state transition function is defined by (4.9), (4.10), and (4.11) as follow:

$$\begin{aligned} Q_{t+1}(s_t, a_t) &= Q_t(s_t, a_t) + \alpha(R(s_t, a_t) \\ &+\gamma V_t(s_{t+1}) - Q_t(s_t, a_t)) \end{aligned} \tag{4.9}$$

Therefore, since there is not a local optimum in our system, we use a trade-off between exploration and exploitation by putting the value of ε to 0. Hence, $V_{t+1}(s_t)$ will be as in (4.10).

$$V_{t+1}(s_t) = argMax_{a \in A}[Q_{t+1}(s_t, a) + \xi H_t(s_t, a)] \tag{4.10}$$

As mentioned above, our objective is to allow the node to take profit of the external information flow of the selected channels to select a free one. For this reason, the sender node must push the actions, which are not chosen by the other sender nodes in the same time-slot, to be preferred in the action choice process. To do that, we put the value of ξ to (0.001) and the heuristic function is calculated by (4.11).

$$H_t(s_t, a_t) = \begin{cases} 1 & \text{if } \beta < Pw(f) \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

Where β is the number of packets to be transmitted by the other neighboring sender nodes using the channel f that is related to the sending action a_t in the slot s_t, and $Pw(f)$ is the power of the chosen channel f. Channel power means how much the channel can be used at the same slot, which depends on the bandwidth of this channel and can take either the value 1 or 2, as it is explained in the next sub-section. The aim of putting the value of the ξ to (0.001) is to add a very small value for the desired actions to push them to be selected among the other actions with the same Q value without disturbing the weight of the previous successful actions since the unstudied choice of a value of ξ can force the node to choose a preferred action, without knowing its reward, rather than another already successful action. Figures 4.19,4.20 show an experiment that we have developed for 100 iterations of Q-values of two actions, the first action is always preferred (i.e. it is not chosen by the other neighboring nodes), while the second one succeeds in all the iterations. We have used $\alpha = 0.6, \gamma = 0.7$ and $R = 0.2$ in two cases; (a) $\xi = 0.05$ and (b) ξ =0.001. The results demonstrate that in (a), the preferred action can disturb the action choice after the first 10 iterations by forcing the node to select the preferred action rather than the successful one, while in (b), the action choice process is performed efficiently during the 100 iterations.

4. The reward function of the node is defined by (4.12).

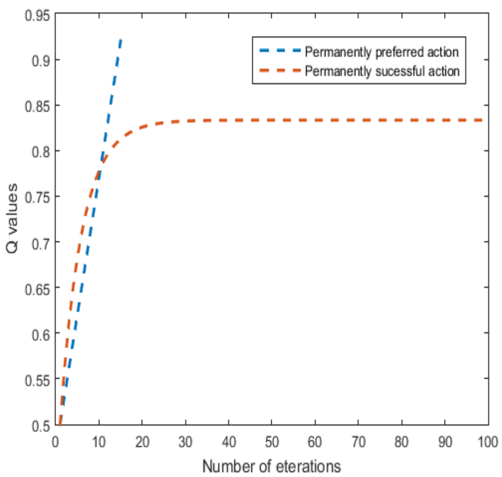


FIGURE 4.19: Comparison of two actions Q-values during 100 iterations with $\xi = 0.05$.

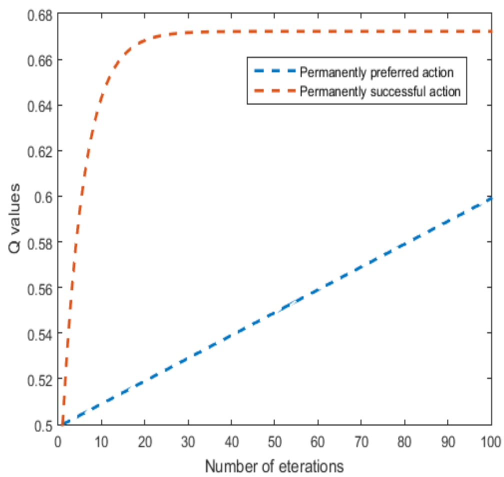


FIGURE 4.20: Comparison of two actions Q-values during 100 iterations with $\xi = 0.001$.

$$R(s_t, a_t) = \begin{cases} 0.2 & \text{if ther is an Ack} \\ 0 & \text{otherwise} \end{cases} \tag{4.12}$$

This means that the reward R takes the value 0.2 if either the sender node receives an Ack or the receiver node send an Ack for the transmitted data message via the channel chosen in

TP sub-period. Otherwise, it takes the value 0.

5. The discount factor γ must be carefully chosen to obtain increasing Q-values in the interval [0,1] since a bad choice of the γ value may result in decreasing Q-values and/or Q-values which do not belong to the interval [0,1]. It is closely related to the reward R as it is presented in figures 4.21,4.22 which show an experiment that we have developed for 100 iterations of a permanent successful action using $\alpha = 0.6$ and ξ =0.001 for two cases: (a) R =0.1 and (b) R =0.2. Hence, there is only the value (γ=0.9) which ensures the increasing Q-values in (a), but it tends to exceed the value 1, however in (b), we found two valid values of γ which are (γ=0.7 and γ=0.8). Therefore, as we have chosen R=0.2 for the successful action, the value of γ can be either 0.7 or 0.8.

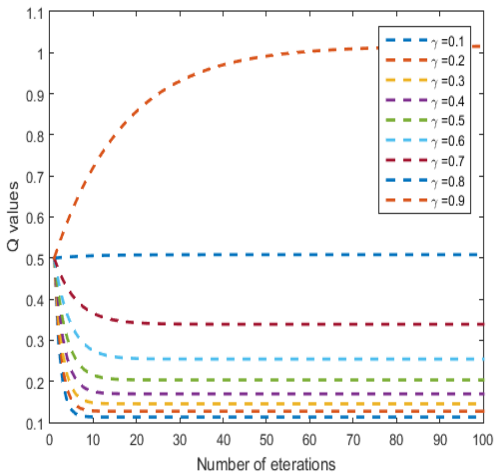


FIGURE 4.21: Comparison of γ values during 10 iterations with $R = 0.1$.

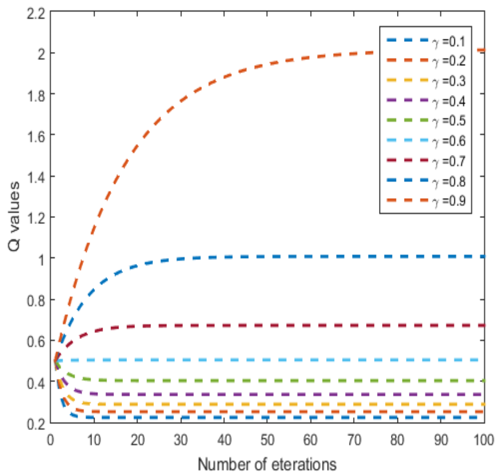


FIGURE 4.22: Comparison of γ values during 10 iterations with $R = 0.2$.

4.4.4 Bandwidth-based scheduling

Herein, we present the bandwidth-based scheduling model that is used in the HARL CA algorithm. As mentioned above, we use a schedule that bases on the effect of the bandwidth in the duration of the communication process, and thus in the duration of the message in the slot. As presented in [91], the Shannon-Hartley theorem determines the Capacity (C) of information transmission over a specific channel. It says that the capacity that is measured by bits per second is a function of the bandwidth of the channel (W), the power of the signal (S), and the noise value (N). This function is described in (4.13) as follow:

$$C = W \log_2(1 + \frac{S}{N}) \tag{4.13}$$

Hence, the increase in the bandwidth involves the increase in the capacity and thus increasing in the transmission duration. Therefore, if the bandwidth is multiplied by 2 ($W1 = 2W$) with the same signal power and noise, the capacity becomes double ($C1 = 2C$), which means that the transmission duration of an information using a specified channel will be reduced by the half if the channel will be changed by another larger in its bandwidth by the multiple. At this end, HARL CA takes into account the difference in the bandwidth of the used channels in consideration in the schedule process. As mentioned before, HARL CA uses two sets of non-overlapping IEEE 802.15.4 channels. Each of them have a specified bandwidth. The first uses a bandwidth of 2MHz in 915MHz band, while the second uses 5MHz in 2.4GHz band. Hence, HARL CA uses two values of the variable Pw which is used in the heuristic function H explained in the previous sub-section. The first value is 1 for each channel with 2MHz bandwidth, while the second is 2 for each channel with 5MHz bandwidth. Hence, the bandwidth-based for scheduling method used in HARL CA performs as follow:

1. It takes the 2MHz bandwidth as reference for the slot duration.

2. A channel with a bandwidth of only 5MHz is used as a dedicated channel that is used in SP sub-period and can be used in TP sub-period. The aim is to specify a small SP sub-period with the highest number of sub-slots.

3. Referring to (4.11), the TP sub-period of a receiver node can perform either only one communication using a channel with 2MHz bandwidth or two consecutive communications for the same sender node or for two different sender nodes using a channel with 5MHz bandwidth. An example is given in Figure 4.23.

4. In order to increase the throughput and delivery rate, the channels with 5MHz bandwidth are always preferred in TP sub-period than those with 2MHz bandwidth.

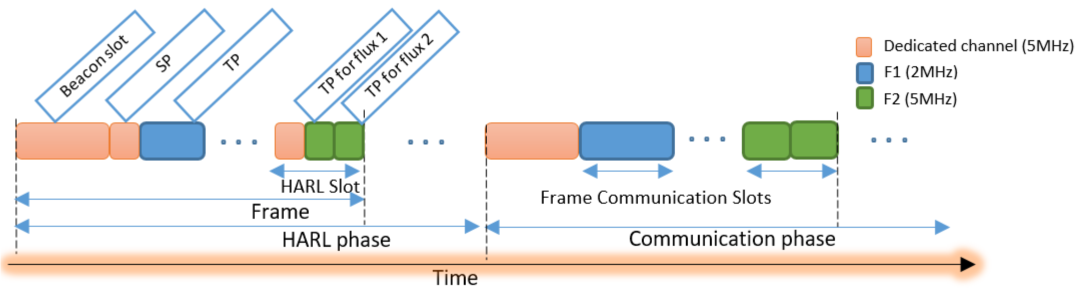


FIGURE 4.23: HARL CA bandwidth-based scheduling

4.4.5 HARL CA Algorithm

At the beginning of each HARL slot, each node starts by looking if its data queue is empty or not. If it is the case, the node chooses the receive action on its default channel and prepares to respond to valid RTSs based on the bandwidth of its default channel as follows:

1. If the bandwidth of its default channel is 2MHz, the node responds only one valid RTS for one packet.

2. If the bandwidth of its default channel is 5MHz, the node responds either one valid RTS for two packets or two different valid RTSs for one packet each, in such a way that, the first RTS is sent in the first half of TP sub-period, while the second RTS is sent on the second half.

If the receiver node has not received any RTS or it has not found any valid RTS until the end of SP sub-period, it turns off its radio until the next slot. However, if the queue of the node is not empty, it selects a random sub-slot in the SP sub-period then it begins listening to RTS control messages of the neighboring nodes until the arrival of its sub-slot. If the node receives, during the listening period, a valid RTS from one of its child nodes towards it, it responds by a CTS and it cancels its own sending of RTS. Otherwise, it selects an action using (4.7) which will be either a sending to one of its parents or a receiving on the default channel. If it is the first case, the node tries to send its RTS control message in the current sub-slot after performing a backoff algorithm to avoid the direct collision. Otherwise, it continues listening to find a valid RTS to respond until the end of the SP sub-period.

If the node which selects a sending action is succeeded in the sending of its RTS, it waits for a CTS from the selected parent in the same sub-slot. If it didn't receive any response, it considers its chosen action as unsuccessful and turns off its radio until the next slot. Otherwise, the node starts sending in the TP sub-period using the selected channel. The different steps are resumed in the flowchart of Figure 4.24.

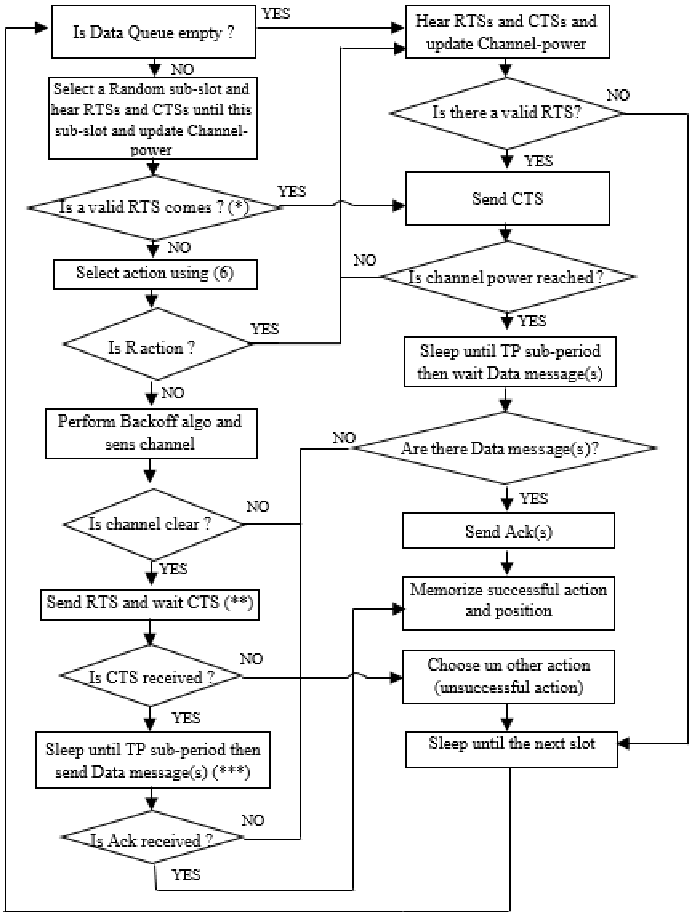


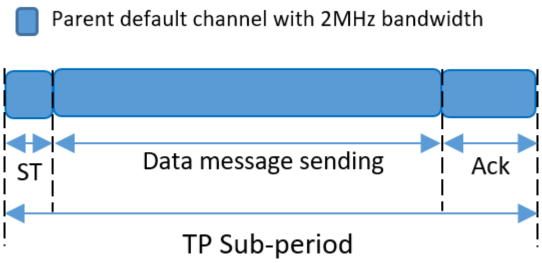
FIGURE 4.24: HARL CA learning steps

As already mentioned, each receiving node verifies the received RTS if it is a valid RTS or not (case mentioned by the symbol (*) in Figure 4.24). To be an RTS valid, the received RTS must verify the following conditions:

1. The receiver node is the concerned parent in the RTS control message.

2. The default channel of the receiver node did not reach its power, i.e. if it is 2MHz bandwidth channel, it must be not chosen by the other neighboring nodes, while if it is 5MHz bandwidth channel, it must be chosen at most once. Therefore, if the default channel is 5MHz, three cases can be happen (case mentioned by the symbol (**) in Figure 4.24):

FIGURE 4.25: TP sub-period of HARL slot



- ☞ The channel is not chosen at all and the sender node is mentioned two packets in the RTS. The receiver sends a CTS indicating two packets to be received in TP sub-period. Hence, the power of the channel is considered as reached.
- ☞ The channel is not chosen at all and the sender node is mentioned one packet in the RTS. The receiver sends a CTS indicating one packet to be received in TP sub-period at the first place. Hence, the power of the channel is considered as not reached.
- ☞ The channel is chosen once and its power is not reached, the receiver node will indicate the second place of the TP sub-period in the CTS control message for only one packet and the power of the channel is considered as reached.

The aim is to avoid indirect collisions.

To give priority to the successful actions, the strategy "win stay, lost shift" used in [62] is realized for the data message sending process in the TP sub-period, which is mentioned by the symbol (***) in Figure 4.24. Hence, each sender node must wait W_{tp} in the Sensing Time (ST) in the TP sub-period as it is shown in Figure 4.25, then it performs the Clear Channel Assessment (CCA) before the sending of its data message. If the channel is clear, the sender node starts sending, otherwise it considers the action as unsuccessful. The W_{tp} is calculated by (4.14).

$$W_{tp} = STe^{-\lambda} \tag{4.14}$$

Where λ represents the number of the success of this action at this slot. In this way, we ensure that the sender nodes which previously succeed wait less than the new ones.

In the algorithm that is presented in Algorithm 3.3, each node finishes its learning phase by comparing at each slot the maximum of the Q-values by a threshold (from line 18 to 25), if it is smaller than, the node considers it as a sleeping slot, otherwise it takes the best action as the slot action with the position in TP sub-period. Note that the position takes diferent values as follow:

- ☞ The value 0 means the used channel is 2MHz, i.e., one packet will be used during the overall of the slot.
- ☞ The value 1 or 2 means the used channel is 5MHz with one packet in the position 1 or 2.
- ☞ The value 3 means the used channel is 5MHz with tow packets.

Furthermore, the node in sleeping slots remains to hear the RTSs on the dedicated channel in the SP sub-slot to discover the coming of the new nodes which can learn only in the sleeping slots of the neighboring nodes. If the node founds new nodes, it continues the performance of the learning phase with those coming nodes, otherwise, it turns off its radio until the next slot.

Algorithm 3.3: HARL CA

1. Initialize:
- Q Table of Q values (Nbr of Slots Nbr of Actions);
- Channel-Power Table (Nbr of Slots Nbr of channels);
- Succ-Action List of successful actions (Nbr of Slots);
- Position Table of successful actions positions (Nbr of Slots);
- P_{Sleep};

LEARNING

2. **while** not end of learning frames
3. **for** each frame
4. **for** each time slot (S_i)except the Beacon slot
5. **if** Succ-Action[S_i]$\neq \varnothing$ **then**
6. Sleep until Position[S_i] in TP sub-period then perform Succ-Action[S_i] using the Equation 4.14
7. **else**
8. do LEARNING scheme depicted in Figure (4.24)
9. **end if**
10. observe Reward
11. update Q Table using the Equation 4.9
12. update Succ-Action List
13. update Position Table
14. update Data queue
15. **end for**
16. **end for**
17. **end wile**

COMMUNICATION

18. **for** each time slot (S_i) in the frame
19. select a* with Q(a*) > Q(a) for all actions
20. **if** Q(a*) <= P_{Sleep} **then**
21. Action[S_i] <– sleep
22. **else**
23. Action[S_i] <– a*
24. **end if**
25. **end for**

4.4.6 Performance Evaluation

Herein, we present the simulation results of our algorithm HARL CA in comparison with both RL MMAC ([62]) and M-LDCS ([90]). The different results are averaged over 10 simulations. Each simulation experiment is run for the same topology over the three protocols. Therefore, the sink is placed in the middle of the area and the nodes are fixed in their random places to execute each of the three protocols in a sequential manner. The leaf nodes generate data messages every 5 seconds. Each frame is divided into 10 slots, and the length of the slot is changed between the three protocols in such a way that in RL MMAC, the token is

40 ms, while in M-LDCS and HARL CA, it is 48 ms. The aim is to use the same packet length considered by 64 bytes in 40ms and exploit the first 8 ms to use 4 sub-slots for the SP sub-period in both M-LDCS and HARL CA. The channels used are divided by half between 5MHz and 2MHz bandwidth. Referring to [62] for the length of the learning period that depends on the maximum path to the sink node, we have used 15 frames which is more than enough. In addition, we have used a tree-based topology for the protocol M-LDCS for two reasons, the first is that M-LDCS focuses on the links that belong to the paths to the sink without mentioning how the node constructs these paths, while the second is to use the same topology for the three protocols to make the comparison more meaningful.

For this comparison, two steps are followed: the comparison of the Learning period then the comparison of the network lifetime during a long period considered by 240 minutes. For the first comparison, four metrics are token over a variation of channels: delivery ratio and packet delivery rate at the sink, the overall energy consumption rate, the energy consumption rate of overhead and the collisions. Note that, we have taken the different measurement in the protocol M-LDCS at the end of the same frame as that of HARL CA's last frame of the learning period since M-LDCS don't use a learning period.

Figure 4.26 shows that HARL CA improves the delivery ratio by an average of 38.74% between the different channels compared to M-LDCS and by 324.25% compared to RL MMAC. This result can be explained by the impact of the overhead and collisions in both M-LDCS and RL MMAC as it is shown in figures 4.29 and 4.28, which effect hardly the total energy consumption as it is shown in figure 4.27. Hence, RL MMAC consumes the

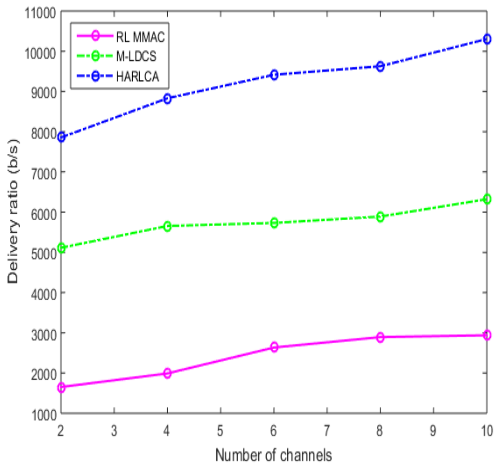


FIGURE 4.26: Comparison of delivery ratio per number of channels.

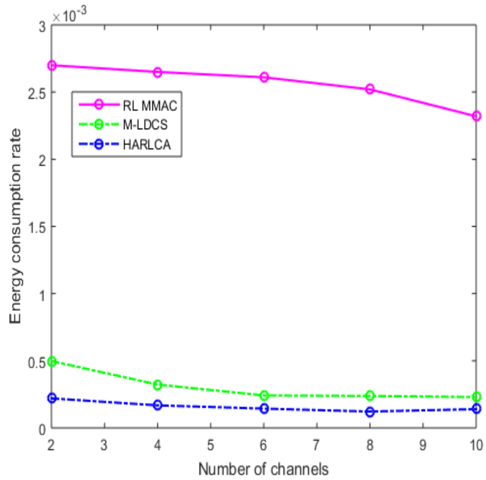


FIGURE 4.27: Comparison of total energy consumption rate per number of channels.

highest level of both overhead and collisions, in such a way that the overhead consumption is measured by 8.4 times the average consumed by HARL CA due to the absence of the RTS/CTS mechanism since RL MMAC exchanges the data packets directly, while the collisions number is measured by 7.2 times the average realized by HARL CA due to the absence of a collision-avoidance mechanism in RL MMAC. However, M-LDCS consumes a level of overhead measured by 1.23 and collisions measured by 2.1 times the average of HARL CA due to the absence of an indirect collision avoidance mechanism in M-LDCS. Furthermore, the bandwidth-based schedule is not used in RL MMAC since it sends only one packet at the slot even with 5MHz channel, and it is only used for the same link in M-LDCS if the node

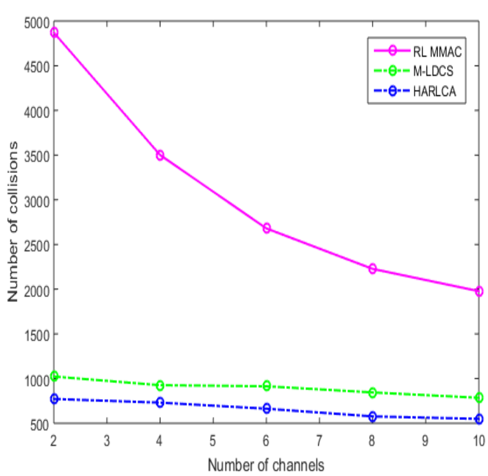


FIGURE 4.28: Comparison of number of collisions per number of channels.

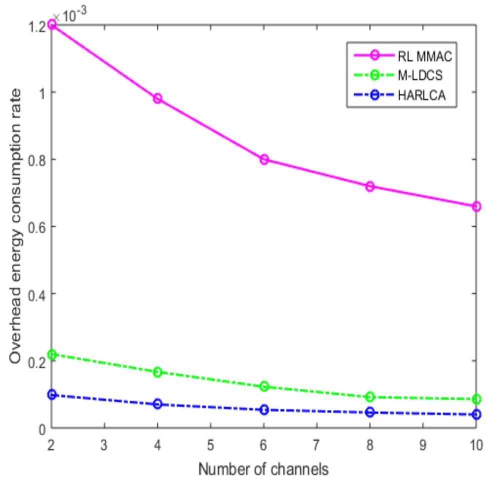


FIGURE 4.29: Comparison of overhead energy consumption rate per number of channels.

has more then one packet, while it is well used in HARL CA in both directions, for the same link if the node has more then one packet, and for two different links if not.

Consequently, the packets delivery rate is improved by HARL CA compared to RL MMAC and M-LDCS as it is shown in Figure 4.30. Therefore, HARL CA can reach the overall delivered packets at the sink over the different number of the used channels due to the reduced overhead and collisions on one side, and the absence of the duplicated packets, as it is shown in Figure 4.31, on the other side.

For the second comparison, the metrics token over a variation of channels in each protocol are: the moment of the first dead node, the rate of the dead nodes and the packet delivery

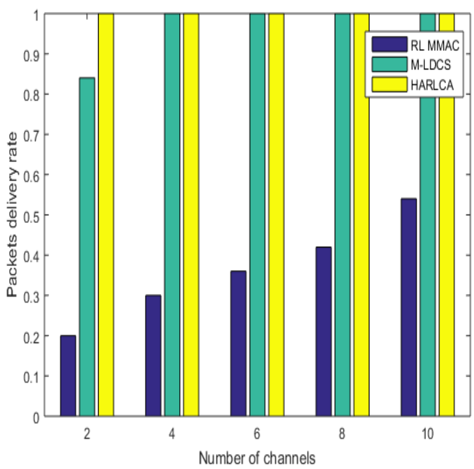


FIGURE 4.30: Comparison of packets delivery rate per number of channels.

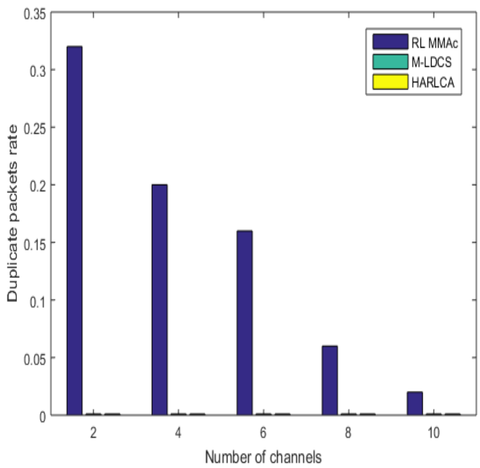


FIGURE 4.31: Comparison of duplicated packets at the sink per number of channels in the learning period.

rate at the sink during the time of simulation that is 240 minutes as it is mentioned above. Hence, HARAL CA shows good results in terms of improvement of the network lifetime and packets delivery compared to both M-LDCS and RL MMAC as it is presented in figures 4.32, 4.33 and 4.34 respectively. These results can be explained by the effect of the continuous overhead and indirect collisions in the increasing of the energy consumption in M-LDCS during the overall of its network lifetime, which is not the case in HARL CA that avoids gradually the collided nodes in the learning period and pushes the overhearing nodes to sleep in the communication period (post-learning period). In addition, the effect of the high-energy consumption in the learning period and the duplicated packets in RL MMAC-compared to HARL CA. However, it is observed that RL MMAC can extend the network lifetime compared to both M-LDCS and HARL CA in the cases of using less than 6 channels due to the very reduced number of packets delivered in the communication period presented in figure 4.34.

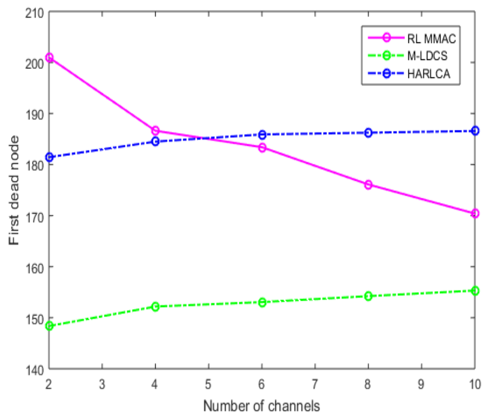


FIGURE 4.32: Comparison of first dead node in 240 minutes per number of channels.

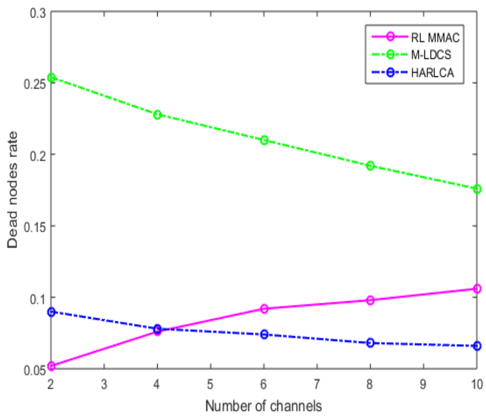


FIGURE 4.33: Comparison of dead nodes in 240 minutes per number of channels.

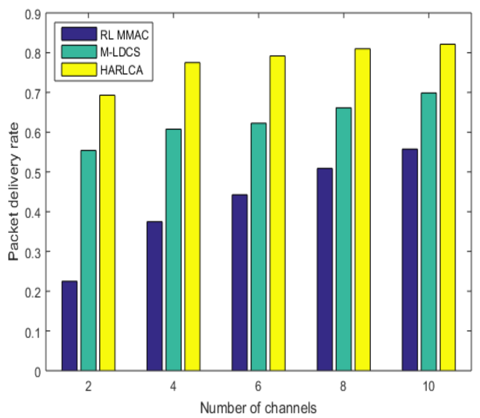


FIGURE 4.34: Comparison of packets delivery rate per number of channels.

4.5 Schedule-based Cooperative Multi-agent Reinforcement Learning for Multi-channel communication in Wireless Sensor Networks

As explained before, the most cooperative RL works are focused on CRSNs since the use of the cooperative RL approach requires a number of cooperative iterations to obtain the best solution which in turn creates a communication overhead and time-wasting. To overcome the problem, the Self-schedule based Cooperative multi-agent Reinforcement Learning for Channel Assignment (SCRL CA) approach is proposed to improve the network lifetime and performance. The proposal addresses both regular traffic scheduling and assignment of the available orthogonal channels in an energy-efficient way through the accelerating of the RL iterations and the reduction of both communication overhead and collisions on one side, and the balance of energy consumption on the other side.

4.5.1 Self-schedule approach

In WSNs, the use of the routing metrics focuses generally on the receiver selection process. Hence, the nodes have to collect these factors periodically from their neighboring nodes then they use it to select the optimal receiver node which results in a high cost in terms of communication overhead and thus energy-wasting. In contrast, the authors of [92] have used these factors in balance by a self-scheduling manner for data routing in WSNs based single-channel to reduce the communication overhead caused by the exchanged information between the neighboring nodes. For this purpose, each node that has to send a data message starts by sending an RTS (Request To Send) control message to its neighboring nodes, and then the receiver nodes must wait a time measured by (4.15) before the response by a CTS (Clear To Send) message. In this way, the node which has waited the smallest time responds first, and the other waiting neighboring nodes learn from this response that the demand is satisfied without sending its CTSs control messages.

$$t = \frac{(\frac{1}{d})\gamma}{Er} \tag{4.15}$$

Where, d is the distance of the node from the sink, E is its self-residual energy, γ represents the number of neighboring nodes and r is the link reliability.

Although the results demonstrate the efficiency of these scheme in term of communication overhead reduction, it has some limitation that can be resumed as follow: first; it doesn't perform the collision-free aspect for the sender nodes since it focuses only on the receiver ones without taking in account that the sender nodes can send RTSs simultaneously. Second; it uses a free waiting time which can be very long even for the smallest one, which in turn increases the wasting time and energy consumption.

4.5.2 System model

We define a WSN as a set of N nodes denoted us N=$\{1, ..., N\}$, randomly dispersed throughout an area. Each node uses a predefined set of orthogonal channels K=$\{f_1, ..., f_k\}$ in a range of communication R, for example the non-overlapping channels proposed by IEEE802.15.4 standard (16 channels in 2.4 GH band and 10 channels in 915 MH band) [3]. All the nodes can communicate with the sink node in a multi-hop fashion through the communication with their M neighbors located within their communication range.

After the initialization phase, all the nodes are synchronized with the global time clock of the sink. Hence, the time is divided into consecutive frame periods which are in turn, divided into a fixed number of slots. Tow types of slots have been considered, Beacon slots for exchanging control information (Beacon messages, new request messages, etc) on the same dedicated channel, and Data slots for data transmission. Furthermore, each node keeps the lowest number D of hops from the sink node as its depth and a list of its parent nodes selected from the list of the neighboring nodes as well as the maximum number of parent lists of the neighboring nodes. The parent nodes are the neighboring nodes that have a smaller D than that of the node.

4.5.3 Problem formulation

To perform both channel and time (channel-time) schedules for each time slot, a cooperative multi-agent reinforcement learning approach is applied. Hence, the channel-time schedule problem can be formulated as POMDP that can be expressed mathematically as POMDP=$\{n, S, A, T, R_{1...N}, \gamma\}$, where;

- ☞ n is the number of agents,
- ☞ S is the set of states of all agents,
- ☞ A is the joint action space composed of local actions for all the agents,
- ☞ T:$SAS \longrightarrow [0,1]$, is the state transition function,
- ☞ R_i:$SAS \longrightarrow \mathbf{R}$ is the reward function of agent i.
- ☞ $\gamma \in [0,1)$, is the discount factor, which represents the relative importance of future and present rewards.

Hence, we define the number of agents n by the number of all the nodes N, so each node is considered as an agent. The set of actions for each node is defined as $A_x = \{f_1, ..., Sf_k, Rf_1, ..., Rf_k\}$, where $A_x$is the set of actions for the node x, Sf_i and Rf_i mean send on channel f_i and receive on channel f_i respectively. Therefore, A_1=A_2=...=A_N. In order to avoid complex calculation that is not desired for the WSN, we use the stateless variant of the Q-learning method that has been demonstrated its efficiency for distributed learning problems [93]. Hence, the state transition function is chosen based on a trade-off between the strategy *"win stay, lost shift"* used in RL MMAC [62] and the stateless Q-learning method. The *"win stay, lost shift"* strategy plays an important role in the learning acceleration process by the fact that if the agent fails in performing action after some successes, it does not lose time in the gradual return. Therefore, T is fined by (4.16) :

$$Q_{x,t+1}(a_t) = (1-\alpha)Q_{x,t}(a_t) + \alpha r_x(a_t) \tag{4.16}$$

where, $Q_{x,t+1}(at)$ means the Q value update at time slot t + 1 (the slot in the current frame) by agent x, after executing at time slot t (the same slot in the last frame), the action a_t. $r_x(a_t)$ means the immediate reward calculated by (4.17) after executing the action a_t at time slot t by agent x. Hence, the reward takes the value (+1) if either the receiver node receives a data message or the sender node receives an Acknowledgement (Ack) message on the chosen channel.

$$r_{x,t}(a_t) = \begin{cases} +1 & \text{if it is a successful communication} \\ -1 & \text{otherwise} \end{cases} \tag{4.17}$$

α is the learning rate parameter which can be set to a value in [0,1]. In order to formulate the *"win stay, lost shift"* strategy that it is mentioned in [28] without formulation, we use the *"win or learn fast"* for variable learning rate method proposed in [**26**], where a small value of

α is used for successful actions and a higher value is used for unsuccessful ones. Therefore, α is done by (4.18) as follow:

$$\alpha = \begin{cases} 0.01 & \text{if r} > 0 \\ 0.1 & \text{otherwise} \end{cases} \tag{4.18}$$

The best joint action to be selected in the next time slot is done by(4.19);

$$a^* \in Max_{a \in A_{av}} [\sum_{x=1}^{NG} Q_{x,t+1}(a)] \tag{4.19}$$

Where, a^* is the best joint action to be selected in next time slot by looking at the maximum of Q values sums after performing actions among the available action set A_{av} by the Neighboring nodes Group (NG) since the cooperation between the agents is reduced to that between the neighboring nodes. To perform (4.19), each node must collect messages from its neighboring nodes according to the cooperation model explained in the next sub-section.

The discount factor γ is not taken since we use the stateless method.

4.5.4 Cooperation model and algorithms

In order to perform the cooperation process in an energy-efficient way, we have used prior coordination based on *"social conventions"* strategy used in [94] to coordinate between Unmanned Aerial Vehicles (UAVs) for field coverage. For that, the UAVs coordinate the action to be selected in advance by the fact that each UAV must not choose the action chosen by the others. To perform this coordination method, specific ranking order is assigned to each UAV. The one with the highest order selects action first and lets the others know its action. The other UAVs then can match their actions with respect to the prior selected ones.

To adapt this method to the WSN in an energy-efficient way, we have used a Self-scheduling mechanism between the neighboring nodes. Hence, before performing actions in the selected channel, the neighboring nodes select actions sequentially by informing one another. To do that, the neighboring nodes must use the same dedicated channel. Therefore, each data time slot in the frame period is divided into two periods: a broadcast period T_B in which all nodes turn back to the dedicated channel, and unicast period T_U in which the communicating nodes use the selected channel for sending data messages. Note that, the dedicated channel can be used in T_U period. The T_B period is divided into two equal and consecutive sub-periods: T_{RTS} and T_{CTS} sub-periods as it is shown in Figure 4.35. T_{RTS} sub-period is used for sending Request To Send (RTS) control messages, while T_{CTS} sub-period is used to send the Clear To Send (CTS) ones by the parent nodes. The sending of control messages is performed in a self-scheduling manner to reduce the communication overhead, balance the remaining energy and avoid collisions. Therefore, in the T_{RTS} sub-period, the

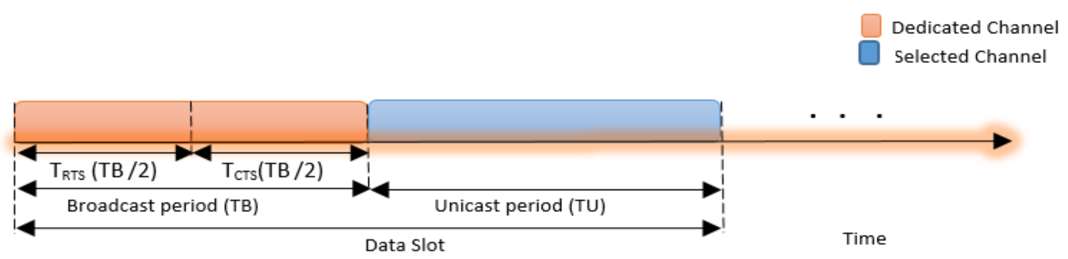


FIGURE 4.35: SCRL CA time slot structure

sender nodes schedule the sending of their RTSs. However, in T_{CTS} sub-period, the receiver nodes schedule and send the CTSs as follow:

Based on the positive number of its parent nodes (Parents Degree (PD>0)), the maximum of parents degree (Max Parents Degree MaxPD) between the neighboring nodes, the data queue size (λ), and the residual energy E_R, each sender node selects a random waiting time (RT_{WS}) bounded by T_{WS} in T_{RTS} sub-period, before transmitting RTS control message. Note that, the node that has no parent, except the sink, can neither send nor receive since it is excluded. The T_{WS} is calculated by (4.20).

$$T_{WS} = T_{RTS}^{(P_{WS}/(S_N+1))} \tag{4.20}$$

Where;

$$P_{WS} = (\frac{1}{\lambda})(\frac{PD}{MaxPD})(\frac{E_R}{E_{Init}}) \tag{4.21}$$

S_N is the number of successful communications started by a value 0, and E_{Init} is the initial energy of the node.

$$T_{RTS} = T_{CTS} = \frac{T_B}{2} \tag{4.22}$$

Hence, by (4.20) we can ensure a collision-free self-scheduling between the sender nodes within the first half of T_B (T_{RTS}) sub-period. Thus, (4.21) gives priority for the node with less residual energy and parents degree in one hand, and with more data messages in its queue in the other hand, to wait less time in order to preserve energy and prefer the sender with more data messages and fewer parents degree to send first, since nodes with more parents degree have more chance to relay its data messages. In addition, the T_{WS} is decreased gradually based on the number of successful communications S_N to give priority to the sender nodes that have succeeded to remain winners. Furthermore, the sender nodes choose their actions by performing the formula (4.19) in a sequential manner. Therefore, the first sender node chooses its action, then it broadcast its choice, the other sender nodes, as well as the other neighboring nodes, learn from this choice by adding the sent Q value to their Q tables. Then the following sender nodes will be forced to choose other actions, among the available actions, sequentially. The different steps in the T_B period are presented in the flow chart of Figure (4.36).

On the other side, the parent nodes receive in T_{RTS} sub-period, the RTS control messages sequentially depending on the priority of the sender nodes. The parent nodes then select a random waiting time (RT_{WR}) between T_{RTS} and T_{WR} times in T_{CTS} sub-period, before the sending of their responses (CTSs). The T_{WR} is calculated by (4.23) as follow:

$$T_{WR} = T_{RTS} + T_{CTS}^{(P_{WR}/(S_N+1))} \tag{4.23}$$

Where;

$$P_{WR} = (\frac{(MaxPD+1) - PD}{MaxPD+1})(1 - \frac{E_r}{2E_{Init}}) \tag{4.24}$$

Hence, by (4.23) we ensure a self-schedule with collision-free between the receiver nodes based on residual energy and parents degree. The aim is to balance the remaining energy between the neighboring nodes on one hand and select the best path to optimize delivery delay on the other hand. Also, the T_{WR} is decreased gradually based on the number of successful communications S_N to give priority to the sender nodes that have succeeded to remain winners. Thus, (4.24) prefers the parent with more energy and parent degree to respond first. Therefore, the other parent nodes, as well as the other neighboring nodes, learn from the previous responses and respond to the other RTS control messages sequentially. To do this, each receiver node updates its RTS queue after the waiting of RT_{WR} during which it listened to

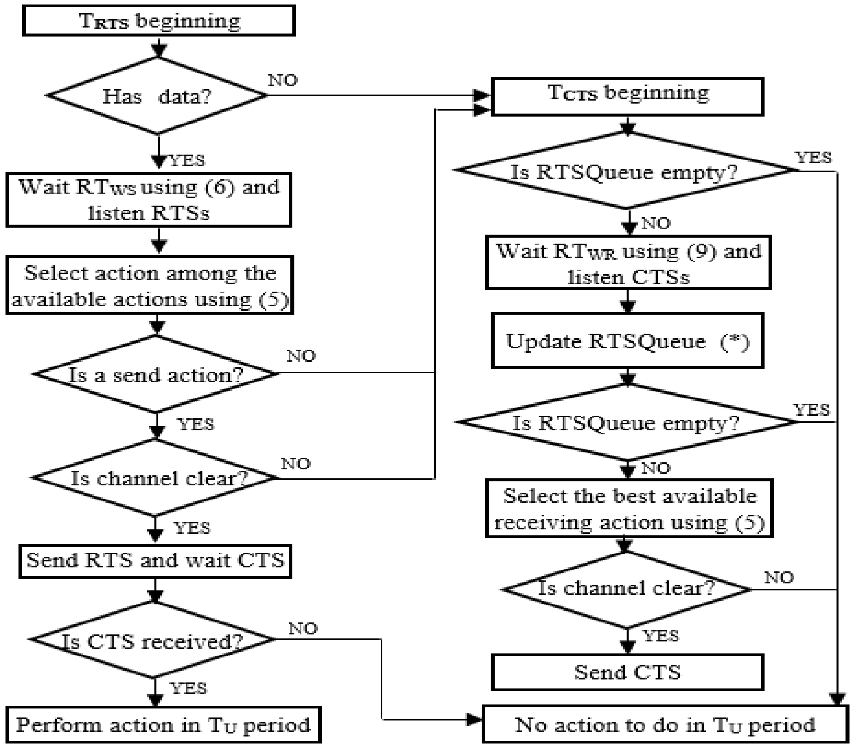


FIGURE 4.36: Cooperation model and scheduling process in T_B period

the other CTSs (the case is mentioned by (*) in Figure (4.36)). The updating process focuses on delating of the RTSs that belong to the following cases:

- The RTSs for which the node is a parent, and are not answered by the previous parent nodes, but they select a channel that is used by the previous responses.
- The RTSs for which the node is a parent, but they are answered by the previous parent nodes,
- The RTSs for which the node is not a parent,

Then, the parent selects the RTS which corresponds to the best available response action. In this way, we ensure collision-free communication that avoids both collision types: direct and indirect collisions. Furthermore, by (4.22) we ensure a self-schedule with collision-free between the receiver and the sender nodes.

As mentioned before, we have used two algorithms: Dynamic Self-schedule based Cooperative multi-agent Reinforcement Learning for Channel Assignment (DSCRL CA) that is presented in Algorithm 3.4, and Static Self-schedule based Cooperative multi-agent Reinforcement Learning for Channel Assignment (SSCRL CA) that is presented in Algorithm 3.5.

In SSCRL CA, we have used sleeping probability (P_{SLEEP}) for each slot that is initialized by the value 0 and increments if there is no action to do in T_U period (line 13). After the learning period, each node looks for each slot at whether its sleeping probability (P_{SLEEP}) is greater than a threshold value Δ. If it is the case, it takes the sleeping action during this slot as the slot action (lines 19, 20), otherwise, it takes the best-performed action in the learning period as the slot action (lines 22, 23). Furthermore, the data message size is increased to satisfy the slot size in the absence of T_B period after the learning phase since the data message size can achieve 128 bytes in IEEE802.15.4 for example [3]. In addition, The new nodes

Algorithm 3.4: DSCRL CA

1. Initialize Q table;
2. **while** not end
3. **for** each frame
4. **for** each data time slot
5. do LEARNING scheme depicted in Figure (4.36)
6. **if** ther is an action to do in T_U period **then**
7. performe action in T_U period
8. observe Reward
9. update Q-table using (4.16)
10. update Data queue
11. **else**
12. Sleep in T_U period
13. **end if**
14. **end for**
15. **end for**
16. **end wile**

Algorithm 3.5: SSCRL CA

1. Initialize Q table,Δ,P_{Sleep}
LEARNING
2. **while** not end of learning frames
3. **for** each frame
4. **for** each data time slot
5. do LEARNING scheme depicted in Figure (4.36)
6. **if** ther is an action to do in T_U period **then**
7. performe action in T_U period
8. observe Reward
9. update Q-table using (4.16)
10. update Data queue
11. **else**
12. Incrimente P_{Sleep}
13. Sleep in T_U period
14. **end if**
15. **end for**
16. **end for**
17. **end wile**
SELECTION
18. **for** each data time slot in the frame
19. **if** $P_{Sleep} > \Delta$ **then**
20. Slot-action <– sleep
21. **else**
22. select a* with Q(a*) > Q(a) for all actions
23. Slot-action <– a*
24. **end if**
25. **end for**

can discover their parents through the beacon messages that are transmitted periodically in Beacon slots, and start the learning process with the sleeping slots of their parents. In order to accelerate the learning process of the new nodes, each parent keeps a blacklist of channels (Bl_{Chs}) for each slot, created during the learning phase, which includes the different used channels by the neighboring nodes in the concerned slot. During the negotiation between the new node and its parents, each parent sends the list of sleeping data time slots as well as the Bl_{Chs} for each data time slot to be excluded from the available actions of the concerned slot.

4.5.5 Performance Evaluation

In this section, we present the simulation results of our protocols in comparison with both RL MMAC [62] which uses the same distributed tree topology in static mode and CMAA [78] that is characterized by its dynamicity.

We run simulation experiments with 100 nodes placed randomly in an area of 500m * 500m. The sink is placed in the middle and the transmission range is fixed at 50 meters. The leaf nodes generate data messages every 10 seconds. The different results are averaged over 10 simulations.

Two scenarios are token, the comparison of learning period between RL MMAC and SS-CRL CA, then the comparison of network lifetime and performance for a long time measured by 300 minutes between the three protocols CMAA, SSCRL CA and DSCRL CA.

For the first comparison shown in figures 4.37-4.40, six metrics are token over the variation of the number of channels:

- End to end packets delivery ratio,
- Total energy consumed ratio,
- Total energy consumed ratio of overhead,
- Number of colliding packets,
- End to end latency,
- End to end hops rate.

To evaluate the impact of the number of time slots per frame, we have used the parameter α which is defined as the ratio of the number of time slots per frame to the number of periodic generated data messages by the leaf nodes. The length of the slot is changed from RL MMAC to SSCRL CA in such a way that in RL MMAC, the token is 40 ms, while in SSCRL CA, it is 60 ms. The aim is to use the same packet length considered by 64 bytes in 40ms and exploit the first 20 ms in SSCRL CA for the T_B period. As it is mentioned in [62]for the length of the learning period that depends on the maximum path to the sink node, we have used 15 frames which is more than enough.

Figures 4.37-4.40 show that SSCRL CA can reach the total end to end packets delivery for all the 10 experiences in α=1.5, while RL MMAC succeded for more than 2 channels in all the experiences and only for two experiences in 2 channels (figure 4.37). Furthermore, SSCRL CA can optimize the delivery packets by more than 50% in the other values of α. The energy consumed in the learning phase is greatly reduced by SSCRL CA in different situations (figure 4.38). This reduction can achieve 80% of the energy rate consumed by RL MMAC. The reason can be explained by the fact of the high communication overhead generated by RL MMAC since it broadcasts the data packets instead of the control packets used in SSCRL CA (figure 4.39), as well as the significant reducion of the collisions performed by SSCRL CA that can be reached 78% due to the self-scheduling mechanism that is used (figure 4.40).

In figures 4.41 and 4.42, we have taken the values of α for which there is a total end to end packets delivery to investigate the latency and the average of the hops taken by all the data messages from the source nodes to the sink in the learning phase of the two protocols. Note that in figure 4.41, we have taken only α=1.5 for RL MMAC since it performs the total reach at this value. Therefore, the latency of SSCRL CA is better than that of RL MMAC if the frame periods are identical since the frame period of SSCRL CA in α=1 is identical to that of RL MMAC in α=1.5. However, if the α values are the same, RL MMAC performs well in latency than SSCRL CA since the time slot in the frame period of RL MMAC is lower than that of SSCRL CA by 33 %. Figure 4.42 shows the path length rate taken by all the data messages to the sink node by the two protocols in the two values of α (1,1.5). Hence, SSCRL CA continually reduces the length of the path according to the number of used channels by an increasing rate that can reach 36.48 % on 8 channels at α=1.5. This can be explained by the effect of the default channel mechanism used by RL MMAC since the tracking of the default channels can result in very long paths, while SSCRL CA gives priority to build the smallest

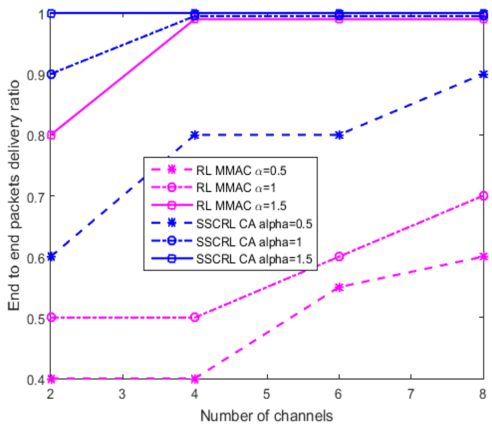


FIGURE 4.37: Comparison of end to end packet delivery ratio.

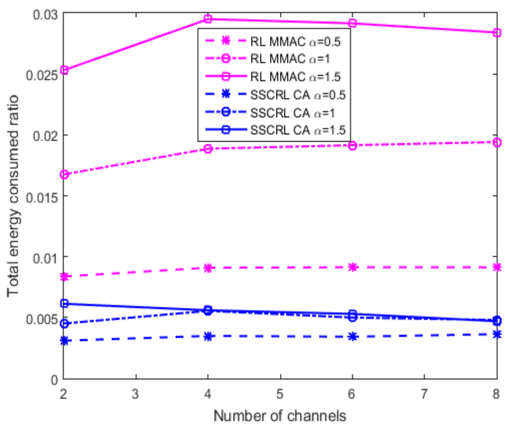


FIGURE 4.38: Comparison of total energy consumption ratio.

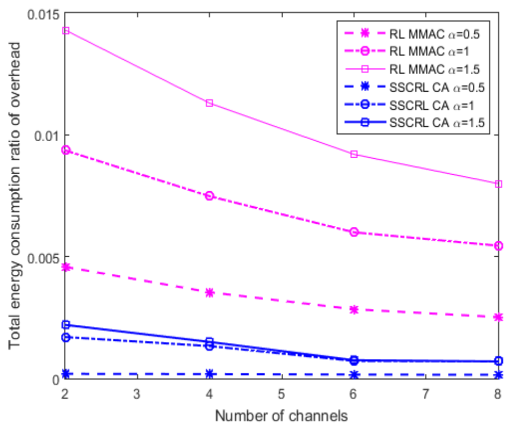


FIGURE 4.39: Comparison of Total overhead energy consumption ratio.

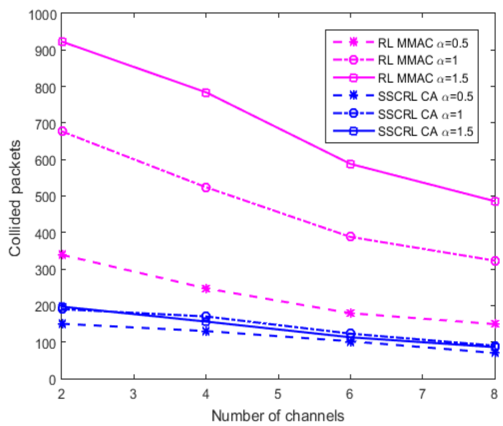


FIGURE 4.40: Comparison of collided packets per number of channels in learning period.

paths to the sink as much as possible based on the dynamic channel selection mechanism that is used.

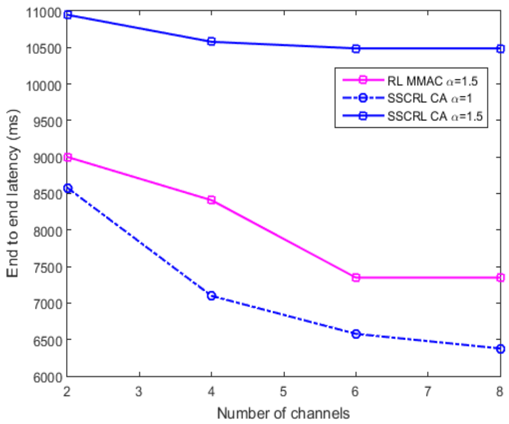


FIGURE 4.41: Comparison of total end to end delivery.

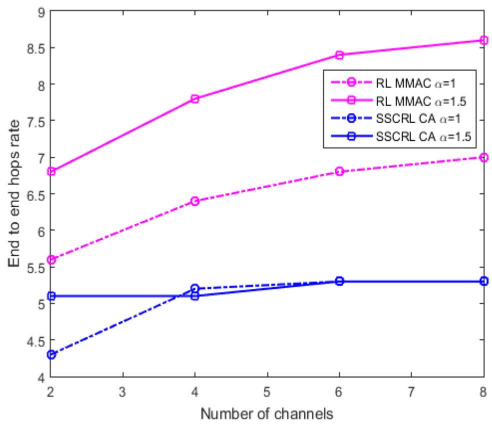


FIGURE 4.42: Comparison of hops rate per number of channels in Learning period.

For the second comparison shown in Figures 4.43-4.46, we set the value of α at 1.5 to investigate the network energy consumption and delivery using the three protocols CMAA, SSCRL CA and DSCRL CA for a long time considered by 300 minutes. Therefore, four metrics are token over the variation of the number of channels:

- Dead nodes,
- The first dead node,
- Delivery ratio in bit per second,
- Total energy consumption ratio of overhead.

The figures4.43 and 4.44 show the dead nodes in 300 minutes of performance using the three protocols. Hence, DSCRL CA is the best one at using of few channels (2,4) in such a way that it does not suffer from any dead node using 2 channels and suffers from the lowest dead nodes ratio using 4 channels. However, unlike the other two protocols, the number of dead nodes in DSCRL CA increases relatively with the increase in the number of the used channels, which makes SSCRL CA the best one at using more than 4 channels. This is due to the increasing level of the communication overhead in T_B period relatively with the increase in the number of the used channels, which leads to an increase of supplementary energy consumption that is shown in figure 4.46, and which has overcome the existed energy balance mechanism in 6 and 8 channels use cases. The delivery ratio in figure 4.45 is optimized especially in SSCRL CA in an increasing manner according to the number of used channels. Hence, The optimization is increased from 6.07% using 2 channels to 60.42% compared to CMAA delivery ratio at 8 channels. The reason returns to the avoidance of the communication overhead as well as the increasing of the packet length after the learning period, which is performed in a successful manner compared to the CMAA as it is explained above. However, the delivery ratio increases slowly in DSCRL CA. It goes from the rate of 10.18 % at 2 channels compared to CMAA to the rate of 15.02 % at 8 channels, and this is due to the costs of supplementary overhead as it is shown in figure 4.46 that's mention an improvement of DSCRL CA compared to CMAA due to the self-scheduling mechanism.

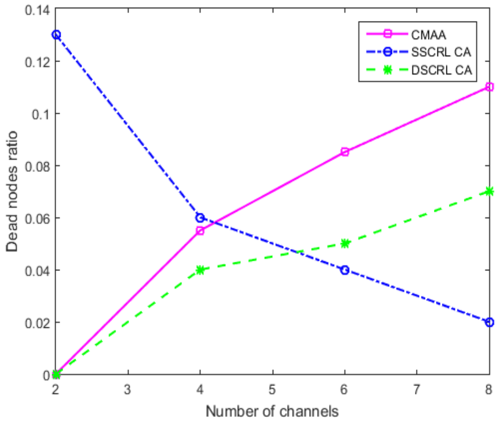


FIGURE 4.43: Comparison of dead nodes ratio.

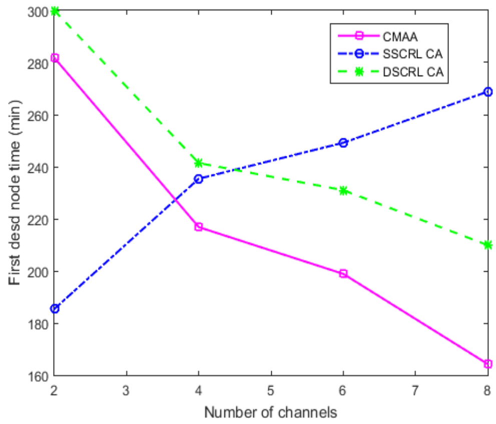


FIGURE 4.44: Comparison of first dead node per number of channels in 300 minutes of communication.

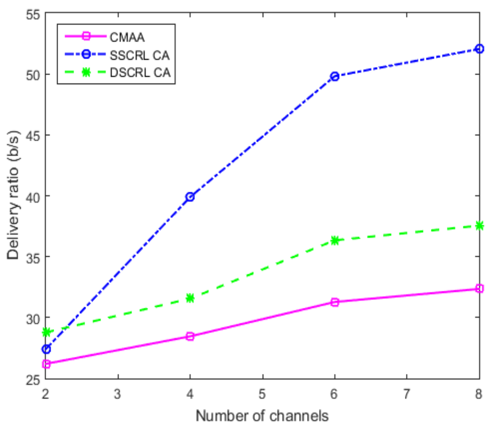


FIGURE 4.45: Comparison of delivery ratio in 300 minutes of communication.

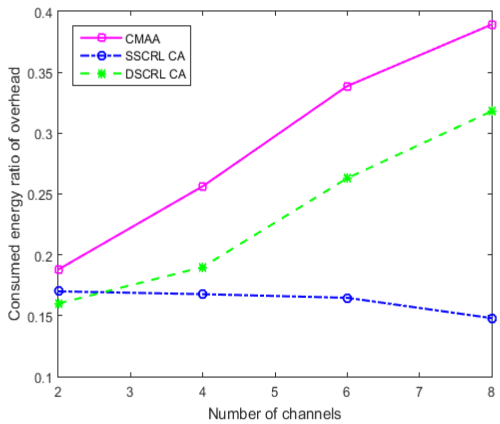


FIGURE 4.46: Comparison of total overhead energy consumption in 300 minutes of communication.

4.6 Conclusion

In this chapter, we have presented our contributions that are MCSP, ERL MMAC, HARL CA and CRLCA protocols. Our proposed methods tackle recent problems in new fields through using new techniques which can be summarized as: first, the proposition of multi-channel collision solution to the WPAN IEEE802.15.4 standard that doesn't specify the use of multi-channel communication in the original standard, second, the use of parent selection in two hops for multi-channel single RL, third, the new use of HARL approach as a modified version of RL approach for accelerating the learning process in an energy-efficient way taking in account the bandwidth aspect in the scheduling process, fourth and finally, the use cooperative RL model with the self-scheduling method. Hence, not only the throughput and delivery rate that are improved, but also the enhancement of the energy consumption through the minimization of collisions and communication overhead on one side, and the use of energy consumption balancing techniques on the other side. The balance of the consumed energy is performed in both cases, the centralized case between the sub-networks in MCSP, and the decentralized one between the neighboring nodes through the use of the self-scheduling method in HARL CA and CRLCA protocols. All these techniques have been successfully evaluated and show outperformed results in different cases through several experiments.

Chapter 5

General Conclusion And Future Works

5.1 General Conclusion

This thesis in intended to contribute to the resolution of multi-channel problems in wireless networks by overcoming the limitations that threaten the wireless communication because of its nature. The ultimate objective of multi-channel communication is to provide an alternative solution to Single-channel communication for some modern applications that need a considerable throughput and delivery rate due to the important volume of the transferred data. However, the use of such alternative solution needs some considerations to take into account in order to adapt it to the specificity of the chosen WN. This consideration becomes more important in WSNs known by their sensitive constraints due to its limited resources. Therefore, the requirement of frequent negotiation between the nodes to schedule the use of channels remains incur an important amount of communication overhead and collisions, which results in more energy consumption, and thus a reduction of both the communication quality and network lifetime, that are considered as the most important factors used in improvement of communication in WSNs.

The main goal of this thesis is to focus on overcoming the challenge of adapting the use of multi-channel solutions to the WSNs in order to improve the throughput and delivery rate, taking into account their constraints by the fact of ensuring the longest network lifetime possible. To do that, this thesis has started with a general introduction in the first chapter to emphasize the multi-channel communication in WNs from the different viewpoints and describe our motivations, then a comprehensive study of different axes strongly related to the thesis objective is made in the second chapter, also a detailed explanation of the different proposed protocols is formed the third chapter and finally, a general conclusion with future works constitutes the forth and final chapter.

In the second chapter, the study has tackled firstly an overview for the WNs field with the aim is to define the main aspects of such field Without digging into the details because it is a very wide field, then a special consideration is given to WSNs, as the area of work of this thesis, to bring out their issues and the opportunity of multiple channels offered in the different standards that are conceived for low powered WNs,

In the third chapter, the multi-channel communication in WSNs is illustrated in order to explain the different dependent problems and their proposed solution by classifying them into non-intelligent and intelligent solutions since the problem of multi-channel scheduling is proven as an NP-Hard problem, then thirdly, a particular interest is given to the RL approach as the state of the art, with a comparison made between the major proposed solutions in each RL type to underline the importance of our proposed protocols.

The fourth chapter has provided a detailed presentation of our proposed protocols that are: Multi-Channel Scheduling Protocol for Wireless Personal Area Networks IEEE802.15.4, which focuses on overcoming the collisions problem by a multi-channel scheduling scheme. The second protocol is Energy-efficient Reinforcement Learning Multi-channel MAC (ERL MMAC) for WSNs, which bases on the enhancement of the energy consumption in WSNs, by reducing collisions and balancing the energy consumption between the nodes in using single-agent RL. The third work represents a proposition of a new heuristically accelerated RL protocol named Heuristically Accelerated Reinforcement Learning approach for Channel Assignment (HARL CA) algorithm, with the aim is to reduce the number of learning iterations in an energy-efficient way taking into account the bandwidth as an important criterion in the scheduling process. The last work is the proposition of a new cooperative multi-agent RL approach for Channel Assignment (CRLCA) in WSNs, which improves the cooperative RL using an accelerated learning model, and overcomes the extra communication overhead problem of the cooperative RL by using a new method for self-scheduling and energy balancing. The proposed CRLCA approach is performed through two algorithms SCRLCA and DCRLCA for Static and Dynamic performance respectively. the proposed protocols and techniques have been successfully evaluated and show outperformed results in different cases through several experiments.

5.2 Future Works

As our work is interested by the study and investigation of the multi-channel communication field, this thesis will require additional efforts and years of training and work. Nonetheless, our short-term goals focus on improving and complementing the work initiated in this thesis, namely:

- ☞ Scaling our protocols to be applied in CRSNs by the fact of taking into consideration both licensed and unlicensed bands in communication unlike most of the proposed CR-based algorithms that focus only on the management of the licensed bands.
- ☞ Propose a new deep RL protocol for the WSNs based on self-scheduling with less computation possible to be adaptive to the nature of such networks.
- ☞ The use of our proposed protocols in real experiments to validate the obtained results and discover possible issues.
- ☞ The investigation of the proposed self-scheduling RL cooperation method in a full dynamic channel assignment environment with a hybrid communication policy rather than the received based one without the use of tree-based architecture.
- ☞ The investigation of the proposed techniques in other wireless network types such as wireless mesh networks.

Bibliography

[1] Mahamat Habib, Senoussi Hissein, Adrien Van D.B, and Thierry Val. "Survey on Multi- Channel Access Methods for Wireless LANs". In: *International Journal of Engineering Research and Technology* 3.12 (2014), pp. 673–680.

[2] IEEE802.11 Working Group The Working Group for WLAN Standards. "802.11-2020 - IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications". In: (Feb 2021). DOI: 10.1109/IEEESTD.2021.9363693.

[3] 802.15 WG Wireless Specialty Networks (WSN) Working Group. "IEEE 802.15.4-2020 - IEEE Standard for Low-Rate Wireless Networks". In: (July 2020). DOI: 10.1109/IEEESTD.2020.9144691.

[4] C/LM LAN/MAN Standards Committee. "IEEE 802.16s-2017 - IEEE Standard for Air Interface for Broadband Wireless Access Systems–Amendment 4: Fixed and Mobile Wireless Access in Channel Bandwidth up to 1.25 MHz". In: (Oct 2017).

[5] Pathak Parth H and Dutta Rudra. "A Survey of Network Design Problems and Joint Design Approaches in Wireless Mesh Networks". In: *IEEE Communications Surveys and Tutorials* 13.3 (2011), pp. 396–428. DOI: 10.1109/SURV.2011.060710.00062.

[6] Yao Junmei, Xu Jun, Che Yue Ling, Wu Kaishun, and Lou Wei. "Revisiting of Channel Access Mechanisms in Mobile Wireless Networks through Exploiting Physical Layer Technologies". In: *Wireless Communcations and Mobile Computing* 2018 (2018), pp. 1–16. DOI: 10.1155/2018/5967194.

[7] Amjad Ali, Wang Huiqiang, Lv Hongwu, and Xiaoming Chen. "A Survey of MAC Protocols Design Strategies and Techniques in Wireless Ad Hoc Networks". In: *Journal of Communications* 9.1 (2014), pp. 30–38. DOI: 10.12720/jcm.9.1.30-38.

[8] K.Y Neha and Priya. I Deva. "A Survey on Channel Coordination for Multi-Channel Wireless Networks". In: *International Journal of Engeneering Research and technology* 2.12 (2013).

[9] Nasipuri Pal A. "A Distributed Routing and Channel Selection for Multi-Channel Wireless Sensor Networks". In: *J. Sensor and Actuator Networks* 6.3 (2017), pp. 113–126.

[10] Maheshwari Ritesh, Gupta Himansho, and Das S.R. "Multichannel MAC Protocols for Wireless Networks". In: *2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*. Vol. 2. 28 sept 2006. DOI: 10.1109/SAHCN.2006.288495. Reston, VA, USA.

[11] Tian-Yun Shi et al. "Low-latency Data Gathering with Reliability Guaranteeing in Heterogeneous Wireless Sensor Networks". In: *International Journal of Automation and Computing* 17 (2020), pp. 439–452. DOI: DOI:10.1007/s11633-017-1074-y.

[12] KH Phung, B Lemmens, M Mihaylov, L Tran, and K Steenhaut. “Adaptive learning based scheduling in multichannel protocol for energy-efficient data-gathering wireless sensor networks”. In: *International journal of distributed sensor networks* 9.1 (2013), pp. 1–11.

[13] Mohamed Sahraoui and Azeddine Bilami. “Decentralized Intelligent Schemes for Channels Assignment In Multi-channel Wireless Sensor Networks:a survay and Futur works”. In: *First International Conference on Digitization and its Applications*. 2020. M’sila, Algeria.

[14] Phung Kieu-Ha, Lemmens Bart, and Mihaylov Mihail. “Adaptive Learning Based Scheduling in Multichannel Protocol for Energy-Efficient Data-Gathering Wireless Sensor Networks”. In: *International Journal of Distributed Sensor* 2013.2 (2013), pp. 1–11. DOI: 10.1155/2013/345821.

[15] Mohamed Sahraoui and Azeddine Bilami. “A new Reinforcement Learning based for Energy-efficient Multi-channel Data Gathering in Wireless Sensor Networks”. In: *2020 4th International Symposium on Informatics and its Applications (ISIA)*. 2020. DOI: 10.1109/ISIA51297.2020.9416546. M’sila, Algeria.

[16] Mohamed Sahraoui and Azeddine Bilami. “Multi-Channel Scheduling Protocol for Wireless Personal Netwoks IEEE802.15.4”. In: *International Journal of Soft Computing and Software Engineering* 3.5 (2013), pp. 1–13. DOI: 10.7321/jscse.v3.n5.1.

[17] Mohamed Sahraoui, Azeddine Bilami, and Abdelmalik Taleb-Ahmed. “Heuristically accelerated reinforcement learning for channel assignment in wireless sensor networks”. In: *International Journal of Sensor Networks* 37.3 (2021), pp. 159–170.

[18] Mohamed Sahraoui, Azeddine Bilami, and Abdelmalik Taleb-Ahmed. “Schedule-based Cooperative Multi-agent Reinforcement Learning for Multi-channel communication in Wireless Sensor Networks”. In: *Wireless Personal Communications* (2021). DOI: 10.1007/s11277-021-09094-8.

[19] Wireless Networking Authors. *Wireless Networking in the Developing World*. Ed. by Jane S Butler. 3rd edition. 2013. 524 pp. ISBN: 978-1484039359. URL: https://www.amazon.com/Wireless-Networking-Developing-World-Version/dp/1484039351.

[20] Kaveh Pahlavan and Allen H. Levesque. *Wireless Information Networks*. Ed. by John Wiley and Sons. 2nd edition. 2005. 768 pp. ISBN: 978-0-471-72542-8. URL: https://www.wiley.com/en-us/Wireless+Information+Networks%2C+2nd+Edition-p-9780471725428.

[21] Ravi Teja. *Wireless Communication: Introduction, Types and Applications*. Ed. by electronicshub. URL: https://www.electronicshub.org/wireless-communication-introduction-types-applications/.

[22] tutorialspoint. *Wireless Communication*. Ed. by tutorialspoint. 2021. URL: https://www.tutorialspoint.com/wireless_communication/.

[23] Kata Kiatmanaroj. “Allocation de fréquence dans les systèmes de communication par satellites de type SDMA”. Doctorat. Toulouse University.

[24] I.F Akyildiz, W Su, and E Sankarasubramaniam Yand Cayirci. “Wireless sensor networks: a survey”. In: *Computer Networks* 38.4 (2002), 393–422. DOI: S1389-1286(01)00302-4.

[25] Ram Prasadh Narayanan, Thazath Veedu Sarath, and Vellora Veetil Vineeth. "Survey on Motes Used in Wireless Sensor Networks: Performance and Parametric Analysis". In: *Wireless Sensor Network* 8.4 (2016), pp. 51–60. DOI: 10.4236/wsn.2016.84005.

[26] Chris K, Joshua D, and Sam B. "Energy Options for Wireless Sensor Nodes". In: *Journal of Sensors* 2008.12 (2008), pp. 8037–8066. DOI: 10.3390/s8128037.

[27] Giuseppe Anastasi, Marco Conti, Mario Di Francesco, and Andrea Passarella. "Energy conservation in wireless sensor networks: A survey". In: *Ad Hoc Networks* 7.3 (2009), pp. 537–568. DOI: 10.3390/s8128037.

[28] IEEE 802.15 WPAN Task Group and IEEE Std 802.15.3TM. "Wireless Medium Acces Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPAN)". In: (2006). URL: http:profsite.um.ac.ir/hyagh^mae/ACN/WSNMAC1.pd.

[29] Eli Dekel. *Low-power Internet connectivity over Wi-Fi*. Ed. by Texas Instruments. 2019. URL: https://www.ti.com/lit/wp/swry019a/swry019a.pdf.

[30] Dola Saha. *Modern Wireless Networks Communication for Internet of Things*. Ed. by University AtAlbani state university of New York. 2019. URL: https://www.albany.edu/faculty/dsaha/teach/2019Spring_CEN574/slides/07_IoT.pdf.

[31] A S Asif Ahmad, Bharath Keshavamurthy, K S Abhay Narasimha, N Mahesh, and M N Suma. "Communication system design for white-fi (802.11af)". In: *2016 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. Nov. 2016. DOI: 10.1109/ANTS.2016.7947805. Bangalore, Indiaa.

[32] Evgeny Khorov, Andrey Lyakhov, Alexander Krotov, and Andrey Guschin. "A Survey on IEEE 802.11ah: An Enabling Networking Technology for Smart Cities". In: *Computer Communications* 58.3 (2015), 53–69. DOI: 10.1016/j.comcom.2014.08.008.

[33] Dmitry Bankov, Evgeny Khorov, Andrey Lyakhov, and Ekaterina Stepanova. "IEEE 802.11ba Extremely Low Power Wi-Fi for Massive Internet of Things - Challenges, Open Issues, Performance Evaluation". In: *arXiv:1909.00594v1* (2019). DOI: 10.1109/BlackSeaCom.2019.8812785.

[34] S Ahmad Salehi, M. A Razzaque, Inmaculada Tomeo-Reyes, and Nasir Hussain. "IEEE 802.15.6 standard in wireless body area networks from a healthcare point of view". In: *2016 22nd Asia-Pacific Conference on Communications (APCC)*. Aug. 2016. DOI: 10.1109/APCC.2016.7581523. Yogyakarta, Indonesia.

[35] Zigbee alliance. "ZigBee Specification". In: (2015). URL: https://zigbeealliance.org/wp-content/uploads/2019/11/docs-05-3474-21-0csg-zigbee-specification.pdf.

[36] Devan P. Arun Mozhi, Hussin Fawnizu Azmadi, Ibrahim Rosdiazli, Bingi Kishore, and Khanday Farooq Ahmad. "A Survey on the Application of WirelessHART for Industrial Process Monitoring and Control". In: *Journal of Sensors* 21.15 (2021), pp. 4951–4964. DOI: 10.3390/s21154951.

[37] David E. Culler and Jonathan Hui. *6LoWPAN Tutorial IP on IEEE 802.15.4 Low-Power Wireless Networks*. Ed. by Arch Rock Corporation. 2018. URL: http://www.ietf.org/internet-drafts/draft-ietf-6lowpan-format-12.pdf.

[38] IEEE Standard for Local and Metropolitan Area Networks Part 15.4. "Low-Rate Wire less Personal Area Networks (LR-WPANs) Amendment3 : Physical Layer (PHY) Specifications for Low-Data-Rate, Wireless, Smart Metering Utility Networks IIEEE Std 802.15.4g 2012 (Amendment to IEEE Std 802.15.4-2011)". In: *IEEE: P Piscataway*. Aug. 2012. NJ, USA.

[39] Ridha Soua. "Wireless Sensor Networks in Industrial Environment: Energy Efficiency, Delay and Scalability". Doctorat. Pierre et Marie Curie University of Paris. URL: https://www.researchgate.net/publication/281533607_Wireless_sensor_networks_in_industrial_environment_energy_efficiency_delay_and_scalability.

[40] Jungmin So and Nitin H. Vaidya. "Multi-channel Mac for Ad Hoc Networks: Handling Multi-channel Hidden Terminals Using a Single Transceiver". In: *Proceedings of the 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc 04*. 2004. New York, NY, USA.

[41] V Raman and N. H Vaidya. *Adjacent Channel Interference Reduction in Multichannel Wireless Networks Using Intelligent Channel Allocation*. Ed. by University of Illinois. URL: https://www.researchgate.net/publication/229012320_Adjacent_channel_interference_reduction_in_multichannel_wireless_networks_using_intelligent_channel_allocation.

[42] Ozlem Durmaz Incel. "A survey on multi-channel communication in wireless sensor networks". In: *Computer Networks* 55.13 (2011), 3081–3099. DOI: 10.1016/j.comnet.2011.05.020.

[43] Abdul-Hamid M.d, Abdullah-Al-Wadud M, and Ilyoung C. "A Schedule-based Multichannel MAC Protocol for Wireless Sensor Networks". In: *Jornal of Sensors* 10.10 (2010), pp. 9466–9480. DOI: 10.3390/s101009466.

[44] Y. Wu, J. A. Stankovic, T. He, and S. Lin. "Realistic and efficient multi-channel communications in wireless sensor networks". In: *Proceedings of the 27th IEEE International Conference on Computer Communications (Infocom 08)*. May 2008. DOI: 10.1109/INFOCOM.2008.175. Phoenix, AZ, USA.

[45] Panos N. Alevizos, Efthymios A. Vlachos, and Aggelos Bletsas. "Inference-Based Distributed Channel Allocation in Wireless Sensor Networks". In: *Jornal of Information Theory* arXiv:1703.06652v1 (2017). URL: https://arxiv.org/abs/1703.06652.

[46] Ridha Soua and Pascale Minet. "Multichannel assignment protocols in wireless sensor networks: A comprehensive survey". In: *Pervasive and Mobile Computing* 16.PartA (2015), pp. 2–21. DOI: 10.1016/j.pmcj.2014.04.004.

[47] Abdulmalik Alwarafy, Mohamed Abdallah, Bekir Sait Ciftler, Ala Al-Fuqaha, and Mounir Hamdi. "Deep Reinforcement Learning for Radio Resource Allocation and Management in Next Generation Heterogeneous Wireless Networks: A Survey". In: *arXiv* arXiv:2106.00574 (2021). DOI: 10.36227/techrxiv.14672643.

[48] Zhicong Chen, Qinghua Li, Lijun Wu, Shuying Cheng, and Peijie Lin. "Optimal data collection of multi-radio multi-channel multi-power wireless sensor networks for structural monitoring applications: A simulation study". In: *Structural Control Health Monitoring* 26.4 (2019), e2328. DOI: 10.1002/stc.2328.

[49] Jianchao Zheng, Honggang Zhang, Yueming Cai, Rongpeng Li, and Alagan Anpalagan. "Game-Theoretic Multi-Channel Multi-Access in Energy Harvesting Wireless Sensor Networks". In: *IEEE Sensors Journal* 16.11 (2016), 4587–4594. DOI: 10.1109/JSEN.2016.2539346.

[50] Asma Bagheri, Ataollah Ebrahimzadeh, and Maryam Najimi. “Game-theory-based lifetime maximization of multi-channel cooperative spectrum sensing in wireless sensor networks”. In: *Wireless Network* 26 (2020), 4705–4721. DOI: 10.1007/s11276-020-02369-1.

[51] Ruben Mennes, Felipe A. P. De Figueiredo, and Steven Latre. “Multi-Agent Deep Learning for Multi-Channel Access in Slotted Wireless Networks”. In: *IEEE Access* 8 (2020), pp. 95032 –95045. DOI: 10.1109/ACCESS.2020.2995456.

[52] Run-Fa Liao et al. “Deep-Learning-Based Physical Layer Authentication for Industrial Wireless Sensor Networks”. In: *Journal of Sensors* 19.11 (2019), p. 2440. DOI: 10.3390/s19112440.

[53] Hongsik Choi, Ju Wang, and Esther A. Hughes. “Scheduling for information gathering on sensor network”. In: *Wireless Networks* 15.1 (2009), 127–140. DOI: 10.1007/s11276-007-0050-9.

[54] Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: An introduction*. Ed. by Massachusetts The MIT Press Cambridge. 2014, p. 352. URL: https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf.

[55] Yi Chu. *Application of Reinforcement Learning on Medium Access Control for Wireless Sensor Networks*. Ed. by UK university of York department of Electronics. 2013. URL: https://etheses.whiterose.ac.uk/4795/1/thesis%20corrected.pdf.

[56] Goncalo Neto. *From Single-Agent to Multi-Agent Reinforcement Learning: Foundational Concepts and Methods*. Ed. by Institute For Systems and Portugal Robotics. 2005. URL: http://users.isr.ist.utl.pt/mtjspaan/readingGroup/learningNeto05.pdf.

[57] Mihail Mihaylov, Yann-Aël Le Borgne, Karl Tuyls, and Ann Nowe. “Decentralised reinforcement learning for energy-efficient scheduling in wireless sensor networks”. In: *International Journal of Communication Networks and Distributed Systems* 9 (2012), 207–224. DOI: 10.1504/IJCNDS.2012.048871.

[58] Muhidul Islam Khan, Kewen Xia, Ahmad Ali, and Nelofar Aslam. “Energy-aware task scheduling by a true online reinforcement learning in wireless sensor networks”. In: *International Journal of Sensor Networks* 25.4 (2017), pp. 244 –258. DOI: 10.1504/IJSNET.2017.087899.

[59] Shahzad Sarwar, Rabia Sirhindi, Laeeq Aslam, and G Mustafa. “Reinforcement Learning Based Adaptive Duty Cycling in LR-WPANs”. In: *IEEE Access* 8 (2020), pp. 161157 –161174. DOI: 10.1109/ACCESS.2020.3021016.

[60] Felipe Leno Da Silva and Anna Helena Reali Costa. “A Survey on Transfer Learning for Multiagent Reinforcement Learning Systems”. In: *Journal of Artifitial Intelligence Research* 64 (2019), pp. 645 –703. DOI: 10.1613/jair.1.11396.

[61] Zhe Chen and Robert C. Qiu. “Q-Learning Based Bidding Algorithm for Spectrum Auction in Cognitive Radio”. In: *2011 Proceedings of IEEE Southeastcon*. March. 2011. DOI: 10.1109/SECON.2011.5752976. Nashville, TN, USA.

[62] Kieu-Ha Phung et al. “Schedule-based multi-channel communication in wireless sensor networks: A complete design and performance evaluation”. In: *Ad Hoc Networks* 26 (2015), pp. 88–102. DOI: 10.1016/j.adhoc.2014.11.008.

[63] Kieu-Ha Phung, Bart Lemmens, Mihail Mihaylov, Lan Tran, and Kris Steenhaut. “Adaptive learning based scheduling in multichannel protocol for energy-efficient data-gathering wireless sensor networks”. In: *International journal of distributed sensor networks* 9.1 (2013), pp. 1–11. DOI: 10.1155/2013/345821.

[64] Waqas Rehan, Stefan Fischer, and Muhammad Maaz Rehan. “Machine-Learning Based Channel Quality and Stability Estimation for Stream-Based Multichannel Wireless Sensor Networks”. In: *Journal of Sensors* 16.9 (2016), p. 1476. DOI: 10.3390/s16091476.

[65] Valerio Di Valerio et al. “CARMA: Channel-Aware Reinforcement Learning-Based Multi-Path Adaptive Routing for Underwater Wireless Sensor Networks”. In: *IEEE Journal on Selected Areas in Communications* 37.11 (2019), pp. 2634–2647. DOI: 10.1109/JSAC.2019.2933968.

[66] Stephen S Oyewobi, Gerhard P. Hancke, Adnan M. Abu-Mahfouz, and J. Onumanyi Adeiza. “An Effective Spectrum Handoff Based on Reinforcement Learning for Target Channel Selection in the Industrial Internet of Things”. In: *Journal of Sensors* 19.6 (2019), p. 1395. DOI: 10.3390/s19061395.

[67] Mu Qiao, Haitao Zhao, Shengchun Huang, Li Zhou, and Shan Wang. “Optimal Channel Selection Based on Online Decision and Offline Learning in Multichannel Wireless Sensor Networks”. In: *Wireless Communications and Mobile Computing* 2017 (2017), pp. 1–14. DOI: 10.1155/2017/7902579.

[68] Jiag Zhu, Yonghui Song, Dingde Jiang, and Houbing Song. “Multi-Armed Bandit Channel Access Scheme With Cognitive Radio Technology in Wireless Sensor Networks for the Internet of Things”. In: *IEEEAccess* 4 (2016), pp. 4609 –4617. DOI: r10.1109/ACCESS.2016.2600633.

[69] Vishnu Ra, Irene Dias, Thulasi Tholeti, and Sheetal Kalyani. “Spectrum Access In Cognitive Radio Using A Two Stage Reinforcement Learning Approach”. In: *IEEE Journal of Selected Topics in Signal Processing* 12.1 (2017), pp. 20 –34. DOI: 10.1109/JSTSP.2018.2798920.

[70] Jing Ma, Tomoya Nagatsuma, Song-Ju Kim, and Mikio Hasegawa. “A Machine-Learning Based Channel Assignment Algorithm for IoT”. In: *International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. Feb. 2019. DOI: 10.1109/ICAIIC.2019.8669028. Okinawa, Japan.

[71] Suzan Ghamry Walid Kand Shukry. “Spectrum access in cognitive IoT using reinforcement learning”. In: *Cluster Computing* (2021). DOI: 10.1007/s10586-021-03306-3.

[72] Amir Alipour-Fanid, Monireh Dabaghchian, Ning Wang, Long Jiao, and Kai Zeng. “Online Learning-Based Defense Against Jamming Attacks in Multi-Channel Wireless CPS”. In: *IEEE Internet of Things Journal* 8.17 (2021), pp. 13278 –13290. DOI: 10.1109/JIOT.2021.3066476.

[73] C Zhang and V Lesser. “Coordinating Multi-Agent Reinforcement Learning with Limited Communication”. In: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems*. 2013. SaintPaul, Minnesota, USA.

[74] M Khan and B Rinner. “Energy-aware task scheduling in wireless sensor networks based on cooperative reinforcement learning”. In: *Proceedings of the IEEE International Conferenceon Communications Workshops*. 2014. Sydney, Australia.

[75] Ibrahim Mustapha, Borhanuddin Mohd Ali, and Aduwati Sali. "An energy efficient Reinforcement Learning based Cooperative Channel Sensing for Cognitive Radio Sensor Networks". In: *Pervasive and Mobile Computing* 35 (2017), pp. 165–184. DOI: 10.1016/j.pmcj.2016.07.007.

[76] Ang Gao, Chengyuan Du, Soon Xin Ng, and Wei Liang. "A Cooperative Spectrum Sensing with Multi-Agent Reinforcement Learning Approach in Cognitive Radio Networks". In: *IEEE Communications Letters* 25.8 (2021), pp. 2604–2608. DOI: 10.1109/LCOMM.2021.3078442.

[77] Sherief Hashima, Ehab Mahmoud Mohamed, Kohei Hatano, and Eiji Takimoto. "WiGig Wireless Sensor Selection Using Sophisticated Multi Armed Bandit Schemes". In: *The 13th International Conference on Mobile Computing and Ubiquitous Networking (ICMU21)*. Sep. 2021. Tokyo, Japan.

[78] Fuqiang Yao and Luliang Jia. "A Collaborative Multi-Agent Reinforcement Learning Anti-Jamming Algorithm in Wireless Networks". In: *IEEE Wireless Communications Letters* 8.4 (2019), pp. 1024 –1027. DOI: 10.1109/LWC.2019.2904486.

[79] Dejan Dasi et al. "Distributed Spectrum Management in Cognitive Radio Networks by Consensus-Based Reinforcement Learning". In: *Journal of Sensors* 21.8 (2021), p. 2970. DOI: 10.3390/s21092970.

[80] Qian Mao, Fei Hu, and Qi Hao. "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey". In: *IEEE Communications Surveys and Tutorials* 20.4 (2018), 2595–2621. DOI: 10.1109/COMST.2018.2846401.

[81] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521 (2015), 436–444. DOI: 10.1038/nature14539.

[82] Yunzeng Li, Wensheng Zhang, Cheng-Xiang Wang, Jian Sun, and Yu Liu. "Deep Reinforcement Learning for Dynamic Spectrum Sensing and Aggregation in Multi-Channel Wireless Networks". In: *IEEE Transactions on Cognitive Communications and Networking* 6.2 (2020), 464–475. DOI: 10.1109/TCCN.2020.2982895.

[83] He Li, Kaoru Ota, and Mianxiong Dong. "Learning IoT in Edge: Deep Learning for the Internet of Things with Edge Computing". In: *IEEE Network* 32.1 (2018), pp. 96–101. DOI: 10.1109/MNET.2018.1700202.

[84] Ruben Mennes, Felipe A. P. DE Figueiredo, and Steven Latre. "Multi-Agent Deep Learning for Multi-Channel Access in Slotted Wireless Network". In: *IEEE Access* 8 (2020), pp. 25032–25045. DOI: 10.1109/ACCESS.2020.2995456.

[85] Alex S.Leong, Arunselvan Ramaswamy, Daniel E.Quevedo, Holger Karl, and Ling Shi. "Deep reinforcement learning for wireless sensor scheduling in cyber–physical systems". In: *Automatica* 113 (2020), p. 108759. DOI: 10.1016/j.automatica.2019.108759.

[86] Yuhan Su, Xiaozhen Lu, Yifeng Zhao, Lianfen Huang, and Xiaojiang Du. "Cooperative Communications With Relay Selection Based on Deep Reinforcement Learning in Wireless Sensor Networks". In: *IEEE Senesors Journal* 19.20 (2019), pp. 9561–9569. DOI: 10.1109/JSEN.2019.2925719.

[87] Jackson Francomme, Gilles Mercier, and Thierry Val. "Beacon Synchronization for GTS Collision Avoidance in an IEEE 802.15.4 Meshed Network". In: *IFAC Proceedings Volumes* 40.22 (2007), pp. 215–222. DOI: 10.3182/20071107-3-FR-3907.00029.

[88] Anis Koubaa, Andre Cunha, and Mario Alves. “A Time Division Beacon scheduling Mechanism for IEEE 802.15.4/Zigbee Cluster-Tree Wireless Sensor Networks”. In: *19th Euromicro Conference on Real-Time Systems (ECRTS’ 07)*. July 2007. DOI: 10.1109/ECRTS.2007.82. Pisa, Italy.

[89] Reinaldo A.C.Bianchi, Luiz A. Celiberto Jr., Paulo E.Santos, Jackson P.Matsuura, and Ramon Lopez de Mantaras. “Transferring knowledge as heuristics in reinforcement learning: A case-based approach”. In: *Artificial Intelligence* 226 (2015), pp. 102–121. DOI: 10.1016/j.artint.2015.05.008.

[90] Xinyan Xu et al. “Throughput-Guaranteed Distributed Channel Assignment and Scheduling Algorithms with Low Complexity for Multichannel Wireless Sensor Networks”. In: *Journal of Sensors* 2019 (2019), pp. 1–12. DOI: 10.1155/2019/6171625.

[91] C.E. Shannon. “A mathematical theory of communication”. In: *The Bell System Technical Journal* 27.3 (1948), pp. 379 –423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[92] J. Kulshrestha and Manas Kumar Mishra. “An adaptive energy balanced and energy eficient approach for data gathering in wireless sensor networks”. In: *Ad Hoc Networks* 54 (2017), pp. 130–146. DOI: 10.1016/j.adhoc.2016.10.013.

[93] Nils Morozs, Tim Clarke, and David Grace. “A novel adaptive call admission control scheme for distributed reinforcement learning based dynamic spectrum access in cellular networks”. In: *Proceedings of International Symposium on Wireless Communication Systems (ISWCS)*. August 2013. DOI: 2-s2.0-84903218425. Ilmenau, Germany.

[94] Huy Pham, Hung Manh.L, David Feil-Seifer, and Luan V.Nguyen. “Cooperative and Distributed Reinforcement Learning of Drones for Field Coverage”. In: *arXiv* arXiv:1803.07250 (2018). URL: https://arxiv.org/abs/1803.07250.