University of Padova
Academic Year 2018-2019 (797^{th})
Department of Information Engineering
Master's Thesis in Telecommunications Engineering

National Institute of Standards and Technology
U.S. Department of Commerce

Scheduling the Data Transmission Interval in IEEE 802.11ad

Author: Tommy Azzino
Advisor: Michele Zorzi*
Co-advisors: Nada Golmie†, Tanguy Ropitault†

*University of Padova, Italy
†National Institute of Standards and Technology, USA

To my parents, who constantly supported and helped me during my studies and career, and without whom I could not have achieved this important goal

Universa Universis Patavina Libertas

Liberty of Padova, Universally and for All

University of Padova's motto

Abstract

IEEE 802.11ad standard enables communications in the mm-wave and unlicensed 60 GHz band. Propagation at mm-wave frequencies accounts for increased path loss and sensitivity to blockage when compared to sub-6-GHz frequencies. To cope with this problem directional transmissions, through beamformed links, are demanded. In order to realize communications in this vast spectrum, the standard introduced a novel Medium Access Control (MAC) layer which enables contention-free and contention-based channel access.

Most of the related literature focuses on the Physical (PHY) layer and the investigation of novel beamforming techniques. However, the increased complexity associated with channel access at MAC must be addressed through the development of a resource scheduling algorithm for IEEE 802.11ad.

This thesis presents and implements in network simulator 3 (ns-3) three different resource scheduling schemes for this technology. Among these approaches, one exploits the directionality of communications to achieve concurrent transmissions over the same medium, through spatial sharing. Another approach exploits Reinforcement Learning (RL) to find the optimal duration of each contention-free access period and, therefore, enhance the overall efficiency of the network.

Contents

Chapter 1

Introduction

As the sub-6-GHz spectrum is constantly more and more crowded with devices communicating at such frequencies, industry and research communities are focusing their attention on the worldwide largely available millimeter-wave spectrum. Moreover, over the last few years, new application usages demand increasingly high throughput and low latency to allow a new generation of services such as: Augmented Reality (AR)/Virtual Reality (VR), mobile offloading, indoor and outdoor wireless backhaul and high-bandwidth connectivity with TVs and monitor displays. To meet the requirements imposed by these applications, mm-wave communication has recently gain huge momentum thanks to the wide spectrum available at those frequencies. The IEEE 802.11ad amendment [1] defines operations in the 60 GHz lndustrial, Scientific and Medical (ISM) unlicensed band and provides mm-wave communications to enable the aforementioned applications. However, signal propagation at the 60 GHz band significantly differs from that at the usual 2.4 GHz and 5 GHz bands and, it is characterized by high propagation loss and significant sensitivity to blockage. At a typical IEEE range of 10 meters we can encounter an additional 22 dB attenuation with respect to the 5 GHz band [2], according to Friis equation. Therefore, efficient use of this vast spectrum requires a fundamental rethinking of the Wi-Fi operations based on legacy IEEE technologies operating in the sub-6-GHz band. To this end, IEEE 802.11ad introduces the important transition from an omni-directional

to a directional usage of the wireless medium. As such, the range of mm-wave communications can be increased through Beamforming (BF) where the transmission power is focused toward a specific spatial direction, realizing the directional communication paradigm. This is accomplished through antenna arrays composed by several antenna elements that can be controlled in order to properly steer beams in the direction of intended users. Directional communications significantly reduce interference among concurrent transmissions, allowing for efficient spatial reuse of the same medium through Spatial Sharing (SPSH).

Because of the peculiarities of propagation at mm-wave frequencies, protocols designed for communications in the sub-6-GHz band cannot be exploited in the 60 GHz band without an adjustment to the new spectrum. In particular, changes are demanded at every layer of the protocol stack, especially at the Physical (PHY) and Medium Access Control (MAC) layers. In contrast with legacy technologies, IEEE 802.11ad implements a hybrid MAC layer where channel access can be realized according to both contention-based or contention-free allocations. Concerning the contention-based approach, devices in the network compete with each other in order to acquire channel access for satisfying their traffic demands. On the other hand, a contention-free period is specifically assigned for the exclusively communication of a pair of nodes, therefore any other station in the network is not allow to transmit during that channel time. From a channel access perspective, the standard provides high flexibility of design, nevertheless, it does not specify any policy which allocates resources according to the rules defined for accessing the medium. To the best of my knowledge, this is the first work whose aim is to provide the implementation of different scheduling schemes that exploit the hybrid MAC layer features introduced by IEEE 802.11ad to fully realize efficient communication in the 60 GHz band.

In particular, in this work we develop three different resource scheduling algorithms for the hybrid MAC access of IEEE 80.11ad using the network simulator 3 (ns-3) [3]. Ns-3 is the de-facto network simulator employed by the industry and academia to develop and test many network technologies. Results obtained with ns-3 are widely accepted and recognized by the re-

search community. First, we design a simple scheduling approach which will be utilized as a baseline performance for the comparison with more advanced strategies. Our second scheme exploits the directionality of communications at mm-wave frequencies to achieve spatial sharing. In particular, this algorithm creates a graph, symbolizing the behavior of interference between contention-free allocations, through which concurrent transmission of different traffic flows can be realized to enhance the overall spectral efficiency of the network. Lastly, our final approach exploits Reinforcement Learning (RL) to optimize the duration for each contention-free period allocated in the network. This, in turn, allows for better resources exploitation and avoids the waste of precious network time. Nowadays, the application of RL, and more generally Machine Learning (ML), techniques constitutes a major trend in the domain of network research. This motivates our decision of investigating RL as a mean for solving 802.11ad-related problems.

1.1 Thesis outline

The reminder of this thesis is organized as follows: Chapter 2 presents the general framework adopted in this work, providing an overview of the IEEE 802.11ad Standard with its main features, as well as, a survey about the main concepts behind RL. In addition, a summary with a brief description of similar works in the literature is provided. Chapter 3 discusses in depth the proposed scheduling approaches, outlining the relevant modeling and the design choices made. Chapter 4 provides a detailed analysis of the simulation results for the developed schemes. Ultimately, Chapter 5 concludes this thesis and highlights some possible future works related to this project.

Chapter 2

Framework

This chapter provides an overview of the technologies and techniques involved in this work. Initially, the IEEE 802.11ad standard is presented, with a focus on the functionalities introduced at the MAC layer. This standard enables communications at millimeter-wave frequencies in the unlicensed 60 GHz band. At this high-frequency spectrum, the adverse signal propagation characteristics require a fundamental rethinking of the communication principles at the basis of previous standards, such as the well-known and worldwide-adopted IEEE 802.11a/b/g/n/ac standards.

Subsequently in this chapter, the theoretical principles of RL are discussed to pave the way for their application in this work. Nowadays, a major trend in the wireless communications research field regards the application of ML techniques to cope with several optimization or NP-hard problems at every layer of the protocol stack. Among several ML methods, RL has emerged thanks to its ability to learn how to interact with the surrounding environment based on trials and errors. Thanks to RL, an agent learns the optimal policy to follow under certain environmental conditions through experience, by optimizing the expectation of the reward function.

In the final part of this chapter, an overview of related work in the literature is provided. Although the introduction of this standard dates back to 2012, there are very few works related to the problem addressed in this project, especially regarding the use of RL-based algorithms and the ns-3

implementation we are going to present later.

2.1 Overview of IEEE 802.11ad standard

The IEEE 802.11ad technology allows wireless devices to communicate in the unlicensed 60 GHz ISM band and provides multi-Gbps data rates for bandwidth-demanding applications. However, signal propagation at these millimeter-wave frequencies is significantly different from that at the usual 2.4 GHz and 5 GHz bands. In order to enable efficient and reliable communication over this vast spectrum, a substantial rethinking of the common Wi-Fi operations is demanded. In particular, the main design change regards a shift from an omni-directional type of communication to a directional use of the wireless medium.

The IEEE 802.11ad amendment [1] introduces a set of functionalities to overcome the challenges associated with the communication at such frequencies, and enables, to Wi-Fi users, the rise of novel applications such as: high speed file exchange between local devices, instant wireless synchronization and cable replacement, as for example the connection to a high definition wireless display. The immediate consequence of the propagation at mm-wave frequencies is increased signal attenuation, with a predicted additional 22 dB attenuation at a range of 10 meters in comparison with a system operating in the 5 GHz band [2]. In addition, oxygen absorption peaks at 60 GHz, making the propagation even more complicated. Another important aspect concerning the communication at 60 GHz regards the quasi-optical propagation behavior where the received signal is primarily composed by its Line of Sight (LoS) component; on the other hand, Non Line of Sight (NLoS) communication is achievable when the environment contains strong reflective surfaces. These peculiarities of mm-waves' propagation require the adoption of a directional communication scheme that makes use of beamforming techniques to cope with the increased signal attenuation.

The passage to a directional communication paradigm entails a complete rethinking of the operations: this was accomplished by the standardizing committee through an adaptation of the 802.11 architecture with the intro-

duction of new features such as the IEEE 802.11ad beamforming mechanism and the hybrid MAC layer design. In the following, after an analysis of the IEEE 802.11ad PHY layer, the major changes and novelties introduced by the standard are presented, with an eye toward the new MAC layer functionalities.

2.1.1 Physical layer

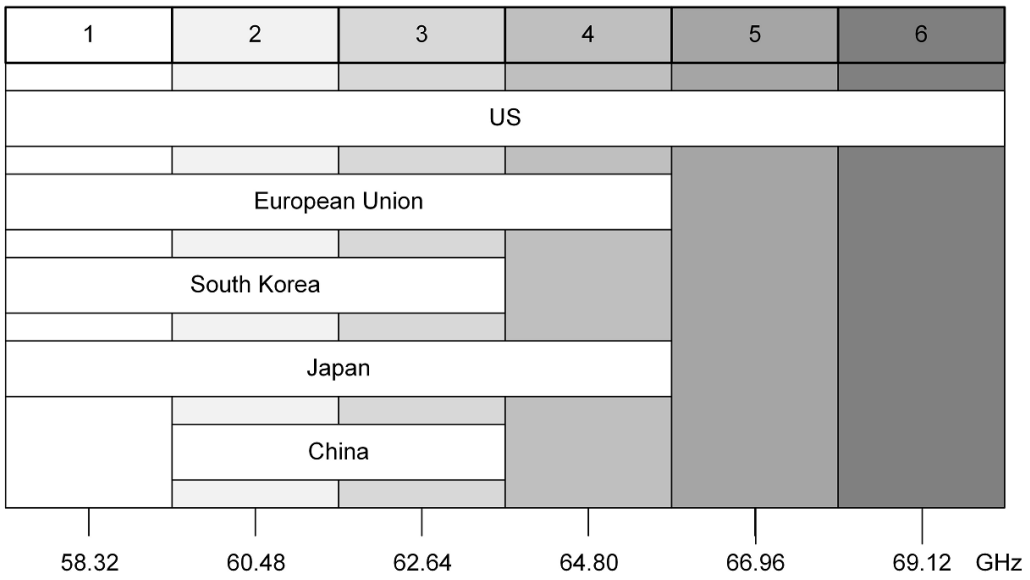

Fig. 2.1: Worldwide 802.11ad available channels at 60 GHz (source: [4])

As reported in Fig. 2.1, six channels are defined in the 60 GHz band, spanning from center frequency 58.32 GHz to 69.12 GHz, and with nominal bandwidth of 2.16 GHz each. Channel 2, available in every region, has been designated as the default channel.

IEEE 802.11ad introduces three different PHY layers with the objective of fulfilling diverse requirements, such as high throughput or robustness. The *Control PHY* represents a reliable physical mean for low Signal to Noise Ratio (SNR) communication during beamforming operations. On the other hand, the *Single Carrier (SC) PHY* guarantees power efficient and low complexity transceiver implementation. Ultimately, the *OFDM PHY* pushes this

technology to the maximum achievable physical rate thanks to the Orthogonal Frequency-Division Multiplexing (OFDM) method. Although providing much higher rates, this latter PHY mode is no longer supported by the standard.

The Control PHY implements Modulation and Coding Scheme (MCS) 0 in order to extend the range of 60 GHz operations and provide the reliability needed to exchange beacon, management and control frames between Stations (STAs) and the Personal Basic Service Set (PBSS) Control Point/Access Point (PCP/AP). Providing a throughput of 27.5 Mbps, the mandatory Control PHY defines the minimum rate at which all devices shall communicate before the establishment of a beamformed link.

The SC PHY, covering MCSs from 1 to 12, permits low complexity and energy efficient transceiver implementation and guarantees data rates of up to 4.62 Gbps. The lowest rate associated with MCS 1 allows for a 485 Mbps rate. Clearly, higher the MCS, higher is the associated achievable data rate. Among all MCSs of the SC PHY, only those ranging from 1 to 4 are mandatory for all devices.

The OFDM PHY (MCSs from 13 to 24) outperforms all other PHY layers in terms of maximum achievable throughput, at the cost of increased transceiver complexity and energy consumption. However, thanks to the adoption of 64-QAM modulation, this modality provides the highest 802.11ad data rates of up to 6.75 Gbps.

Due to the necessity of implementing low complexity and energy efficient equipment characterized by a competitive cost, nearly all 802.11ad-compatible devices adopt only the Control and SC PHY layers. As a consequence, the OFDM-based PHY layer is currently not supported.

2.1.2 Beacon interval

Legacy IEEE 802.11 technologies, working at sub-6-GHz frequencies, organizes the medium access, in time, through periodical Beacon Intervals (BIs). Each BI starts with the omni-directional transmission of a beacon frame by the PCP/AP. A beacon frame announces the existence of a certain Wi-Fi

network and carries useful management information. The remaining time associated with the current BI is entirely devolved to the transmission of data, typically following a contention-based access scheme, where STAs and the PCP/AP compete between each other to acquire a Transmission Opportunity (TXOP). The length of a BI is expressed in multiples of a Time Unit (TU) (1 TU = 1024 μs) and typical values are chosen in the range of 100 ms. Having a longer BI duration increases the connection delay for new nodes in the network, but on the other hand, limits the continuous transmission of management frames, thereby increasing the overall throughput performance.

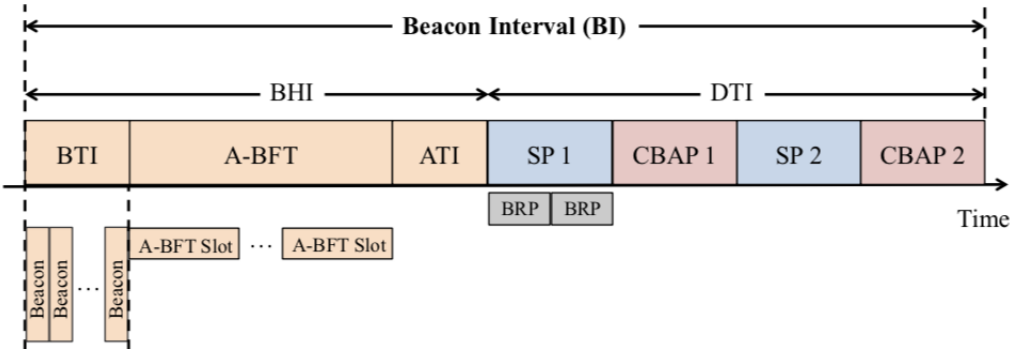

Fig. 2.2: IEEE 802.11ad beacon interval structure (source: [5])

To address the challenges of mm-wave propagation, IEEE 802.11ad introduces a modified concept of beacon interval, as depicted in Fig. 2.2. Specifically, the 11ad BI is composed by two main time intervals namely: the Beacon Header Interval (BHI) and the Data Transmission Interval (DTI). The former is intended for exchanging beacon, management as well as network announcement frames, employing a sweeping directional mechanism as explained later in this section. Following the BHI, there is the DTI, where STAs and the PCP/AP can access the medium for data transmission according to different schemes.

The BHI is further divided into three sub-intervals. During the Beacon Transmission Interval (BTI) one beacon frame for each antenna sector is transmitted by the PCP/AP to cover all possible spatial directions (the antenna's azimuth space is divided into several spatial sectors where directional frames are transmitted). This sweeping procedure allows to overcome the in-

creased attenuation, especially when unassociated devices need to be discovered by the network controller. Furthermore, beacon frames carry important network announcements such as the schedule in the following DTI and other medium access parameters, as described later. Most importantly, the BTI serves to conduct beamforming training of the PCP/AP's antenna sectors. As an integral part of this latter BTI's purpose, the Association Beamforming Training (A-BFT) sub-interval is exploited by STAs to find the optimal antenna sector for communication with the PCP/AP. Finally, the optional Announcement Transmission Interval (ATI) is used by the PCP/AP to exchange management frames with beam-trained associated STAs. In fact, while communication during BTI and A-BFT intervals occurs exploiting the Control MCS (MCS 0) to increase range and robustness of transmitted frames, the transmission of management frames during the ATI leverages beam-trained STAs, and therefore, can be more efficient thanks to higher order MCSs.

The DTI, as depicted in Fig. 2.2, is made up of Contention-Based Access Periods (CBAPs) and scheduled Service Periods (SPs), where stations can exchange data frames. During a CBAP multiple nodes compete to acquire the channel following the Enhanced Distributed Channel Access (EDCA) function; conversely, a SP consists of a specific contention-free time interval intended for the exclusive communication between dedicated pair of nodes.

As seen so far, the introduction of a directional communication paradigm requires an increased overhead during the BHI in comparison to legacy Wi-Fi technologies, in order to realize communication in the mm-wave spectrum. This problem gets worse when considering that directional beacon frames, transmitted during the BTI at the lowest MCS to provide resilience and extended range, contain several additional information including network scheduling and beamforming data. Moreover, the exchange of frames during the A-BFT, which also employs MCS 0, creates additional non-negligible overhead. This poses a relevant problem especially when reduced BI durations are adopted for delay sensitive applications, such as video streaming. The IEEE 802.11ad standard provides several strategies to address this problem. First, it is possible to split the beacon sweeping procedure for all PCP/AP's antenna sectors among consecutive BIs. Second, it is possible to

periodically have some BIs without the A-BFT sub-interval. Despite a clear overhead reduction, both methods produce additional association delay. An optimal strategy exploits the possibility of sending management frames during the ATI, where already beam-trained stations can transmit with a more efficient MCS. In this way, it is possible to move information from the spectrally inefficient beacon frames to the frames transmitted during the ATI period. As we will see later in this work, scheduling and other management information are transmitted from the PCP/AP to associated STAs and vice versa, if the case, during the ATI.

2.1.3 Beamforming training

IEEE 802.11ad introduces a directional communication scheme to alleviate the increased attenuation due to the propagation at mm-wave frequencies. In order to exploit the benefits of beamforming antenna gain, the standard rises the concept of *virtual* antenna sectors that discretize the antenna azimuth; a specific sector focuses antenna gain in a particular spatial direction. The beamforming training procedure aims at finding the optimal transmit and receive antenna sectors for a certain pair of stations and, at a first stage, utilizes virtual sectors to reduce the space of possible antenna array configurations to be explored.

The beamforming procedure is composed by two sub-phases. Initially, a Sector-Level Sweep (SLS) phase provides an initial coarse-grain antenna sector between a pair of stations. Later, a fine-tuning of the selected sectors, during the Beam Refinement Protocol (BRP) phase, provides the best beam for the communication between pairs of stations. Typically, during the SLS phase, only the transmit sector for each node is trained, whereas in the following BRP phase both transmit and receive sectors are trained to find the best configuration. In the following, we will focus on the SLS phase with a description of the beamforming training procedure when a STA is associated with the PCP/AP and, later, provide a high-level description of BRP mechanism.

As depicted in Fig. 2.3, primarily, the SLS phase is carried in the BTI and

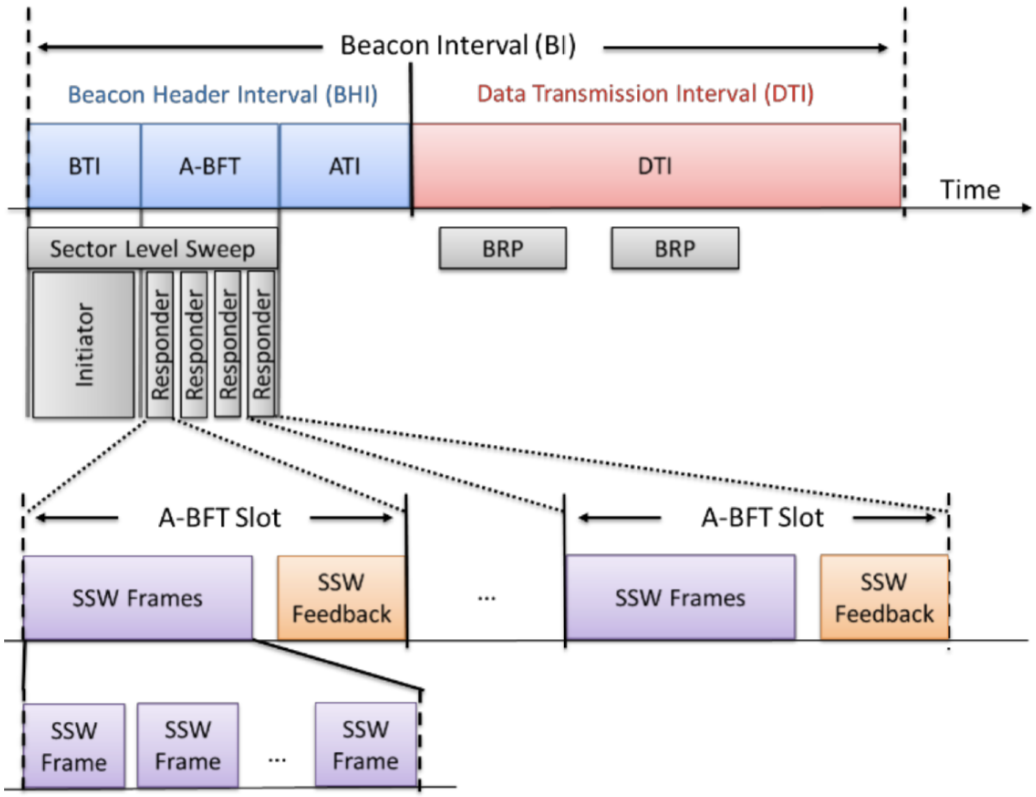

Fig. 2.3: IEEE 802.11ad beamforming training (source: [5])

A-BFT sub-intervals of the BHI; in this case, the PCP/AP finds the optimal sector for the communication with its paired STAs, contrarily, every STA finds the optimal sector for transmissions directed to the PCP/AP. However, the SLS phase can be initiated outside the BHI, as it is used to find the optimal antenna sectors between STAs that intend to communicate directly with each other. In details, during the SLS phase, a pair of stations exchange a series of Sector Sweep (SSW) frames (or beacons during SLS beamforming training at the PCP/AP) covering all virtual sectors to find the one providing highest signal quality. During the SLS phase, a station that transmits first is called the initiator, while the other one is depicted as responder. There exist two different modalities: Transmit Sector Sweep (TXSS) and Receive Sector Sweep (RXSS). In the former case, frames are transmitted over different sectors while the responder is listening in quasi-omnidirectional mode. Every

frame is marked by an antenna and sector IDs that are used to successfully identify the best antenna sector. In the latter case, transmission occurring over the best founded sector allows for searching the best receive sector at the responder station. Totally, there are four possible combinations: TXSS or RXSS at both initiator and responder, TXSS at initiator while RXSS at responder and RXSS at initiator while TXSS at responder. The beamforming training operations during the association phase of a station are represented in Fig. 2.3: the PCP/AP employs its beacon sweeping procedure during the BTI to start a SLS phase as initiator for all STAs. Thus, every beacon frame incorporates SSW specific information. The A-BFT interval implements a contention-based slotted period where each STA, acting as responder, tries to acquire a specific A-BFT Slot to complete the SLS procedure with the PCP/AP. In particular, each A-BFT Slot represents a fixed time interval where the connecting STA transmits SSW frames to the PCP/AP; in this case, each SSW frame contains the information regarding the best transmit antenna sector from the PCP/AP to that STA, founded through the optimal beacon frame of the previous phase. A slot ends with the transmission by the PCP/AP of a SSW Feedback frame that signals the successful completion of the SLS procedure (i.e. no collision with other STAs that are requiring an A-BFT slot) and reports to the responder the best antenna configuration for transmissions directed to the PCP/AP.

The BRP phase can be defined as a refinement process with respect to the sub-optimal sectors founded during the SLS phase; in fact, as they are computed adopting imperfect quasi-omnidirectional patterns, they can lead to poor performance. Thorough BRP, antenna weight vectors can be optimized independently from the pre-defined SLS sectors resulting in additional beamforming gain and consequently, throughput gain. Although there are several optional beam refinement mechanisms, the mandatory beam refinement transactions represents an iterative procedure through which both initiator and responder can request training for their receive or transmit antenna. In a nutshell, BRP transaction tests a set of directional transmit or receive antenna patterns against the best known configuration. Given that a BRP procedure follows the SLS phase, a reliable communication exists, there-

fore several different configurations can be tested using the same transmitted frame, resulting in less overhead for the BRP transaction. As indicated in Fig. 2.3, typically the BRP phase is conducted in the DTI immediately after a SLS phase in the BHI. In particular, a STA that acquires the channel can request a series of BRP transactions after a BRP setup phase.

2.1.4 Data transmission interval

As opposed to legacy Wi-Fi technologies, IEEE 802.11ad employs a hybrid MAC layer to satisfy the envisioned use cases and realize the communication in the 60 GHz band. The exchange of data between stations during the DTI can occur according to three different mechanisms: through contention, thanks to a scheduled channel time allocation and, ultimately, by means of a dynamic channel time allocation.

Concerning the first case, stations compete with each other to acquire access to the medium following the IEEE 802.11 EDCA, which defines traffic categories to support Quality of Service (QoS). Channel access is granted after a station obtains a TXOP by winning an instance of the EDCA. A time interval in the DTI that satisfies these characteristics is defined as Contention-Based Access Period (CBAP). In a contention-based channel access approach dealing with directional communications, the deafness problem arises. In this case, a deaf node does not receive directional information from other nodes due to misaligned antenna patterns. This fact represents a problem for carrier sensing during contention-based access and can lead to an increased number of collisions. As such, for example there can be interference in the network even if the medium was considered to be idle. To cope with this problem, the contention-based access in IEEE 802.11ad is adapted for a directional medium usage through support to multiple Network Allocation Vector (NAV) timers; thus enabling physical and virtual carrier sensing.

On the other hand, a scheduled channel time allocation consists of a specific time interval of the DTI dedicated exclusively to a pair of communicating nodes. Re-occurring every BI, this pseudo-static amount of time is referred to as Service Period (SP). Clearly, there can be as many SPs as the

pairs of communicating nodes that require a specific time interval. According to this definition, the pseudo-static channel time allocation closely matches the well-known Time Division Multiple Access (TDMA) mechanism, where channel access is slotted in time and at each user is assigned a particular recurring time slot. The schedule of SPs is communicated by the PCP/AP to all STAs in the network, therefore every node knows when it needs to wake up to transmit or receive data from another node and when to go into sleep mode; this allows for very efficient power saving. According to IEEE 802.11ad amendment [1], STAs can reserve a dedicated SP through a specific request sent to the PCP/AP. In this request, the transmitting STA specifies the receiving node and a set of parameters that define its traffic requirements, such as: duration of the SP and traffic type (asynchronous or isochronous). The asynchronous mode is intended for non-recurring traffic as the download of a file, whereas the isochronous mode represents re-occurring traffic such as the one associated with a wireless display application. The duration of an allocated SP is computed after a beamformed link is available between the communicating stations, and requires the knowledge of the physical layer data mode adopted in the network. The actual schedule that contains the arrangement of CBAPs and requested SPs is announced by the PCP/AP in an Extended Schedule Element (ESE) during the BTI or ATI.

Ultimately, channel time can be allocated in a dynamic way according to a polling mechanism. This scheme provides high flexibility for resource allocation and suits directional communications. As a matter of fact, during dynamic allocation in the DTI, the PCP/AP acquires the channel so that STAs can listen in directional mode for any information coming from it. The PCP/AP sends a series of Polling frames to associated STAs, afterwards, each polled station answers with a Service Period Request (SPR) to require channel time. The PCP/AP elaborates the requests and allocates the available channel time according to the traffic characteristics from each request. The PCP/AP announces each allocation with a separate grant period, where the STAs involved in the allocation receive a Grant frame, meaning they are allowed to communicate. This procedure is reliable since in every moment individual directionally addressed frames are sent to each node. However, a

polling procedure exhausts its effect in the current BI, implying that a new polling phase is required between the same two stations in the following DTI. This slight drawback, that accounts for increased overhead, makes dynamic allocation suitable for bursty type of traffic. In case of communication with the PCP/AP, the grant period consists of only one Grant frame transmitted to the non-PCP/AP node. Finally, the standard enables dynamic allocation in both CBAPs and SPs.

2.1.5 Scheduling the DTI

In this sub-section, we will discuss how the DTI can be scheduled, in particular we will see: how the DTI can be organized between SPs and CBAPs, how a STA can ask for specific allocation time (i.e. a SP), and how the schedules for the following DTI are announced in the network. By the end of this sub-section, it will be clear why it is necessary to develop a scheduling algorithm that takes advantage of the new features introduced by IEEE 802.11ad at MAC layer. Moreover, the increased complexity at MAC channel access requires an effective resource scheduling algorithm capable of fully realizing multi-Gbps and robust communication at 60 GHz. This motivates the work conducted in this project.

First, we already know that IEEE 802.11ad exploits a hybrid MAC resulting from the possibility of combining CBAPs and SPs. However, the standard does not provide specific guidelines regarding any number or order that define how these channel access methods can exist together; leaving to the network designer a great degree of freedom. Therefore, the need of developing a scheduling algorithm that effectively realizes the co-existence between these two channel access techniques.

Second, STAs can reserve a dedicated SP by sending an Add Traffic Stream (ADDTS) Request frame of type management to the PCP/AP. Such a request carries a Directional Multi-Gigabit (DMG) TSPEC element that contains all parameters of the requested allocation. The PCP/AP sends back an ADDTS Response frame to either accept or reject the requested allocation based on its admission control policy (i.e. scheduling algorithm) and the

available time resources. Further, a STA can request an allocation through a SPR frame, however this strategy is not treated in this work.

The structure of a DMG TSPEC element is shown in Fig. 2.4. The 3-octet DMG Allocation Info field defines: *Allocation ID*, *Allocation Type* (CBAP or SP), *Allocation Format* (asynchronous or isochronous) and a series of flags to indicate if the allocation is: pseudo-static (a non pseudo-static allocation lasts for one BI and then deleted), truncatable (i.e. the allocation can be truncated in time if necessary), extendable (i.e. the allocation can be extended in time if necessary). Most importantly, the DMG Allocation Info field contains *Destination* and *Source* Associaton Identifiers (AIDs) of the stations involved in the allocation, that, together with the Allocation ID, uniquely identify a particular allocation.

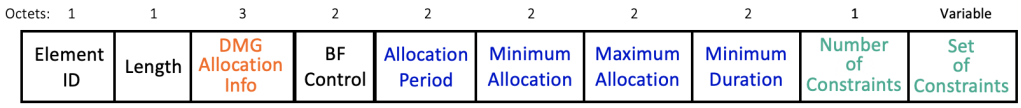

Fig. 2.4: IEEE 802.11ad DMG TSPEC element

The BF Control field contains information for beamforming training. If the traffic required is isochronous then the parameter: *Allocation Period* represents the period over which the allocation repeats (fraction or multiple of a BI), *Minimum Allocation* represents the minimum accepted allocation duration in μs in each Allocation Period and the *Maximum Allocation* parameter represents the requested allocation duration in μs in each Allocation Period. If the traffic is asynchronous then: *Allocation Period* represents the period over which the *Minimum Allocation* applies, and *Minimum Allocation* represents the minimum accepted duration in μs that the STA expects. The *Maximum Allocation* parameter is not exploited in the case of asynchronous traffic. At the end, a number of up to 15 constraints can be specified; each constraint can represent specific traffic requirements requested by the STA. The parameters *Minimum Allocation* and *Maximum Allocation* are two very important values the requesting STA must specify in its request and they must be computed according to a fixed rate (i.e. the data rate provided by the selected PHY mode).

Finally, the are two main ways to announce the schedule of the following DTI. The first approach consists on simply allocating the entire DTI as a CBAP, this is accomplished by sending DMG Beacon frames during the BTI with the special field, called *CBAP Only*, set to 1. In this way, through the transmission of directional beacon frames, every STA in the network is aware about the need of competing with other STAs in order to acquire a TXOP during the next DTI. Clearly, the usage of this approach prevents the possibility of scheduling dedicated SPs. Alternatively, the DTI can be scheduled through an Extended Schedule Element (ESE), which can be attached to either an Announcement frame or a DMG Beacon frame. The ESE contains the allocations for the current DTI, bearing in mind that this element is transmitted during the BTI or ATI sub-intervals before the DTI. In a broader way, we can also state that an ESE defines the organization of the DTI between CBAPs and/or SPs. To this end, it is composed by a series of *Allocation Fields* (from 1 up to 15), each of which has the structure reported in Fig. 2.5.

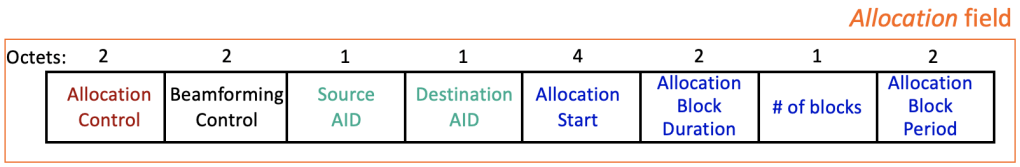

Fig. 2.5: IEEE 802.11ad allocation field of the Extended Schedule Element

The Allocation Control field identifies important parameters of the allocation that were defined in a DMG Allocation Info element of the DMG TSPEC. Among all the parameters of the Allocation Field, the most important one identifies the type of allocation, namely: CBAP or SP. Source AID identifies the station that initiate channel access, while Destination AID identifies the station communicating with the source. This latter can be a broadcast AID. Each allocation is scheduled with a specific start time defined by the parameter *Allocation Start* and a fixed duration expressed in μs according to the *Allocation Block Duration* field. The duration range in μs is $1 - 32767$ for a SP and $1 - 65535$ for a CBAP. The data exchange (ACKs included) shall be completed before the end of a CBAP/SP. The *Number of*

Blocks field identifies how many blocks make up the allocation and, *Allocation Block Period* the time in μs between the start of two consecutive blocks. In this work, we will consider the allocation of a unique block for every kind of allocation, thus the parameter Allocation Block Duration represents the duration of the single-allocated block during one DTI.

We just saw how scheduled allocations are announced to the STAs in the network and how a STA can require a dedicated channel time to the PCP/AP. Nevertheless, how these allocations are accepted and organized in time by the PCP/AP still remain an open question, hence the urge of developing a resource scheduling policy for IEEE 802.11ad. In the reminder of this section, we will introduce a novel technique that allows the design of advanced and efficient scheduling algorithms.

2.1.6 Spatial sharing

Highly directional communications of IEEE 802.11ad devices strongly reduce interference outside the beam direction. This allows enormous spatial re-utilization of the same frequency and enables a significant increase in throughput performance. To exploit this peculiarity of directional transmissions, IEEE 802.11ad amendment introduces a 11ad-unique mechanism called SPSH.

The SPSH procedure relies on measuring the interference between communication links separated in time and space and, it is exclusively intended in the presence of SPs. The PCP/AP can decide to overlap those scheduled links in time and measure how the introduced interference impacts the quality of each link. If the PCP/AP decides the amount of interference is negligible, then it can schedule the assessed links concurrently in time and, therefore achieve spatial sharing whilst improving the overall network spectral efficiency. The SPSH technique is executed in two phases as described below.

During the initial *assessment* phase, the PCP/AP starts a radio measurement procedure with the intended STAs to assess the possibility to perform spatial sharing. This procedure is initiated with the transmission from the

PCP/AP of a Directional Channel Quality Request as part of the Radio Measurement Request frame to the participating STAs. Every STA involved must have conducted beamforming training before any measurement is performed. During this procedure we identify a candidate SP which is to be assessed with respect to an existing SP. A candidate SP can be a new SP scheduled in the next BI or a SP with allocated channel time in the DTI. If the candidate SP is already allocated then the PCP/AP sends a Directional Channel Quality Request to the STAs involved in the existing allocation to assess the possibility of spatial sharing with the existing SP. With a Directional Channel Quality Request the PCP/AP asks to report the Average Noise plus Interference Power Indicator (ANIPI) measure after those STAs involved perform channel measurements using the antenna configuration selected for the communication with its peer STA. When a STA completes the required measurements, it sends back to the PCP/AP a Directional Channel Quality Report, as part of a Radio Measurement Report frame, containing the results of these measurements.

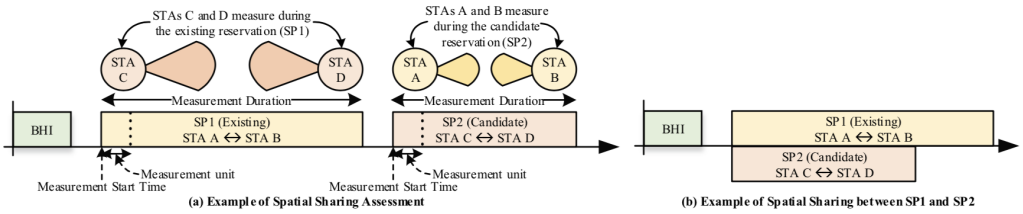

Fig. 2.6: IEEE 802.11ad spatial sharing mechanism (source: [1])

In the final *execution* phase, based on the reported channel measurements, the PCP/AP evaluates the channel quality between candidate and existing SPs and decides whether or not to implement spatial sharing. If spatial sharing can be achieved, these SPs are overlapped in time to maximize the performance; however, how to enhance the performance based on spatial sharing depends on the specific implementation of the manufacturer. To sustain a reliable communication between the SPs involved in the spatial sharing mechanism, the PCP/AP periodically sends a Directional Channel Quality Request to the STAs participating in those SPs. In this case, STAs are required to report the Received Signal-to-Noise Indicator (RSNI) mea-

sure. Upon reception of these measurements, the PCP/AP decides whether to continue with spatial sharing or not. This maintenance procedure accounts for sudden changes in the wireless propagation environment that might cause the occurrence of interference between time-overlapping SPs.

In Fig. 2.6, the SPSH interference assessment procedure over two BIs is depicted for two pairs of communicating STAs, with an existing scheduled SP in the DTI. As we can see from the assessment phase in the first BI, during the allocated SP between STA A and STA B, STAs C and D performs interference measurements with their antenna pointed to each other as if they were to communicate between each other. Similarly, during the allocated SP between STA C and STA D, STAs A and B performs the same measurements, this time with their antenna pointed to each other. If spatial sharing can be achieved, from the following DTI, the two SPs are allocated concurrently over the same channel. After this, the PCP/AP starts a periodical reporting phase upon which it decides whether to sustain spatial sharing.

2.2 Reinforcement Learning

RL is successfully used in many fields such as robotics as it allows the design of sophisticated and hard to solve algorithms [6]. The main advantage of a RL algorithm is its capability of learning through the interaction with the surrounding environment based on its own experience. Although RL can be used to learn an optimal policy that maximizes a certain future reward under the knowledge of a complete model environment, the most promising RL application relies on the ability of learning the optimal policy without the availability of a complete model of the environment (this means solving the so-called *Control Problem*, using RL terminology). The basic mechanism behind RL is based on trials and errors where the RL agent gradually learns how to maximize the reward by performing different actions and making mistakes. The agent memorizes the consequences associated with any action taken and, in the future, tries to avoid those actions that resulted in low reward or that led to a low rewarding state.

The architecture of RL is depicted in Fig. 2.7 where an agent and the

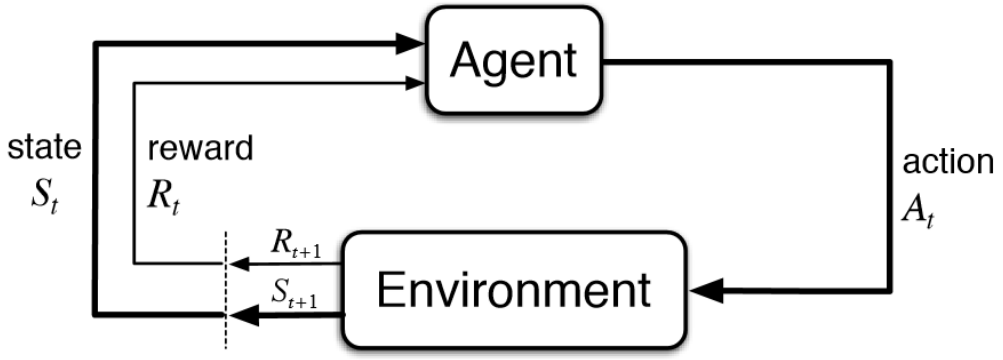

Fig. 2.7: The reinforcement learning concept (source: [7])

environment are considered as two separated entities interacting with each other. Basic RL is modeled as a first-order Markov Decision Process (MDP) characterized by the 4-element tuple: $(S, A_s, P_a(s_t, s_{t+1}), R_a(s_t, s_{t+1}))$. Each element is described as follows:

- S is a finite set of states that represent the environment
- A_t represents a set of actions the agent can undertake against the environment
- $P_a(s_t, s_{t+1})$ is the transition probability from state s_t to state s_{t+1} when the agent chooses action $a \in A_s$
- $R_a(s_t, s_{t+1})$ is a scalar number representing the reward the agent earns when undertaking action $a \in A_s$ in state s_t and reaching the new state s_{t+1}

Basically, the agent's environment is a Markov chain whose transition probabilities depend on both the agent's actions and the environmental consequences associated with an action choise.

In RL, a policy $\pi(s)$ is a function that maps every state to an action. Generally, it represents the action an agent shall take when the environment is in a certain state s; it usually depends on the expected future reward of each action, that the learning agent has estimated during previous iterations. When solving the *Control Problem*, the agent's objective is to learn

the optimal policy $\pi^*(s)$ based on its interaction with the unknown environment to explore. The optimal policy $\pi^*(s)$ maximizes the expected future reward for each state. The future reward at each time step can be written as: $G_t = \sum_{k=0}^{\infty} R_{t+k+1}$. Again, an agent must select the actions that maximize the future rewards, however some actions might have long-term consequences. Given that the maximum reward can be delayed, the agent cannot act in a greedy fashion at all time (i.e. take an action associated with maximum reward at current time), it has to plan ahead by sacrificing immediate reward in order to gain *future* reward. On the other hand, rewards that come sooner are more probable to happen since they are more predictable than the long-term future reward. To account for this trade-off, in RL, we introduce the concept of discounted future reward written as follows: $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$; where $\gamma \in [0, 1]$ is a discount rate as described below:

- The larger the γ, the larger the discount, meaning the learning agent cares more about future reward
- A smaller γ, instead, accounts for bigger discount, meaning the agent cares more about immediate reward

Ultimately, we can formalize the final RL control problem in the following way: the agent's objective is to find the optimal policy $\pi^*(s)$ that maximizes the expected discounted future reward written as: $E_\pi[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}]$.

A task is an instance of a RL problem, we can define two types of tasks namely: episodic and continuous. In the former case, there is a specific starting point and ending point (terminal state) which create an *episode*. Typically in this case, we let the agent perform various episodes through which it finds the optimal solution to the problem. In the latter case, continuous tasks run forever (i.e. there is no terminal state). The agent keeps running until we decide to stop him.

The employment of RL seems very useful for solving networking problems as it can provide a near optimal solution in the case where solving an optimization problem with standard methods is almost impossible or time

consuming. Additionally, RL can provide solution in the case of problems without an accepted one. In the following, we will present and discuss the methods used to find a solution for the RL problem. Moreover, we will see the importance of environment exploration to let the agent interact and learn with the environment.

2.2.1 Exploration and exploitation

The most important strategy to enable the RL agent's learning process is exploration. In fact, if we let the agent take optimal actions with limited or, in the worst case, no knowledge about the surrounding environment, it may never learn the whole system and get stuck on a sub-optimal solution. In order for the agent to converge to an optimal policy as per its objective, it needs to conduct an exploration phase where, by performing random actions, the agent increases its knowledge of the environment. Strictly speaking, during the exploration phase the agent takes random actions, even if they might be non-optimal, and learns from the effects those actions cause to the environment and the reward associated with them. On the other hand, we do not want our agent to take exclusively random actions since, most of the time, they are not the optimal choise. Therefore, in contrast with the exploration phase, there are must be a subsequent exploitation phase where the RL agent puts into practice what it learned during exploration. At this stage, the agent uses its knowledge to perform the optimal action following the computed optimal policy for any state. As a consequence, the balance between exploration and exploitation is one of the central problems in the successful application of RL.

There are several policies proving different balances between exploration and exploitation. As already mentioned, an agent adopting a greedy policy tries to maximize its reward according to its current knowledge, without any exploration. The greedy policy turns out to be optimal only if the agent already posses a perfect knowledge of the environment. Given this, a greedy policy is not suitable for learning how to interact with the environment, since the agent needs to able to converge to an optimal policy with restricted or

no initial knowledge of the environment.

The Softmax policy works by converting each action's expected future reward to a probability distribution over the action space. The action to take is chosen at random according to the resulting distribution, which is given by the following equation:

$$P(a_i) = \frac{e^{\frac{Q_t(a_i)}{\tau}}}{\sum_{k=1}^{|A_s|} e^{\frac{Q_t(a_k)}{\tau}}} \tag{2.1}$$

where the parameter τ can be adjusted to control the exploration rate and $Q_t(a)$ represents the current expected future reward associated with action a. Clearly, actions with high expected future reward have more probability to be chosen, driving the exploration towards those actions that seems more promising.

Among the exploration/exploitation strategies, one very common approach is the ϵ-greedy approach where, basically, the parameter ϵ controls the trade-off between exploration and exploitation as follows: with probability ϵ the agent performs a random action, whereas with probability $1 - \epsilon$ the agent greedily selects the action to take based on the current policy. If ϵ is large, the probability of selecting a random action is higher, so there is more exploration than exploitation. Conversely, if ϵ is small, the probability of exploiting the computed policy to chose an action is higher, thus our agent will exploit what has learned. As a common practice, the value for ϵ is initially kept high since we want our agent to explore the environment, then it is decreased over time to switch from higher exploration to higher exploitation.

Both Softmax and ϵ-greedy policies use time-dependent parameters to adjust the exploration rate over time, reducing exploration as the agent gains more knowledge of the environment and converges towards the optimal policy.

2.2.2 Temporal difference learning

In order to find the solution to the RL (or MDP) problem dynamic programming is used. In particular, there are two ways to enable agent's learning:

the *Monte Carlo Approach* where rewards are collected at the end of the episode and then the expected discounted future reward is evaluated. With this approach, when an episode ends (i.e. the agent reaches the terminal state) the agent looks at the total discounted reward to see how well it performed. However, this approach is not feasible when dealing with continuous tasks RL. Another more promising approach is called *Temporal Difference Learning*. TD Learning enables the agent to learn through every single action it takes, namely at every step of the RL algorithm. Specifically, TD updates the agent's knowledge of the environment at every time-step (action) rather than at the end of every episode. The updating process at each time-step follows this equation:

$$NewEstimate \leftarrow OldEstimate + \alpha[Target - OldEstimate] \tag{2.2}$$

where the value of $Target - OldEstimate$ is defined as the target error and α represents a learning rate, whose value lies between 0 and 1. This equation permits to achieve the *Target* by making updates at every time-step where an action is taken by the agent. Since our objective is to find the optimal map between states and actions, the *Target* of Eq. 2.2 is of course the expected discounted future reward, therefore: $Target = E_\pi[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}]$. In summary, the agent has no knowledge of the states, rewards and transitions so it interacts with the environment, making random or informed actions, to learn new estimates after taking every action, with the ultimate goal of learning the optimal mapping between states and actions (i.e. the optimal policy $\pi^*(s)$).

2.2.3 Q-learning

Q-learning is an off-policy method to solve the *Control Problem* through TD learning. The algorithm builds a table called *Q-table* that stores the expected discounted future rewards (or *Q-values*) for all the possible state-action pairs. The off-policy attribute simply means that the agent implementing Q-learning is not following any policy during the learning process, contrarily, it is trying to improve its policy at each step of the algorithm.

The Q-table is a mapping between states and actions, thus its dimensionality is represented by the product between the states set and the actions set cardinalities as: $|S| \times |A|$. Each Q-table score is called a Q-value and represents the maximum expected future reward that the agent earns if it takes that action at that state following the best policy given. To learn each value of the Q-table, we use the Q-learning algorithm whose objective is learning the action-value function (or Q-function). This function takes two inputs: a state and action, and returns the expected discounted future reward of that action at that state; we can define this mathematically as follows:

$$Q(S_t, A_t) = E[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t, A_t] \tag{2.3}$$

The Q-learning algorithm uses TD learning as anticipated in Eq. 2.2 to estimate the Q-function at each time-step according to an updating formula, that will eventually converge to the true expected future reward [8], defined by the following equation:

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \max_A Q(S_{t+1}, A) - Q(S_t, A_t)] \tag{2.4}$$

where the targeted expected future reward is represent by the term: $R_{t+1} + \gamma \max_A Q(S_{t+1}, A)$ where, in turn, R_{t+1} is the reward for taking that action at that state and the term $\max_A Q(S_{t+1}, A)$ is the maximum expected future reward, over all possible actions, given the new state S_{t+1} discounted by a factor γ. The learning rate α is a parameter that adjusts the updating speed and it is gradually decreased during the learning process to allow better convergence.

The on-policy counterpart to Q-learning is the State-Action-Reward-State-Action (SARSA) algorithm. It is an on-policy algorithm because the next action to take based on the new state S_{t+1} is imposed by a known policy. Therefore the *max* term of Eq. 2.4 is replaced by the Q-value associated with the new state-action pair as follows:

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha[R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)] \tag{2.5}$$

The main difference between Q-learning and SARSA is that the former does not follow a pre-defined policy to find the next action A_{t+1} but rather it chooses the action in a greedy fashion.

Algorithm 2.1 Q-learning (off-policy TD) algorithm for estimating π^*

initialize $Q(s, a)$, for all $a \in A$, $s \in S$, arbitrarily
repeat (for each episode)
  Initialize S
  repeat (for each step in the episode)
    Choose A from S using policy derived from Q (ϵ-greedy approach)
    Take action A and observe reward R and new state $S^{'}$
    Update $Q(S, A) = Q(S, A) + \alpha[R + \gamma \max_a Q(S^{'}, a) - Q(S, A)]$
    $S \leftarrow S^{'}$
  until End of episode
until Episodes completed

Finally, the Q-learning algorithm pseudo-code is reported in Algorithm 2.1 , and it will be used in the reminder of this work as the basis for the development of our RL-based scheduler. The pseudo-code can be translated into plain English steps as follows:

1. Initialize the Q-values of the Q-table for every state-action pair
2. Observe the current state S
3. Chose an action A based on the action selection policy (ϵ-greedy, soft-max, etc), meaning that during the exploration phase we select a random action to gain knowledge of the environment. While, after we find the optimal policy π^*, we exploit the Q-table to select the action that yields the maximum expected future reward given the current state S
4. Take the action A and observe the associated reward R, as well as, the new state $S^{'}$
5. Update the Q-value for state S and action A using the observed reward R and the maximum reward for the next state $S^{'}$, according to Eq. 2.4
6. Set the state S to the new state, and repeat the process until a terminal state is reached or the task is finished

2.3 Related Work

In the literature, the majority of works are focused on the introduction of the standard and its features like in [2] and [4]. Some other works concentrate their efforts in the beamforming for IEEE 802.11ad, which is one key concept to enable efficient communications in the 60 GHz band. In particular, the authors of [9] proposed an efficient codebook-based MIMO beamforming training scheme for estimating the best antenna configuration in mm-wave wireless local area networks. Their scheme, based on the Discrete Fourier Transform (DFT), has been proved to outperform the standard IEEE 802.11ad beamforming procedure in terms of beam gain and training time. Similarly, in [10], the authors proposed a hybrid BF scheme compatible with IEEE 802.11ad standard to overcome the fact that 11ad serves one user at a time. They argued that there is an optimal number of users that a hybrid BF-enabled mm-wave communication systems should support.

In [11], the authors studied the impact of MAC layer parameters on the buffer size at the transport layer, using a real IEEE 802.11ad testbed. They found out that using large buffer sizes with TCP is harmful despite the multi-Gbps data rates. However, in this work, the authors did not deal with the design of MAC layer scheduling techniques, since they used the proprietary algorithms implemented in the devices used as testbed.

There are some works in the literature that provide an analytical model for the hybrid MAC of IEEE 802.11ad. For example, the authors of [12] introduced an analytical model for the access during SPs. They analyzed the delay performance of the system with varying arrival rate of SP packets. They also provide a possible CBAP and SP allocation strategy to achieve a trade-off between SP delay and CBAP throughput. Similarly, in [13], the authors introduced an analytical model for the performance analysis of IEEE 802.11ad using a 3D Markov chain. They showed how the number of antenna sectors as a high impact on the network throughput, and how the delay at MAC layer is affected by the duration of the contention-based period. Despite being interesting and inspiring, these analytical approaches do not offer the enhanced and robust results of an end-to-end simulation

using ns-3. Moreover, they make several assumptions regarding the model design that can lead to misinterpreted results. With this respect the ns-3 module, which we are going to introduce later in this thesis, provides a more accurate IEEE 802.11ad implementation. In addition, there are no works in the literature that investigate and, possibly, develop the spatial sharing feature described in the IEEE 802.11ad amendment.

Concerning RL, there are several efforts, in the literature, that successfully apply RL techniques in the networking/wireless field; especially in the context of cellular network traffic scheduling [14] and for improving the performance of TCP congestion control [15]. Nevertheless, there are no works that tries to apply RL for the optimization of a problem related to IEEE 802.11ad.

To the best of my knowledge, this is the first work that implements and develops, in ns-3, different resource scheduling algorithms for IEEE 802.11ad: one of which employs and extends the spatial sharing mechanism to enhance the overall network spectral efficiency, while another approach exploits RL to optimize the allocation of SPs in order to increase the performance during hybrid MAC channel access.

Chapter 3

Data Transmission Interval Scheduling

In the previous chapter we have seen all the theoretical ingredients that are going to be used in this work. In this chapter, we will present and discuss the implementation of three resource scheduling algorithms. Therefore, the contributions advanced in this work can be split into three scheduler development directions, namely: the *Default scheduler*, the *Spatial sharing scheduler* and the *RL-based Default scheduler*.

Before delving into the details of these three approaches, we will introduce the *ns-3 802.11ad module* where all schedulers have been developed and their performance has been assessed. Moreover, we will introduce *ns-3 gym*: an important tool which enables to interface the most famous open-source RL toolkit (i.e. *OpenAI Gym* [16]) with the ns-3 environment.

3.1 NS-3 modules overview

In this section, we will briefly explore the functionalities of the IEEE 802.11ad ns-3 module, in particular we will discuss the most important features of the standard that are implemented in this module. In addition, we will rise the need for a scheduling algorithm to enhance the capabilities of the available module. Then, we will present the ns-3 gym framework, as a powerful mean in

order to develop RL-based algorithms fully integrable with the ns-3 network simulator.

3.1.1 NS-3 802.11ad model

The ns-3 802.11ad module, described in [17]-[18] and publicly available at [19], implements the IEEE 802.11ad standard in the network simulator ns-3, covering every layer of the protocol stack with the implementation of a channel model at 60 GHz. It provides support to many essential 802.11ad features such as: beamforming training and steering, hybrid MAC, relay operations and fast session transfer. This model is a direct evolution of the classic ns-3 Wi-Fi module [20], which is well suited for the simulation of legacy wireless technologies that operate with omni-directional antenna patterns exploiting the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) scheme. In particular, the ns-3 Wi-Fi module offers an accurate implementation of the MAC layer with the availability of the following features: Ready To Send (RTS), Clear to Send (CTS), Normal Acknowledgment (ACK) and Block ACK, implementation of the EDCA function and MSDU/MPDU aggregation/de-aggregation. The authors of [17] extended this module by adapting it to the IEEE 802.11ad characteristics. In particular, they developed hybrid channel access support through a SP-class that co-operates with the usual CSMA/CA access scheme at MAC layer. They provide functionalities such as: DMG beaconing, SLS and BRP beamforming protocols and support to the 11ad BI structure, inclusive of the ATI. To deal with directional communications, they also introduced a directional antenna model which divides the 2D plane into a tunable number of virtual sectors with equal apertures and coverage range. It is also possible to define parameters of the radiation such as: main lobe gain, side lobes gain and the omni-directional gain. Finally, they adapted the PHY layer for communications in the 60 GHz band with suitable BER-SINR curves for computing the Packet Error Rate (PER) according to the selected MCS.

In the latest version available [18], the authors provide support to dynamic channel access by exploiting the described polling mechanism. More-

over, they provide all the placeholders for the integration of SPSH as the simple mechanism outlined by the standard and discussed in 2.1.6. However, the spatial sharing features they provide are limited as the entire mechanism is not well integrated with the module. In addition, even if they supports both SP and CBAP channel access schemes, they do not provide any scheduler implementation that organizes channel access during the DTI. These limitations motivate the work conducted in this project aimed at providing an enhanced version of the currently available module.

3.1.2 NS-3 gym

Given that ns-3 represents the de-facto framework used by the industry and academia to conduct end-to-end simulations of many networking and wireless scenarios, the authors of [21] developed the ns-3 gym toolkit to fully enable the application of RL techniques with the network simulator ns-3. Also, the authors were motivated by the fact that one of the major trend in nowadays network research regards the use of machine learning techniques like for example RL. This open-source package was designed to facilitate and shorten the time required for prototyping new RL-based networking solutions. It allows for fast-prototyping, scalability and low overhead for the adaptation of existing ns-3 scripts to their architecture.

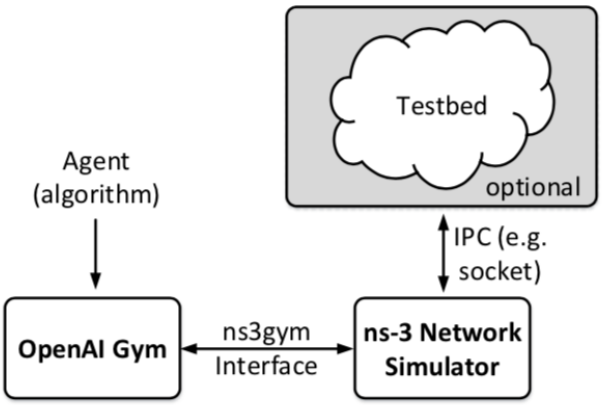

Fig. 3.1: Ns-3-gym framework architecture (source: [21])

As depicted in Fig. 3.1, the ns-3-gym framework interfaces two main blocks: OpenAI Gym and ns-3. According to the RL terminology, the Gym framework implements the agent while the network simulator acts as the environment with which the agent interacts. The interface provided by the authors takes care for the management of the ns-3 simulation process and delivers state, action and reward information to the agent, following the OpenAI format. This split between agent and environment (i.e. ns-3) allows for a separate implementation of the agent, which can be written in *Python* and can be implemented using the most advance machine learning libraries (such as *Tensorflow*, *Keras*, etc).

3.2 Design assumptions

In this section, we will state the design assumptions that holds for all the developed approaches. In particular, we solely allow for the allocation of contiguous time block, whether they are SPs or CBAPs, for the communication between pairs of stations during the DTI. This means that two specific stations will have their dedicated slot of time once every BI. As a consequence of this first assumption, when a STA sends an ADDTS request, the parameter *Allocation Period* is set to 1. Another assumption is that STAs, through an ADDTS request, can only require a SP type of allocation for the exclusively communication with another node; thus the parameter *Allocation Type* is always set to SP for any STA's request, meaning it is asking the allocation of a contention-free time period. Furthermore, we consider the allocation of Asynchronous (ASYN) and Isochronous (ISOC) traffic, where the requesting STA has the possibility of specifying an associated User Priority (UP), represent by a number between $0-7$, from lowest to highest user traffic priority. Recalling what we have seen so far, the *Minimum Allocation* and *Maximum Allocation* fields of the DMG TSPEC element in an ADDTS request represent the most important parameters a STA specifies for its SP allocation request. The computation of the *Minimum Allocation* duration takes into account the transmitting station's application layer data rate as well as the beacon interval duration and the data rate supported by the network accord-

ing to the selected MCS. In this work, the last SC PHY mode, represent by MCS 12, is considered. The coarse calculation of the *Minimum Allocation* duration does not consider the header overhead associated with each layer of the protocol stack nor it considers a possible estimate of the channel quality between those stations exchanging data during the SP. In addition, it does not consider the possible re-transmission of MAC-layer packets thanks to the ACK mechanism. This is why the parameter *Minimum Allocation* of the ADDTS request represents a quite conservative allocation duration that might not be enough to meet traffic requirements such as: the reliable exchange of information and the delay performance associated with the communication.

On the other hand, the parameter *Maximum Allocation* is set to a value that provides the theoretical certainty about the successful transmission of data between the stations involved during the SP. In fact, this value is set to the double of the *Minimum Allocation* value. However, even allocating according this time value can be sub-optimal: if a station terminates its transmission early with respect to the end of the dedicated SP, then the remaining time is wasted as it cannot be used to serve other traffic in the network.

3.3 Default scheduler

The Default Scheduler (DES) represents a first approach aimed at developing a scheduling policy in IEEE 802.11ad. Driven by the objective of keeping this first implementation intentionally simple, this scheme organizes the channel access during the DTI as follows: until at least one ADDTS Request frame from any associated STA is received by the PCP/AP, the entire DTI is allocated as CBAP with Source and Destination AIDs set to the broadcast value, so that every station in the network can compete in order to acquire channel access by winning an instance of the EDCA function. As soon as the first ADDTS request is received by the PCP/AP, the initial portion of DTI, with duration specified by the parameter *FirstCBAPDuration*, is always allocated as CBAP with Source and Destination AIDs set to broadcast. This allocation is necessary to always preserve backward compatibility

with previous Wi-Fi technologies; moreover, the exchange of ADDTS Request and Response frames, as well as other Management frames, follows a contention-based approach where STAs compete to send their requests to the PCP/AP, and vice versa. The PCP/AP accumulates those ADDTS requests received during the current DTI and analyzes them at end of it, so that new allocations or modifications to the existing ones are announced in the following BTI through an ESE inside the periodic DMG Beacon frames. The employed admission policy acts as described in the following. Initially, all received requests from STAs, together with those the PCP/AP itself desires to allocate, are ordered considering primarily their *Allocation Format* (by design choice ASYN traffic has higher priority than ISOC traffic). Given its nature, ASYN traffic is requested for non-recurrent type of communication, so we aim at satisfying first this kind of traffic knowing that it is not going to last in time as an ISOC type of allocation. Secondly, requests are ordered considering their associated UP; clearly, requests with higher value are examined before those with less priority. The third discriminating parameter is *Maximum Allocation* duration: requests with lower value for this field have higher priority. In the successive step, the remaining DTI time is computed as the difference between DTI's duration and the total duration of existing allocations. Then, each allocation request in the ordered list is accepted if the *Maximum Allocation* time specified in the request does not exceed the available remaining time. The remaining DTI time is updated once a request is accepted and allocated. This policy leads to two possible outcomes: if the request is accepted, a SP consisting of one contiguous block with duration *Maximum Allocation* is reserved for the communication between the requesting pair of stations; after that, an ADDTS Response frame with status code *SUCCESS* is sent to both nodes to indicate that the request has been accepted. On the other hand, if the request is not accepted by the PCP/AP due to a lack of remaining DTI time, an ADDTS Response with status code *REJECTED FOR DELAY PERIOD* is sent to the station which transmitted the request. This particular response carries a Traffic Stream (TS) Delay element that indicates in multiples of a TU (1 TU = 1024 μs) the elapsed time starting from the reception of this response, after which the station can

retry to reserve the rejected allocation by sending a new ADDTS request to the PCP/AP. Lastly, after all requests received in the current DTI are processed, any remaining DTI time is allocated as CBAP with Source and Destination AIDs set to broadcast. Anytime a new set of ADDTS requests received during the DTI is evaluated, the last CBAP allocation, if present, is removed for the calculation of the DTI remaining time.

The DES scheme allows only to modify the duration of existing allocations. In particular, whenever a received ADDTS request matches an existing allocation (i.e. it has the same tuple: [Allocation ID, Source AID, Destination AID], it is interpreted as a modification request. If the *Maximum Allocation* duration of the modification request is greater than the actual allocation duration, the request is accepted if their difference is available as remaining DTI time.

A flow chart describing the operations of the DES algorithm is depicted in Fig. 3.2.

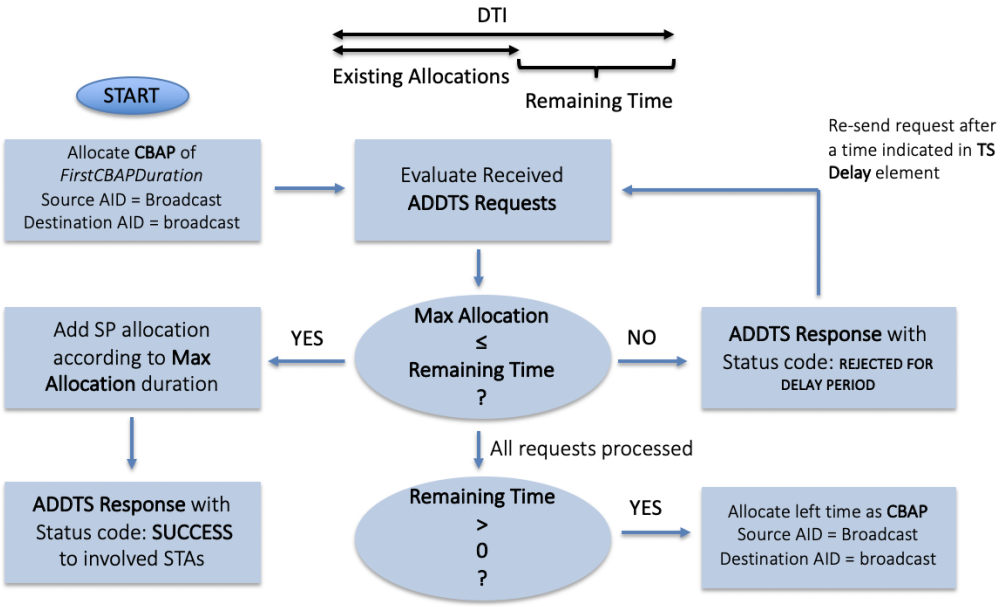

Fig. 3.2: Flowchart of the operations conducted by the DES once a set of ADDTS requests is received by the PCP/AP

3.4 Spatial sharing scheduler

The Spatial Sharing Scheduler (SPS) extends the SPSH mechanism introduced by the standard thanks to the application of well-known graph theory concepts to the Interference Graph (IG). Therefore, before delving into the details of this scheduling algorithm, we are going the discuss the properties and how is constructed the IG.

3.4.1 Interference graph

In order to properly understand the mechanism behind the SPS, it is necessary to introduce the concept of interference graph. The IG is an un-directed graph that represents the relationship in terms of interference power among the SP allocations in the network. Specifically, a vertex of the graph corresponds to a specific allocation, and therefore it is identified by the tuple: [Allocation ID, Source AID, Destination AID]. First, the graph is constructed with just vertexes representing all the SP allocations granted by the PCP/AP. Second, an edge between a pair of vertexes is drawn when the ANIPI measure, reported by those STAs involved in the two allocations represented by each vertex in the pair, goes beyond a predefined threshold for at least one of the STAs. In the event that two allocations share one node or both, an edge is added by default, without requiring any measurement report. The updating process of the IG is based on an enhanced version of the spatial sharing mechanism specified by the standard. For the sake of clarity, we name this algorithm as Report Interference Measure Procedure (RIMP).

The behavior of this procedure is described according to the following example. Given the network scenario of Fig. 3.3 as reference, suppose that STA A with STA B, as well as STA C with STA D and STA E with STA F, have been granted a dedicated SP for their communication. The RIMP algorithm is initiated by the PCP/AP which sends to each STA a Directional Channel Quality Request indicating the following parameters: *Measurement Method* (ANIPI), *Measurement Start Time* and *Measurement Duration*. In this scenario, a request received by STA A and STA B asks them to report

the two ANIPI values, with antenna pointed to each other, measured during the SP reserved for communications between STA C and STA D, as well as the one between STA E and STA F; vice versa in the case of requests received by STAs C-D and STAs E-F. After completion of each measurement period, a STA sends back to the PCP/AP a Directional Channel Quality Response which carries the results for each measurement. Upon reception of all reports, the PCP/AP eventually adds edges to the IG as described earlier. Given that, from Fig. 3.3, the communication between C-D and E-F spatially overlaps, the resulting IG would have been composed by the only edge between C-D and E-F, symbolizing considerable interference if their SPs were to be allocated concurrently over time. Finally, this procedure can be extended to any number of SP allocations in the network.

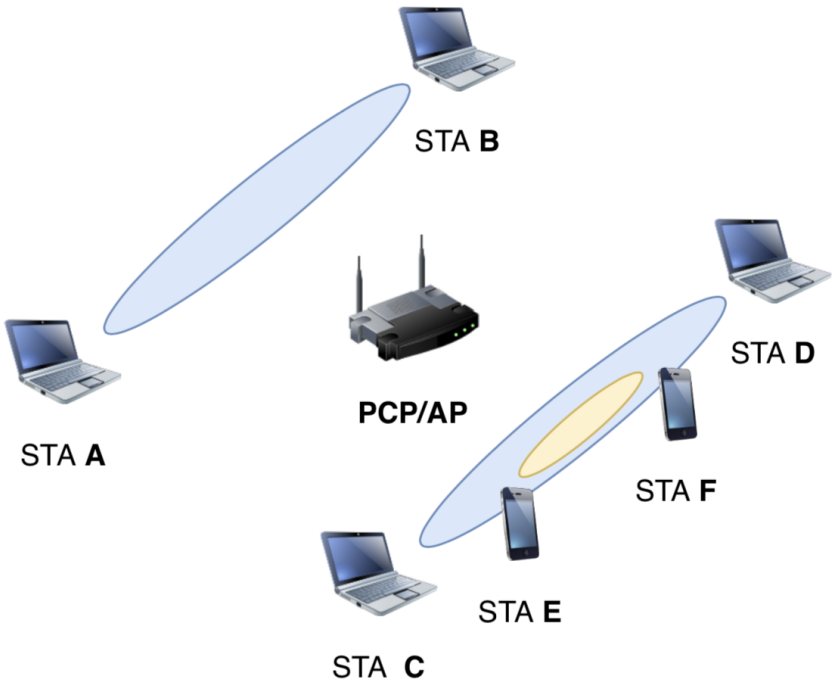

Fig. 3.3: Example topology for the RIMP. In this figure, the elliptical colored regions represent the directional communication (i.e. antenna's beam direction) between two STAs during their dedicated SP, where their antenna are directed toward each other

3.4.2 SPS operations

SPS's operations are described in the following. This scheme shares some similarities with the DES scheme. As such, until no ADDTS requests are received, the entire DTI is allocated as CBAP. In addition, when the first request is received by the PCP/AP, the initial portion of DTI, with duration *FirstCBAPDuration*, is allocated as CBAP, similarly to the DES algorithm. However, the duration of this first contention-based period is higher than the one for the DES. This is due to the fact that all Directional Channel Quality Request/Report frames follow EDCA transmission rules, leading to an increased amount of contention-based traffic to be handled during this period. The PCP/AP can only receive ADDTS requests during the latter period, therefore all accumulated requests are elaborated at its end. The list of received requests is ordered according to the same procedure described for the previous scheme.

Then, the SPS performs the following operations. First, all existing allocations, if any, are restored to their original starting time and duration (existing allocations might have been affected by previous SPS operations). Subsequently, the remaining DTI time is computed considering the total duration of restored allocations and a request is accepted if the *Minimum Allocation* time required does not exceed the available remaining time. As earlier, this condition creates two possible outcomes that this scheme handles as the previous one. After all requests are processed, in the next step, a new measurement procedure (RIMP) is started in order to create a IG which reflects the new set of allocations. Measurements are conducted in the successive DTI, therefore all reports should be available at the PCP/AP in 2 BIs. If an expected measurement report is not received, interference is assumed, therefore an edge is added between the involved allocations. Once all reports have been received and processed, the SPS scheme applies its peculiar characteristic: from the IG, the set of connected components is computed. Each connected component represents a set of SP allocations that are not interfering with each other; therefore, it is possible to consider the entire DTI duration (excluded the first CBAP period) for the allocation of the dedicated

SPs belonging to a component. Considering again Fig. 3.3, clearly one connected component is represented by the SP allocation between STA A and STA B, while the other one is composed by SP allocations between STAs C-D and STAs E-F, because eventually their concurrent transmission causes interference to each other. Consequently, STAs A-B can communicate concurrently over the same channel with the set [STAs C-D, STAs E-F]. Most importantly, the duration of each allocation can be augmented in order to guarantee higher throughput and lower delay per allocation, together with higher overall spectral efficiency. The extra time given to each allocation is computed based on the difference between the DTI duration, excluded the first CBAP period, and the duration of the allocations in the considered connected component.

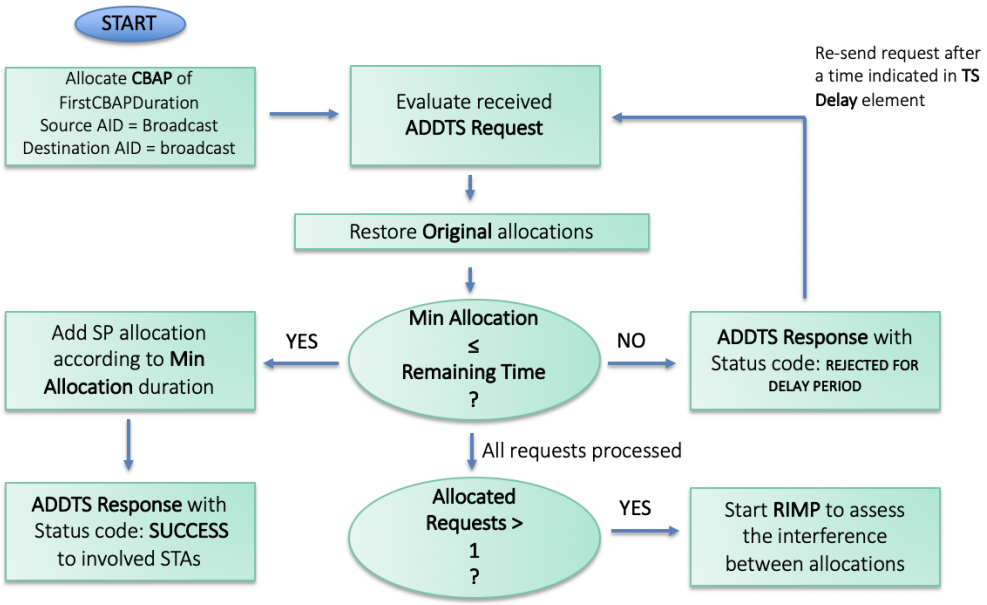

Fig. 3.4: Flowchart of the operations conducted by the SPS once a set of ADDTS requests is received by the PCP/AP

Once concurrent transmission is achieved, it is necessary to keep track of sudden changes in network interference due, for example, to user mobility. This is accomplished by periodically sending a different Directional Channel Quality Request to each stations. In this case, STAs are requested to report

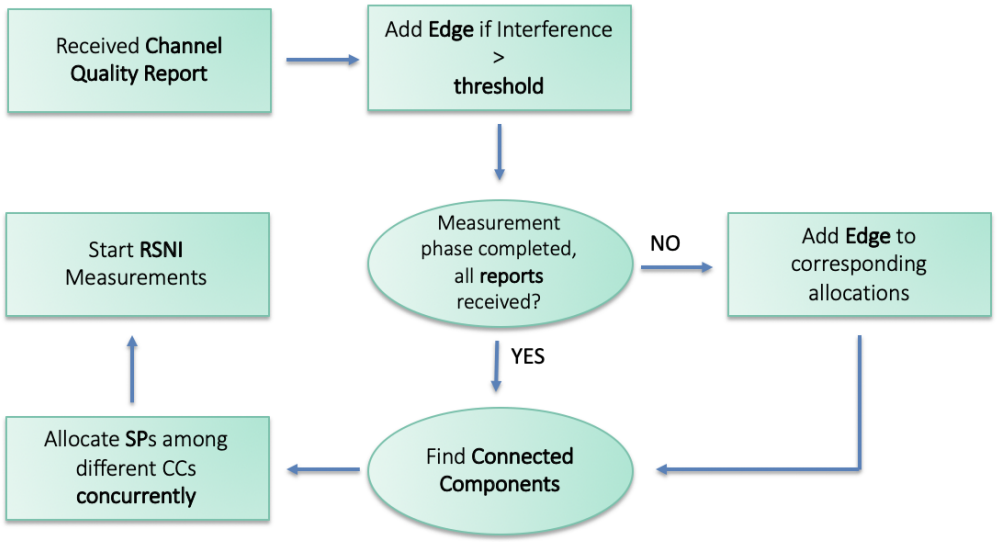

Fig. 3.5: Flowchart of the concurrent SPs allocation thanks to the RIMP

the RSNI measure to the PCP/AP. Once all reports are collected, if the Signal to Interference plus Noise Ratio (SINR) for one particular STA goes below a given threshold, a new RIMP routine is started in order to compute the new IG, with updated network's interference information. Then, the same steps described earlier are performed.

The operations of the SPS algorithm, together with the concurrent allocation of SPs through the RIMP, are depicted in Figs. 3.4 and 3.5.

3.5 RL-based Default scheduler

One drawback concerning the allocation of SPs following the DES algorithm regards the possible waste of DTI time when the allocated SP time is based on the *Maximum Allocation* field specified in the transmitting STA's ADDTS request. On the other hand, even the allocation considering the *Minimum Allocation* parameter could display degraded performance in terms of throughput and latency. In order for the PCP/AP to allocate each SP's time length in the most optimal way, namely: avoiding waste of DTI time while,

at the same time, guarantying the same performance offered when allocating considering *Maximum Allocation*, a variation of the DES scheme based on RL has been introduced in this work. In particular, we exploit RL to learn the optimal length for each SP allocation based on the network load at the transmitting station of the allocated pair. The network load at each node is represented by its queue size at MAC layer. As seen in section 2.2, RL represents a promising technique to solve problems that are hard to engineer. In RL, an agent interacts with the environment (i.e. it observes *states* and performs *actions*); when an action is undertaken, the environment moves to a *new state* as a consequence of the action taken and the agent earns a reward associated with that action. The ultimate goal of a RL agent is to learn, from its experience with the environment, the best action to take in particular state (i.e. the action that yields the maximum expected discounted future reward). As a consequence, when formulating a RL problem we need to define several important elements as follows. First, the *observation/s* represents what the agent is watching from the environment, in other words, the state an agent is observing. Second, it is important to define the set of *actions* an agent can exploit when interacting with the environment. Third, the measure of the reward produced by taking a certain action in a particular state represents an important parameter to define, since it affects the learning process of the agent. Finally, we need to define what is called the *gameover*: a condition which causes the end of a RL episode.

The RL-based DES follows the operations of the DES scheme. The proposed RL mapping for the problem addressed by the RL-based DES is the following:

- **Observation:** queue length, in number of packets, for the transmitting station of each SP allocated by the PCP/AP in the current DTI.

- **Actions:** the actions' set is composed by the integers between 1 and 10. An integer in this set represents an increment in the SP duration starting from the value associated with the *Minimum Allocation* parameter of a requested SP. In particular, every time an action is taken

the SP duration is changed according to the following formula:

$$NewSPduration = MinAllocation \times \left(1 + \frac{Action}{10}\right) \tag{3.1}$$

where if the action taken is 10, then the new SP duration equals the *Maximum Allocation* duration; whereas, if the action taken is 1 the SP's new duration is close to the duration specified by the parameter *Minimum Allocation*.

- **Reward:** number of packets received by each receiving station in the SP allocation, during the DTI where the action taken manifests its effect. However, the actual reward depends on the number of received packets and the action that was taken and led to that amount of received packets. This artifact was introduced to take into account the following case: if, for example, taking actions 4 and 6 leads to the same amount of received packets, then we want the reward associated with action 6 to be less than the one associated with action 4. Given that with action 4 we allocate a shorter SP, saving DTI time while preserving throughput (i.e. the number of packets transmitted), we want the reward associated with this action to be higher. Thus, the new reward is computed thanks to the following expression:

$$NewReward = \#ReceivedPackets \times \left(1 - \frac{Action}{100}\right) \tag{3.2}$$

- **Gameover:** end of ns-3 simulation.

Clearly, the RL agent runs on the PCP/AP, as a matter of fact, the PCP/AP needs to collect observations (queue sizes of transmitting STAs) and rewards (number of packets received at receiving STAs) from the stations in the network. In order to enable this, the RL-based DES employs the exchange of Vendor Specific Request/Report frames during the ATI period between the PCP/AP and the STAs involved.

The standard foresees the availability of Vendor Specific Request/Report frames, that each manufacturer can shape according to its own preferences.

In this specific case, with a Vendor Specific request the PCP/AP asks to a STA to report one between the following two values: number of received packets in the previous DTI and the current queue size at MAC layer. Whereas, with a Vendor Specific report, a STA reports the information required by the PCP/AP.

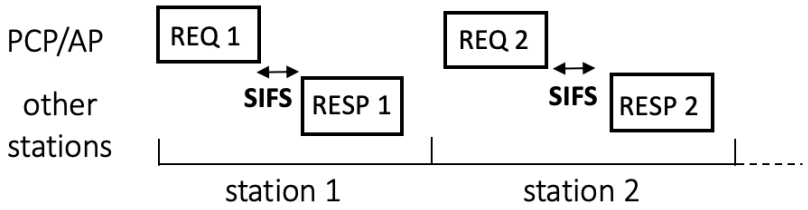

Fig. 3.6: Request/response mechanism during the ATI period

Through a request/response mechanism during the ATI, each STA, between those participating in an allocated SP, is questioned with an individually addressed request which is followed by the STA's response. If a STA is the transmitting node of the allocated pair, then it reports its queue size at MAC layer; whereas, if a STA is the receiving node of the pair, it sends back the number of packets received. The request/report mechanism between STAs and PCP/AP during the ATI is depicted in Fig. 3.6. Given that STAs already performed beamforming training before the ATI, then the exchange of frames exploits higher MCS other than the Control MCS; thus the communication is more efficient.

The RL agent interacts with the environment during the ATI. In particular, once all expected Vendor Specific reports are received by the PCP/AP, the agent gets a reward (R) associated with the action (A) undertaken during the previous DTI (according to Eq. 3.2), and a new state ($S^{'}$) which represents the consequence of taking that action in the previous state (S). Based on this information, the Q-tables (one for each SP allocation) are updated following the Q-learning algorithm as per Eq. 2.4. After the Q-tables are updated, the agent takes a new action (ϵ-greedy policy) which modifies the SP duration of each allocation. At this point, in order to make the entire operations more reactive, instead of waiting the next BTI to announce the new allocations, the PCP/AP broadcasts, during the same ATI, an An-

nouncement frame containing an ESE with the updated allocations' duration. Thanks to this procedure, modifications to existing SP allocations, in the form of increased or decreased allocated time, take their effect from the immediately subsequent DTI; thus in the successive ATI, we measure the reward associated with the exchange of data during the previous DTI, and so on.

Finally, the entire states' space, represented by the number of packets queued at MAC layer, is divided in intervals with a certain granularity considering the maximum number of packets that can be queued. So, to every queue size value is assigned a particular interval. This enables to reduce the dimensionality of the states' space (i.e. the dimension of the Q-table) and allows for faster exploration of the environment.

Chapter 4

Simulations and Results

The performance of all three scheduling algorithms has been assessed through extensive ns-3 simulation campaigns under different meaningful scenarios. Given its simplicity and low complexity the DES is employed as performance baseline for the comparison with more advanced approaches such as: the SPS algorithm and the RL-based DES. As a consequence, the assessment of performance has been conducted following two main directions. First, we made a comparison between DES and SPS in a specific scenario through which it is possible to appreciate the performance enhancements achievable thanks to spatial sharing, as well as, some possible drawbacks due to the RIMP in terms of reactiveness and associated overhead. Second, we exclusively compared the DES and RL-based DES, adopting a realistic IEEE simulation scenario, to observe the performance of the RL-based variation we introduced later in the previous chapter.

In the remainder of this chapter, we are going to present and discuss the results obtained through extensive ns-3 simulations as per the two evaluation directions just described. In particular, in the first section, we will highlight the performance comparison between DES and SPS policies. In the last section, after an overview of the IEEE scenario developed in ns-3, a performance comparison between DES and RL-based DES is offered. Thus, showing how the use of RL-based techniques can improve the behavior of a simple resource scheduling scheme for IEEE 802.11ad.

4.1 Spatial sharing vs Default

In this section, we will illustrate and discuss the performance of the SPS against the DES. First, we introduce the ns-3 scenario we developed for this comparison while, later, the results are presented and discussed.

4.1.1 Simulation scenario and settings

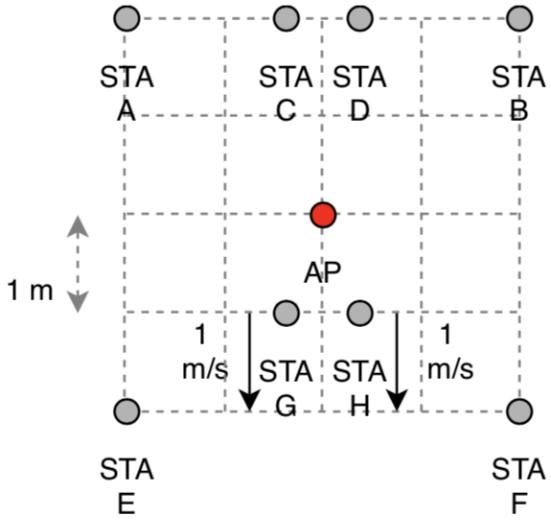

Fig. 4.1: Ns-3 simulation scenario for the comparison between DES and SPS. The 2D positional grid reported in the figure is composed by dashed squares of side 1 m. STAs are represented by gray dots, whereas the PCP/AP is the red dot. For reference the PCP/AP is located at position (0,0) on the grid

For this performance evaluation we decided to adopt a scenario where the potential benefits and possible drawbacks of the developed SPS scheme can be analyzed. With this aim, we introduce the scenario depicted in Fig. 4.1. Accordingly, we consider a total of 4 dedicated allocations, namely the SPs between: STAs A and B, STAs C and D, STAs E and F, and STAs G and H. The transmitting stations for each SP allocation are STAs: A, C, E and G; while the receiving nodes are STAs: B, D, F, H. For the sake of clarity, in the remainder of this section, we refer to each SP allocation respectively as: SP1, SP2, SP3, SP4; as also reported at the bottom of Table 4.1. All

stations are at fixed position, as per the scenario of Fig. 4.1, for the entire simulation duration, except for STAs G and H that behave as follows. At 3.5 seconds after the simulation starts, they begin moving, from their initial position of Fig. 4.1, with constant speed of 1 m/s following the direction depicted by the arrows in the figure; then, at 4.5 seconds after the start of simulation, they stop and last, at the position reached, until the end of the simulation (after stopping, STAs G and H are horizontally aligned with STAs E and F). The movement of STAs G and H has been introduced to account for the following fact: initially, SP3 does not interfere with SP4, so each one of them represents a distinct connected component of the IG. On the other side, SP1 interferes with SP2 if allocated concurrently, therefore, they together represent another distinct connected component, leading to a total of 3 connected components. While STAs G and H move towards the LOS between STAs E and F, the concurrent allocation of SP3 and SP4, as a result of the initial computation of the IG, causes interference between both of them. This interference reaches its maximum value when, at 4.5 s, the 4 STAs are located at the same horizontal coordinate. In this case, the reception of below-threshold RSNI reports by the PCP/AP should trigger a new RIMP routine, that maps the interference changes occurred in the network and reschedule SPs concurrently according to the updated IG.

Parameter	Value	Parameter	Value
802.11ad carrier frequency	60.48 GHz	TX power P_{TX}	10 dBm
802.11ad bandwidth	2.16 GHz	RX noise figure	10 dB
CCA threshold	-79 dBm	Energy detection threshold	-76 dBm
PHY MCS data mode	12	Beacon interval duration	50 TU
BTI duration	500 μs	Slots per A-BFT	8
Sector sweep frames per slot	8	ATI duration	0 μs
MAC layer queue size	5000 packets	MSDU aggregation size	7935 bytes
First CBAP duration	2500 μs	Interference threshold (ANIPI)	-40 dBm
SINR threshold (RSNI)	5 dB	RSNI report periodicity	2 BIs
Application data rate	650 Mbps	App layer payload size	1472 bytes
App start time	3.0 s	Default SP duration	7500 μs
STAs speed	1 m/s	Simulation duration	8 s
SP allocation STAs A-B	SP1	SP allocation STAs C-D	SP2
SP allocation STAs E-F	SP3	SP allocation STAs G-H	SP4

Table 4.1: Ns-3 first simulation campaign settings

The simulation starts with the association of each STA with the PCP/AP. Then, 4 non pseudo-static SPs are allocated in order to perform beamforming training for each pair of STAs; this is needed for the successive communication during each SP. In order to simulate a high-rate data transfer, at 3 seconds after the start of simulation, each transmitting STA initiates a constant User Datagram Protocol (UDP) flow of app layer packets according to a data rate of 650 Mbps (same data rate for each transmitting node) and sends an ADDTS request to the PCP/AP with the following main characteristics: *Allocation Format* = ISOC, *Maximum Allocation* = *Minimum Allocation* = 7500 μs. Importantly, we assume that each SP request demands the allocation of 7500 μs; in turn, this time slot recurring every BI is not enough to satisfy an application data rate of 650 Mbps. However, we decided to force this value for the duration of each allocated SP to highlight the performance enhancements when achieving spatial sharing through the SPS algorithm.

Based on the choice of scheduler during the simulation configuration, the PCP/AP organizes the DTI accordingly. Moreover, in the case of SPS scheduling, the PCP/AP starts a RIMP to compute the IG and allocates SPs concurrently; later, it starts a periodical RSNI reporting phase to keep track of the network's interference and, eventually, initiate a new RIMP. The most important ns-3 simulation parameters are reported in Table 4.1.

4.1.2 Simulation results

In the following, we analyze the results in terms of throughput and delay for the DES vs the SPS schedulers. The depicted throughput, in all figures, is the end-to-end throughput at application layer. The results reported have been averaged considering a simulation campaign over 40 independent runs in ns-3.

Fig. 4.2 depicts the throughput vs simulation time for the allocation between STAs A and B (i.e. SP1) when exploiting the DES or SPS schemes. At the start, once spatial sharing is achieved after the RIMP, we can see that the SPS scheme is able to guarantee the expected throughput performance required by the application layer data rate. On the other hand, the DES scheme

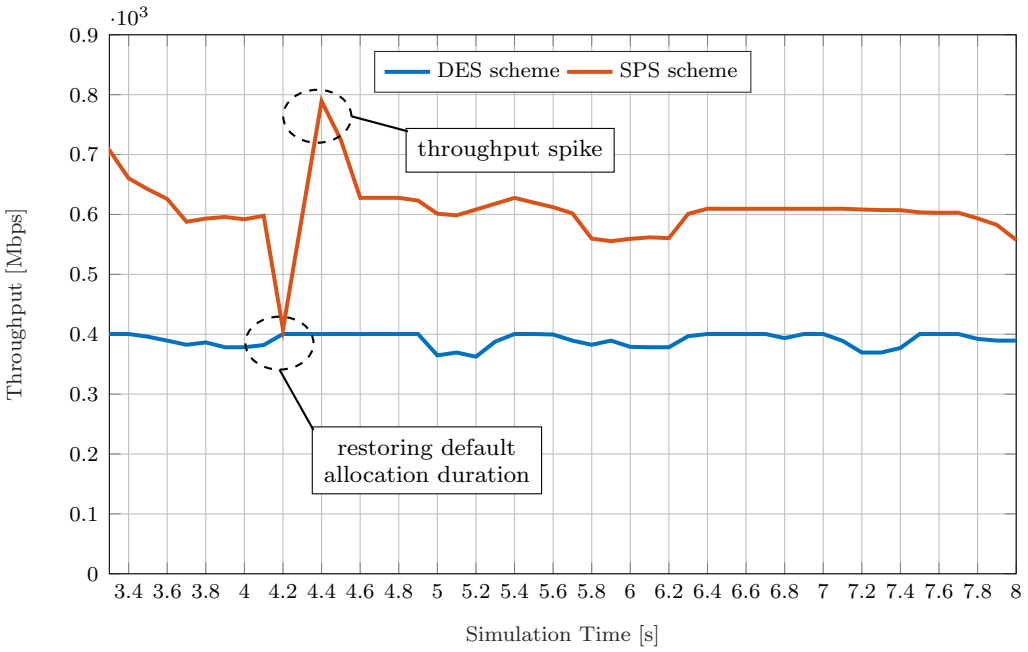

Fig. 4.2: Throughput vs simulation time for the allocation between STAs A and B: SP1

with predefined allocation duration does not meet the required performance, as expected; and, given the always fixed allotted time, the throughput over the simulation duration is almost constant except for some fluctuations due to the loss of packets and varying channel conditions. Furthermore, considering the SPS policy, we can clearly see from the throughput at 4.2 seconds the effect of restoring the original allocation duration for every SP allocation, due to the start of a new RIMP routine. In fact, the throughput value for the SPS scheme drops to the same value guaranteed by the other scheme, as the original allocation duration for the SPS is the same as the allocation duration for the DES. At this specific time, the reception at the PCP/AP of a RSNI report, with value below the *SINR threshold*, triggered the start of a new RIMP phase. This, in turn, was caused by the upcoming interference due to the movement of STAs G and H toward the directional communication between STAs E and F. Even if STAs A and B are located spatially apart from STAs E, F, G and H, once a single RSNI report does not meet the threshold requirement, a new RIMP routine embraces all stations participating in the

spatial sharing mechanism (in this case all STAs with an allocated SP).

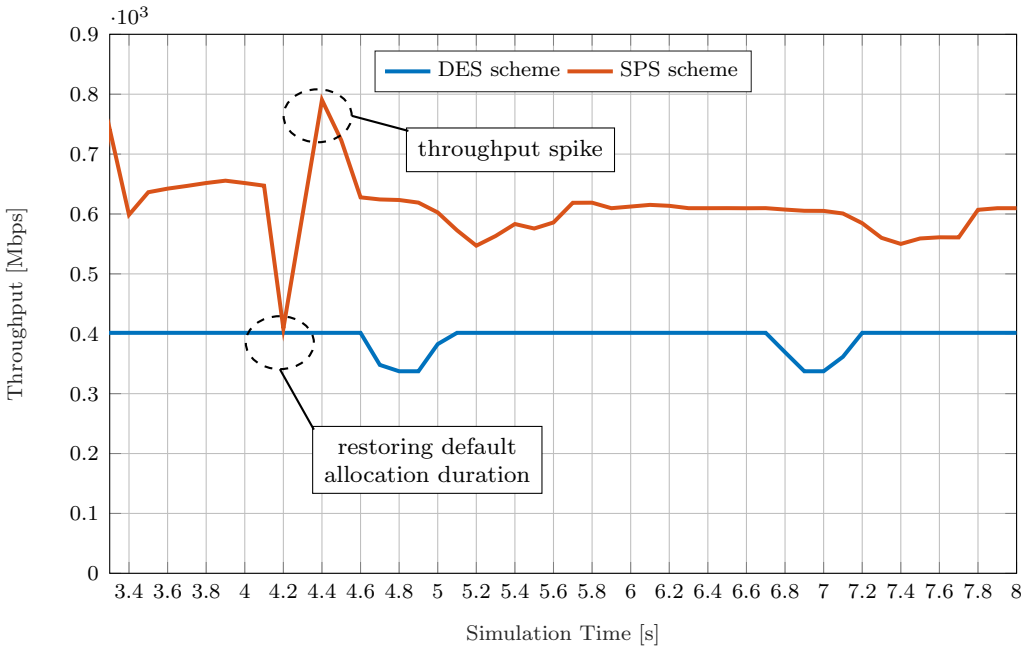

Fig. 4.3: Throughput vs simulation time for the allocation between STAs C and D: SP2

The same considerations hold for the SP allocation between STAs C and D (i.e. SP2), as shown in Fig. 4.3. From this figure, we can see that the throughput performance over time for the SPS scheme is higher when compared to the DES. Again, at 4.2 seconds after the simulation start, we can see the throughput decrease due to the initiation of a new RIMP. After this latter procedure is completed with the computation of a new IG, the new duration for each existing SP allocation is computed according to the new set of connected components. After this phase, the SPS scheme achieves again spatial sharing gain and so, it outperforms the other scheme. The presence of a spike in throughput after the new RIMP procedure is due to the buffering of packets during the new interference assessment procedure where the allocation duration is restored to its original and shorter value. The large amount of queued packets is released as soon as more channel time is reserved to the allocation thanks to spatial sharing, thus the spike of throughput. This same effect is visible as well for SP1 in Fig. 4.2.

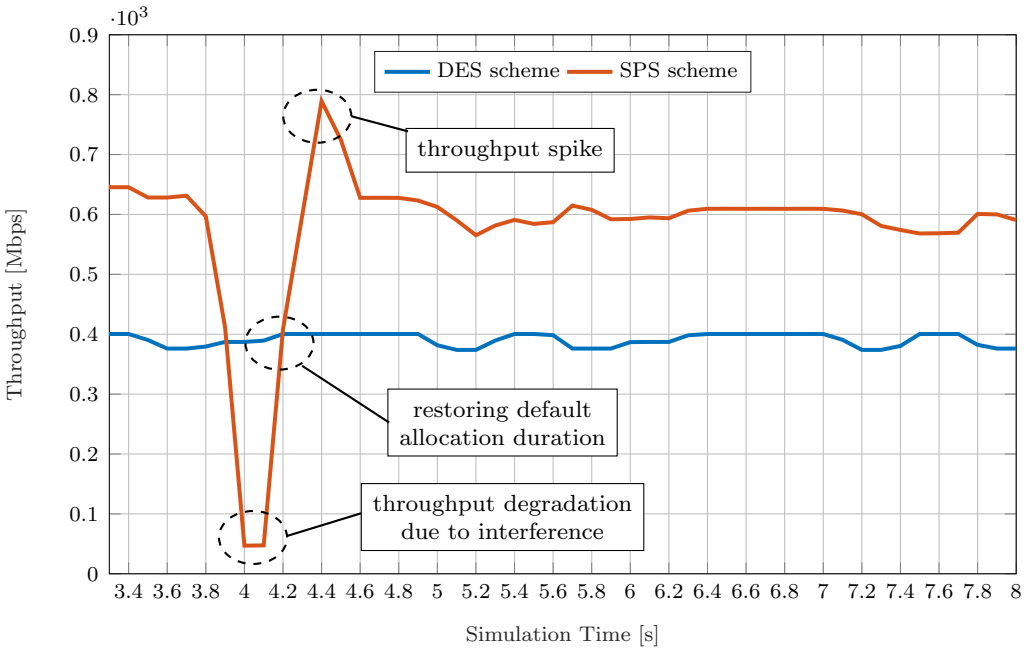

Fig. 4.4: Throughput vs simulation time for the allocation between STAs E and F: SP3

From Fig. 4.4, we can see the behavior of the two schedulers over time for the allocation between STAs E and F (i.e. SP3). For the SPS scheme, after the RIMP phase and the subsequent re-allocation of concurrent transmissions, the SPS scheme is able to guarantee higher throughput performance. When STAs G and H start moving towards STAs E and F at time 3.5 seconds, the throughput of this allocation starts decreasing until it reaches a very low value due to upcoming interference, indicating that those SPs between STAs G and H, and STAs E and F cannot be allocated concurrently anymore due to the fact that they now interfere with each other. This movement does not affect the DES scheme because each SP allocation has its own dedicated time inside the DTI, which does not overlap with any other allocated time. After detecting a SINR value under the predefined threshold thanks to the periodical RSNI reports, original allocations are restored, as we can see from the throughput value at 4.2 seconds. At this time instant, both schemes behave equally, thus they show the same throughput value. Successively, a new RIMP phase is initiated. After that, SP3 is scheduled concurrently according

to the new IG.

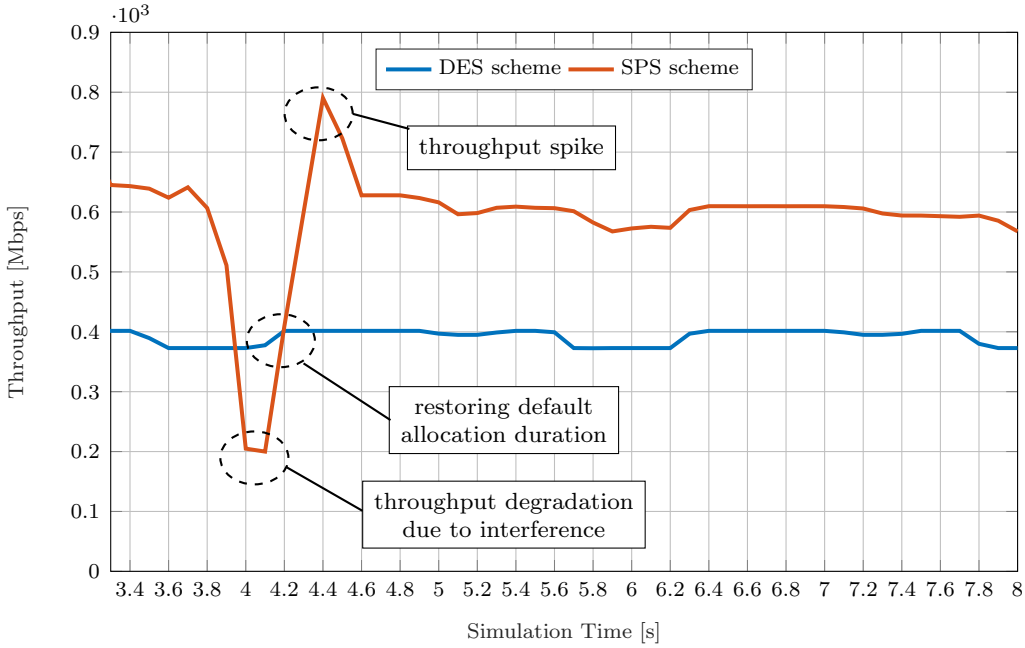

Fig. 4.5: Throughput vs simulation time for the allocation between STAs G and H: SP4

From Fig. 4.5, we can make the same considerations previously done also for the allocation between STAs G and H (i.e. SP4). However, in this case the throughput during the interfering phase is slightly higher with respect to SP3's case because STAs G and H are located closer than STAs E and F; therefore, STA H is able to retrieve more packets (i.e. the effect of interference is less pronounced compared to that experienced in SP3). Even in this case, we can see the effect of restoring the original allocation duration, which anticipates the start of a RIMP phase. In both Figs. 4.4 and 4.5, there is a spike of throughput after the end of the RIMP, which can be justify according to the same consideration offered before.

Fig. 4.6 depicts the average throughput over the entire simulation duration for each SP allocation. As we can see, the SPS scheme is able to guarantee better throughput performance compared to the DES, even on average. Considering the DES scheme, the average throughput is basically the same for every allocation, given that 4 pseudo-static SPs with the same

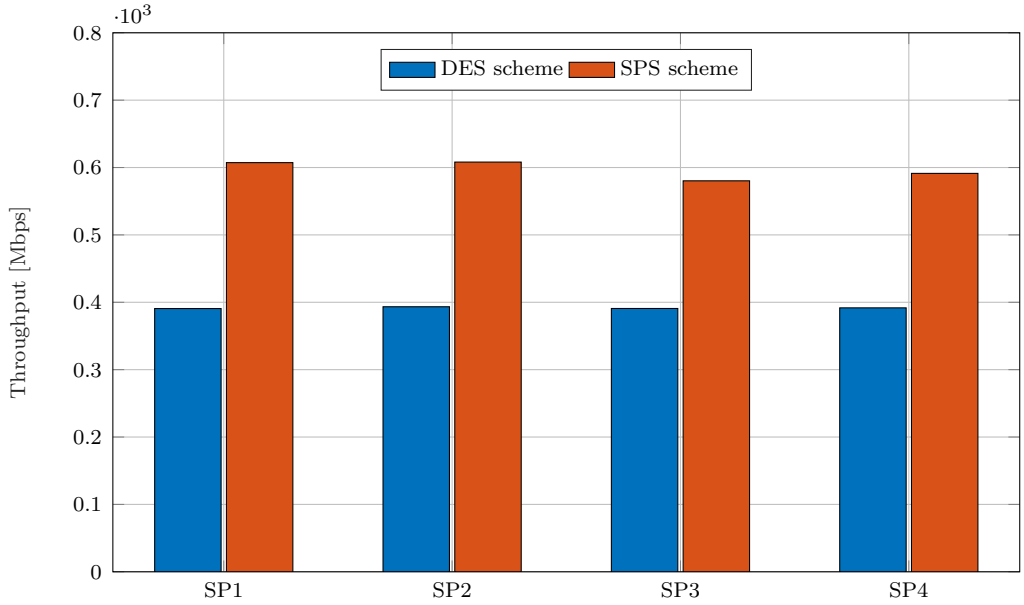

Fig. 4.6: Average throughput over the entire simulation duration for each allocated SP

duration are allocated during every DTI. Whereas, in the case of the SPS scheme, throughput for SPs 3 and 4 is, on average, slightly lower than the one for allocations: SP1 and SP2. This can be justified by the throughput degradation caused by the upcoming interference between the concurrently scheduled SP allocations (i.e. SP3 and SP4), when STAs G and H moves toward the line of communication between STAs E and F.

The graph of Fig. 4.7 depicts the end-to-end delay at application layer averaged over the simulation duration for each SP allocation. Clearly, the SPS algorithm outperforms the DES one even in terms of delay performance, thanks to the superior amount of time reserved for each allocation through the spatial sharing mechanism. On the other hand, the delay associated with the DES scheme is significantly high with respect to the other scheme. In fact, given the non-sufficient allocated time granted by the DES as a result of the simulation choices, many packets are buffered at MAC layer before being transmitted, thus increasing the end-to-end delay measured.

Finally, in this section we have seen that the SPS scheme guarantees higher throughput performance, hence spectral efficiency, thanks to an en-

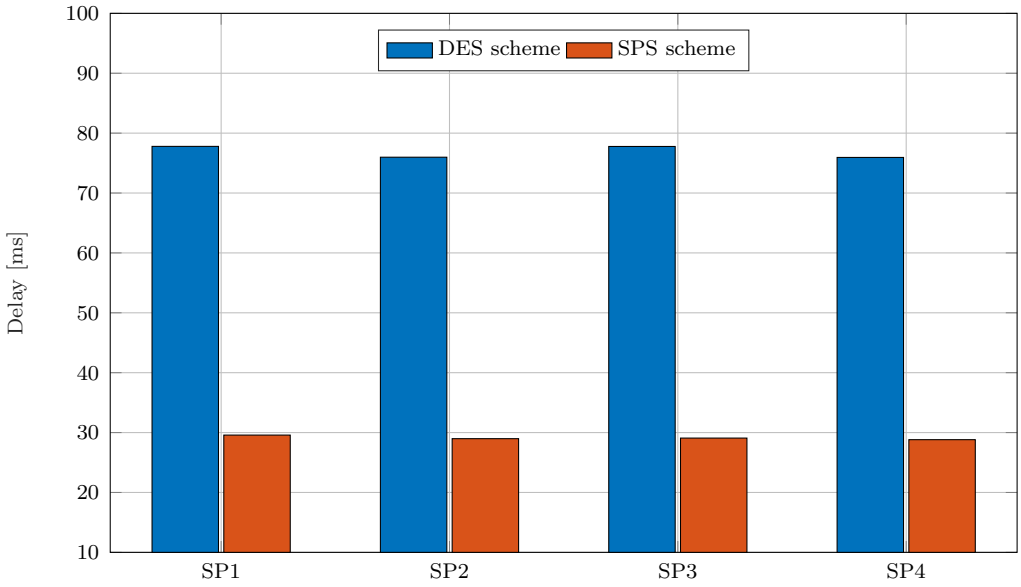

Fig. 4.7: Average delay over the entire simulation duration for each allocated SP

hanced implementation of the spatial sharing feature introduced by the IEEE 802.11ad standard. This mechanism builds an interference graph exploiting the exchange of Radio Measurement requests/reports between those STAs involved in allocated SPs and the PCP/AP. After the IG is built, existing SPs are allocated concurrently thanks to the computation of the IG's connected components. This permits to expand the contention-free allocated time, with clear benefits in terms of throughput and delay performance, as we previously demonstrated. However, there is an overhead associated with the RIMP and the RSNI report phases, which needs to be investigated in terms of co-existence with other sources of traffic in the network. Moreover, the performance of the SPS scheme is interference-limited: in the case of sudden changes in the network's interference, the algorithm takes time to detect those changes and react accordingly, as we have seen when considering the results of Figs. 4.4 and 4.5. This behavior might not be good in the case of delay-sensitive applications. In addition, the impact of the SPS's operations needs to be evaluated in the case of high mobility scenarios where changes in the interference footprint occur more frequently and thus the reaction time represents an important constraint to the overall performance.

4.2 RL-based Default vs Default

In this final section, we perform a comparison between the DES scheme and the RL-based DES. First, we introduce the IEEE office conference room scenario, then the results are presented and discussed.

4.2.1 Simulation scenario and settings

The ns-3 simulation scenario developed for this last comparison reproduces the IEEE office conference room scenario described in [22]. This scenario, whose topological plan is depicted in Fig. 4.8, comprises several different traffic types to mimic the real behavior of communications at 60 GHz inside a conference room.

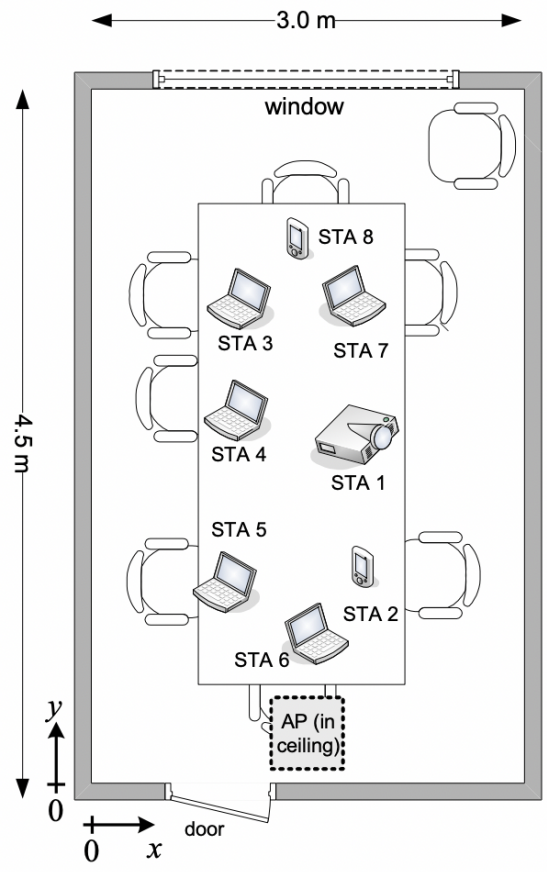

Fig. 4.8: IEEE office conference room scenario: floor plan

In particular, a laptop is transmitting Lightly Compressed Video (LCV) to a projector. Those laptops connected between each other are performing Local File Transfer (LFT). In addition, some laptops are transmitting or receiving a LFT from the PCP/AP, whereas, some other devices are producing web browsing traffic with the web content located in a remote server. The exact configuration for each device in the scenario is reported in Table 4.2 where the considered room dimension is in meters: $3.0 \times 4.5 \times 3.0$. In our ns-3 implementation we neglected the third dimension (i.e. height) since the available ns-3 channel/phy model is 2-dimensional.

Device	Type	2D location [m]	Transmitting traffic	Receiving traffic	Web browsing
PCP/AP	60 GHz access point	(1.50, 0.50)	> 1 LFT	> 1 LFT	N.A.
STA 1	Projector	(1.75, 2.30)	N.A.	LCV from STA 2	No
STA 2	Mobile device	(1.90, 1.50)	LCV to STA 1	LFT from PCP/AP	No
STA 3	Laptop	(1.35, 3.00)	LFT to STA 5	LFT from STA 5	Yes
STA 4	Laptop	(1.30, 2.40)	LFT to PCP/AP	N.A.	Yes
STA 5	Laptop	(1.25, 1.40)	LFT to STA 3	LFT from STA 3	Yes
STA 6	Laptop	(1.55, 1.20)	N.A.	N.A.	Yes
STA 7	Laptop	(1.85, 3.10)	LFT to STA 8	LFT from PCP/AP	No
STA 8	Mobile device	(1.60, 3.25)	N.A.	LFT from STA 7	No

Table 4.2: Office conference room scenario configuration. LCV stands for Lightly Compressed Video, LFT stands for Local File Transfer

Every station in this scenario remains at the indicated position for the entire simulation duration. The LCV stream has a target bitrate of 600 Mbps and the traffic model implementation guidelines are provided in [22]. Moreover, each LFT has been modeled according to a constant flow of UDP packets of payload size 1472 bytes. The application layer data rate for each LFT is uniformly chosen at random between the following values: $[50, 100, 150, 200]$ Mbps, at the start of simulation. Ultimately, the web browsing model implementation, based on HTTP protocol, has been carried according to the specifics contained in [22].

From the ns-3 simulation point of view, we allocated for each type of traffic, whether it is a LFT or the LCV stream, a dedicated SP inside the DTI following the rules defined by the DES algorithm. Given the configuration reported in Table 4.2, there are a total of 6 LFTs and 1 LCV stream, leading to a total of 7 SP allocations. Once the simulation starts, each transmitting

node of the pair sends to the PCP/AP an ADDTS request, where the parameters *Minimum Allocation* and *Maximum Allocation* duration are computed as described in section 3.2; then, the PCP/AP evaluates all received requests according to the DES policy. On the other hand, web browsing traffic is satisfied observing contention-based transmission rules, therefore the exchange of HTTP requests/responses occurs during a CBAP in the DTI.

When the RL-based DES scheme is exploited, the optimal duration for each allocated SP is learned by the RL agent, as per the operations illustrated in section 3.5. In particular, the agent needs an exploration period where, through random actions, it learns the optimal duration for each SP according to the MAC queue size. Once, the agent has had enough learning time, then it can start exploiting its knowledge. To effectively allow for agent's learning, we use 10 episodes (training episodes) of the ns-3 simulation where the agent explores the set of actions at its disposal and learns from the interaction with the environment. Then, during the last simulation episode (test episode) the agent performs actions according to the best policy founded during its training phase.

The main ns-3 parameters for the comparison between DES and RL-based DES schedulers are reported in Table 4.3.

Parameter	Value	Parameter	Value
802.11ad carrier frequency	60.48 GHz	802.11ad bandwidth	2.16 GHz
TX power P_{TX}	10 dBm	RX noise figure	10 dB
CCA threshold	-79 dBm	Energy detection threshold	-76 dBm
PHY MCS data mode	12	Beacon interval duration	41 TU
BTI duration	500 μs	Slots per A-BFT	8
Sector sweep frames per slot	8	ATI duration	300 μs
MAC layer queue size	5000 packets	MSDU aggregation size	7935 bytes
First CBAP duration	2500 μs	Video stream datarate	600 Mbps
App layer payload size	1472 bytes	App start time	3.0 s
Simulation duration	10 s	Core network latency	10 ms
Core network datarate	10 Gbps	Number queue states	50
Training episodes	10	Test episodes	1

Table 4.3: Ns-3 second simulation campaign settings

4.2.2 Simulation results

In the following, we analyze the results in terms of throughput, delay, Packet Loss Rate (PLR) and DTI occupancy for the DES vs the RL-based DES schemes. All the results reported have been averaged considering a simulation campaign of 40 independent runs over the described ns-3 simulation scenario. In the following figures, SP1 is associated with the LCV stream, while, SP2, SP3, SP4, SP5, SP6 and SP7 refer to LFTs.

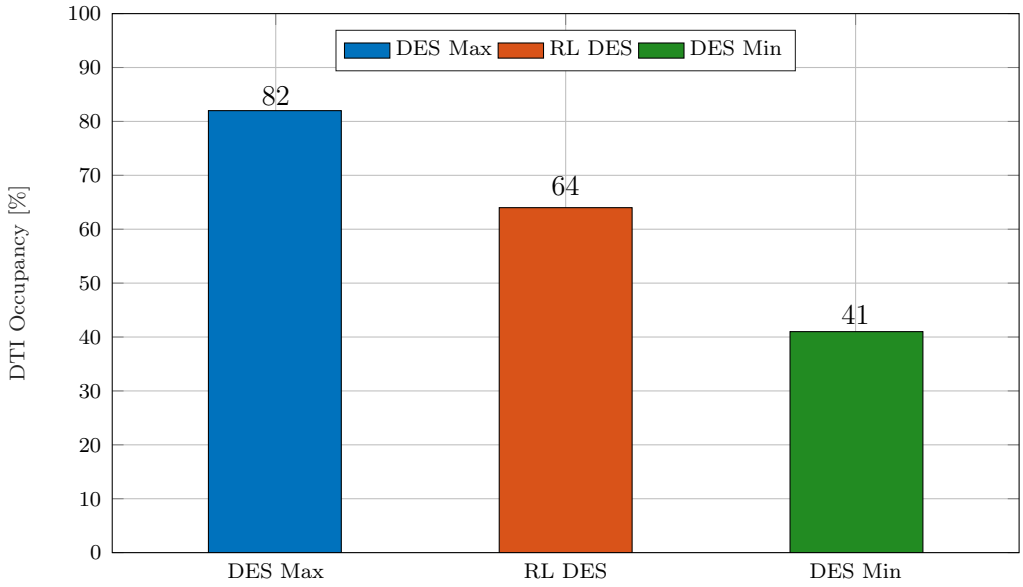

Fig. 4.9: Overall DTI occupancy for the allocated SPs, expressed as a percentage of the total DTI duration. For reference, a value of 100 would have indicated that the allocated SPs together occupy the entire DTI

Fig. 4.9 depicts the total DTI occupancy for all the allocated SPs according to different scheduling strategies. In particular, for this comparison, we consider the RL-based variation of the DES scheme and two characterizations of the DES policy: one where SPs are allocated according the *Maximum Allocation* parameter as usually done by this scheme (*DES Max* in figure), whereas, the other one where SPs' duration is forced to the value represented by the parameter *Minimum Allocation* (*DES Min*). As we can see, since the *Maximum Allocation* parameter is, by design, as twice as the *Minimum Allocation* value, then the total DTI occupancy for the SPs allocated according

to the DES Max scheme is twice as the one for the DES Min. In addition, the total duration of the SPs following the DES Max scheme is very close to the DTI duration. On the other hand, from the DTI occupancy associated with the RL-based DES policy, we can see that the allocated SPs occupy 18 % less DTI time with respect to the DES Max mechanism.

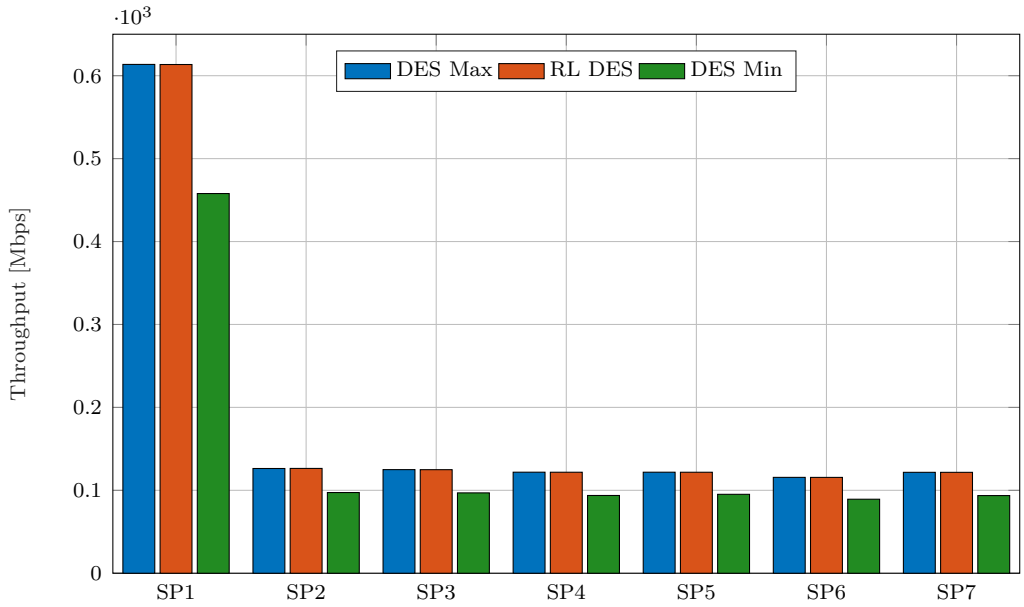

Fig. 4.10: Average throughput over the entire simulation duration for each allocated SP

In Fig. 4.10, it is shown the average end-to-end throughput for each SP allocation according to the three schemes just introduced. We can see that both DES Max and RL-based DES guarantee the same required throughput performance for every type of SP allocation. Contrarily, the DES Min provides lower throughput performance for every SP allocation involved. As a secondary comment, the average throughput between all the SPs allocated for a LFT is represented by nearly the same value because, recalling that the application data rate for a LFT is randomly chosen from a set of 4 values, then the average over 40 runs approaches the expectation of this random experiment.

From both Figs. 4.9 and 4.10, we can conclude that the RL-based DES algorithm is able to provide lower DTI occupancy while, at the same time,

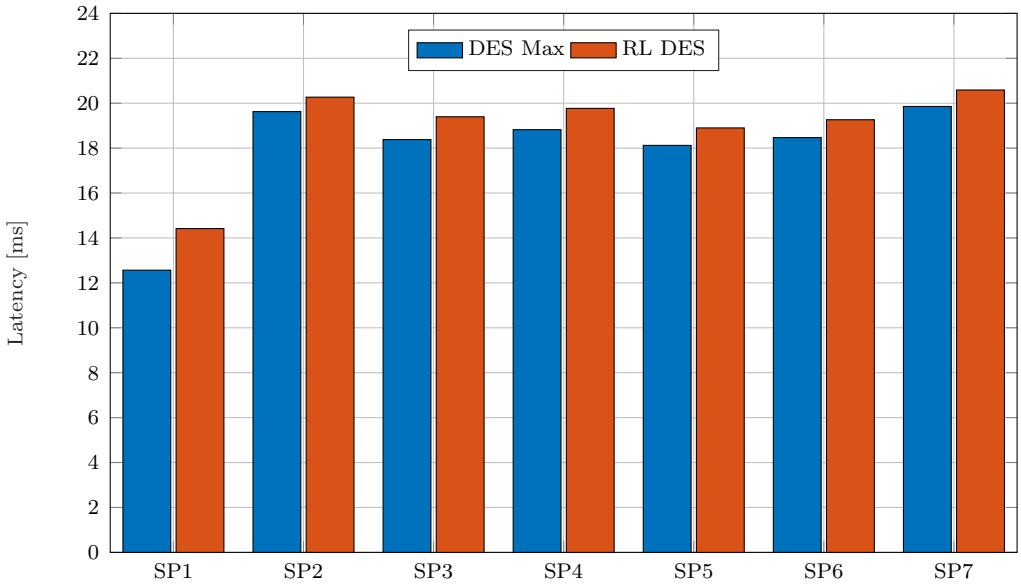

Fig. 4.11: Average end-to-end latency over the entire simulation duration for each allocated SP

preserve the required throughput performance. It follows that, thanks to the RL-based DES, there is more DTI time available to be granted for other sources/types of traffic. Instead, despite providing the lowest DTI occupancy, the DES Min variation does not meet the requirements in terms of throughput, as expected.

Fig. 4.11 shows the average end-to-end delay experienced by each SP allocation. In this figure, the delay associated with the DES Min scheme is not shown due to the tremendous performance degradation caused by the buffering of packets at MAC layer. This, in turn, can be related to the insufficient SP duration granted by the DES Min. As we can see, the RL-based DES scheme provides slightly worse performance in terms of delay. However, we can quantify the average loss to be less than 1 ms, therefore perfectly acceptable. This low degradation, in terms of delay performance, can be harmful in the case of delay-sensitive applications, such as video streaming. For the LCV stream of this scenario, the performance breakdown is less than 2 ms, as depicted in Fig. 4.11. In any case, when dealing with delay constrained traffic, one can always tune the BI duration in order to meet any requirement

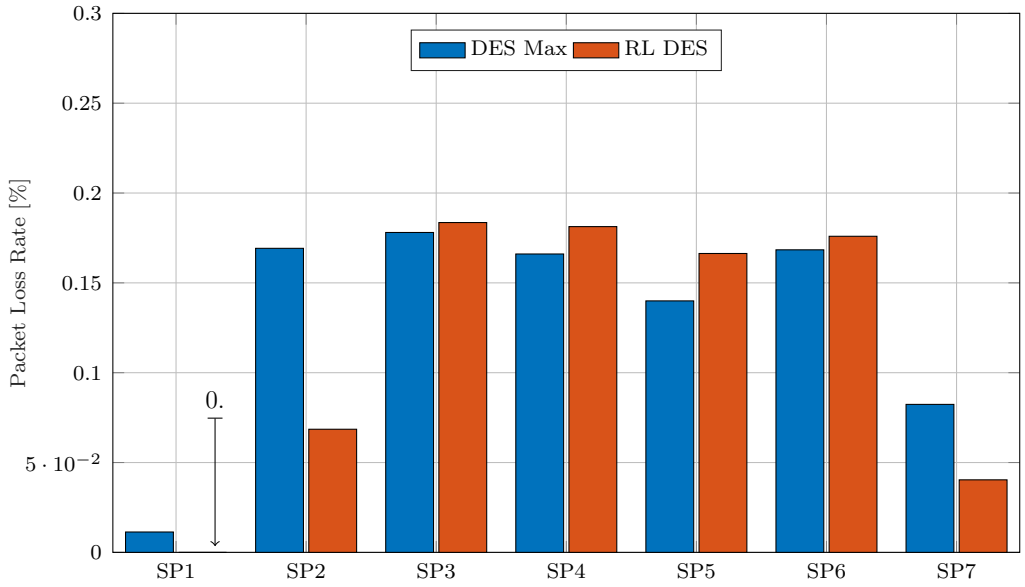

Fig. 4.12: Average PLR over the entire simulation duration for each allocated SP. The PLR is computed as percentage of the number of non-received packets out of the total number of transmitted packets

in terms of end-to-end latency. This performance degradation introduced by the RL-based DES scheme can be justified according to the fact that our RL agent is learning the optimal policy considering as reward the number of received packets. Therefore, it is not optimizing the SP duration for each allocation considering even the delay resulted from each action performed. In fact, as future work, we will consider the computation of the reward according to a formula that considers both metrics (i.e. number of received packets and delay performance), where each metric can be weighted according to the type of traffic served by the considered SP allocation.

In Fig. 4.12, the average end-to-end PLR for each SP allocation is depicted. Even for this evaluation, the results related to the DES Min scheme have been ignored due to the large performance degradation associated. In fact, many packets are not even received since they are discarded after spending too much time buffered at intermediate layers of the stack. We can see that both algorithms provide close performance in terms of PLR. Considering the y-axis scale of the figure, we can see how these results are close to

each other and, basically, they indicate that both approaches offer the same performance. In addition, for the LCV stream, we can see that the RL-based DES scheme is able to guarantee 0 PLR, which is an important result in the case of video streaming applications.

Ultimately, Fig. 4.13 represents a heatmap of the Q-table for SP1 (i.e. the SP associated with the LCV stream) at the end of the 10 training episodes. This Q-table represents the optimal policy our agent will employ during the exploitation phase, thus during the last episode (test episode).

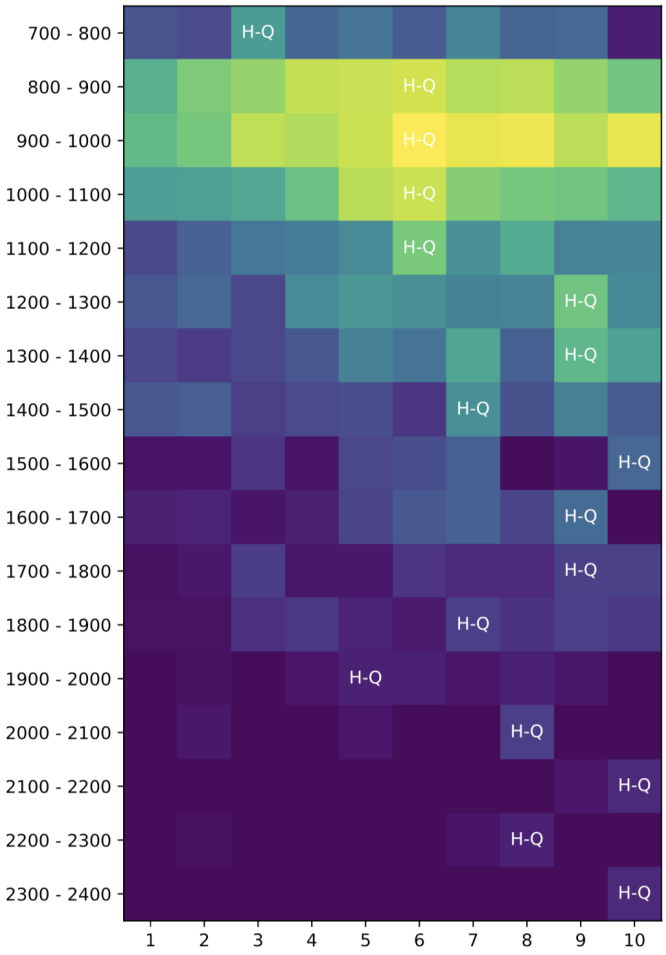

Fig. 4.13: Heatmap of the Q-values for the LCV's SP allocation. Each row represents a particular state (a queue size interval), whereas, each column represents an action. The H-Q symbol identifies the action with highest Q-value for that state

Recalling the queue size parameter of Table 4.3, the Q-table of Fig. 4.13 represents a reduced version of what the original one should be; this is due to the fact that many states are never visited during the learning process and, therefore, the associated Q-value for each action remains always equal to 0. As a matter of fact, this heatmap depicts the intensity of the Q-values for each state-action pair, for those states that are visited at least one time during the exploration phase of the RL agent. A row of the heatmap refers to a state representing a queue size interval, while, each column identifies an action from the actions' space. The optimal policy for the RL agent is to choose, at each particular state, the action that yields the maximum expected future reward; as a consequence, after the training phase, the agent selects at each step the action with highest Q-value. In turn, the symbol *H-Q* in the figure highlights the action to which correspond the highest Q-value for each state. From Fig. 4.13, we can see that, most of the time, the states observed by the RL agent lie between queue sizes 800 and 1200, meaning that these states are observed many times by our RL agent. Moreover, for this particular set of states the best action corresponds to action 6. This means the optimal SP1 duration is close to the mid value between the parameters *Minimum Allocation* and *Maximum Allocation*. Looking at the actions corresponding to the highest Q-value for states after queue size 1200, we can see that their associated SP duration is more close to the *Maximum Allocation* value than the *Minimum Allocation* one. In general, actions selected by the optimal policy are equal to or higher than 6, confirming that a value close to the *Minimum Allocation* does not provided the required performance in terms of throughput.

In conclusion, we saw that the RL-based DES algorithm is able to effectively learn the optimal duration for each allocated SP, which enables to preserve useful DTI time to be allocated for other traffic in the network while, at the same time, guarantees the expected throughput performance. This scheme introduces 1 ms of latency on average, due to the fact that it optimizes the policy according to the number of received packets as reward. However, this performance degradation is sweetened by a $\sim 20\%$ increase in DTI available time to satisfy new traffic and enhance network performance.

Chapter 5

Conclusions and Future Work

In this work, we addressed the lack of a resource scheduling algorithm for the hybrid MAC access of IEEE 802.11ad by developing different schemes characterized by a their own behavior and complexity. The implementation has been carried exploiting the complete and powerful ns-3 network simulator which provides robust and meaningful results related to the features implemented. We developed and assessed the performance of three scheduler implementations.

The first scheduling algorithm (DES scheme) has been used as a baseline for the comparison with more complex and advanced approaches. The SPS algorithm (our second approach) exploits a mapping between interference among different SP allocations through the IG to achieve the concurrent transmission of spatially separated links in the network. In fact, this scheme enhances and improves the spatial sharing mechanism introduced by the standard. Simulation results proved that the SPS scheme is able to guarantee better throughput performance and spectral efficiency thanks to the transmission of multiple flows at the same time, compared to the firstly developed DES scheme. However, in order to achieve and sustain spatial sharing, an overhead is associated with the exchange of radio measurements frames between STAs and the PCP/AP. Moreover, the scheduling algorithm itself needs a small time to react to sudden changes in the network's interference, however, this impact must be evaluated in scenarios characterized by higher

mobility than the one used in our evaluation.

In our last approach, we used RL to learn the optimal duration for each allocated SP. In fact, one of the major drawbacks connected to the DES scheme was associated to a waste of DTI time when each SP allocation is granted according to the *Maximum Allocation* parameter. Through results obtained over a realistic IEEE scenario, we showed that the RL-based DES algorithm is able to guarantee an inferior DTI occupancy for the allocated SPs with respect to the DES scheme, while at the same time, can guarantee the same throughput and close performance in terms of delay. Thus, this RL-based scheme enables to satisfy more diverse traffic than its non-RL counterpart.

Future work can be devoted into different directions according to the type of scheduling policy. Concerning the SPS scheme, it will be important to characterize the overhead associated with the measurement phase for constructing and updating the IG. In addition, the performance of this scheme should be evaluated in the case of high interference and/or mobility scenarios.

Considering the RL-based strategy, it would be interesting to introduce a new function which computes the reward associated to a state-action pair considering the number of received packets and the delay associated. Thanks to this method, the RL agent can learn how to optimize each SP duration according to the metric of interest for that specific allocation. For example, if an application is delay sensitive, we can optimize considering more the reward associated to actions that yield lower delay; whereas, if throughput performance is the main metric to consider, we can optimize by giving more importance to the number of packets received for the computation of the reward.

Bibliography

[1] "IEEE 802.11ad, Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band," *IEEE 802.11 Working Group*, 2012.

[2] T. Nitsche, C. Cordeiro, A. Flores, E. W. Knightly, E. Perahia, and J. C. Widmer, "IEEE 802.11ad: Directional 60 GHz Communication for Multi-Gbps Wi-Fi," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, 2014.

[3] G. F. Riley and T. R. Henderson, "The ns-3 Network Simulator," in *Modeling and Tools for Network Simulation*, pp. 15–34, Springer, 2010.

[4] B. Schulz, "802.11ad - WLAN at 60 GHz: A Technology Introduction," *White Paper, Rohde & Schwarz*, 2017.

[5] Y. Ghasempour, C. M. da Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-Generation 60 GHz Communication for 100 Gb/s Wi-Fi," *IEEE Communications Magazine*, vol. 55, pp. 186–192, 2017.

[6] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement Learning in Robotics: A Survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[7] R. S. Sutton and A. G. Barto, "Introduction to Reinforcement Learning," *MIT Press*, 1998.

[8] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[9] L. Zhou and Y. Ohashi, "Efficient Codebook-based MIMO Beamforming for Millimeter-wave WLANs," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, pp. 1885–1889, 2012.

[10] W. Wu, Q. Shen, M. Wang, and X. S. Shen, "Performance Analysis of IEEE 802.11ad Downlink Hybrid Beamforming," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2017.

[11] H. Assasa, S. K. Saha, A. Loch, D. Koutsonikolas, and J. Widmer, "Medium Access and Transport Protocol Aspects in Practical 802.11 ad Networks," in *2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks"*, pp. 1–11, 2018.

[12] C. Hemanth and T. G. Venkatesh, "Performance Analysis of Service Periods (SP) of the IEEE 802.11ad Hybrid MAC Protocol," *IEEE Transactions on Mobile Computing*, vol. 15, no. 5, pp. 1224–1236, 2016.

[13] K. Chandra, R. V. Prasad, and I. Niemegeers, "Performance Analysis of IEEE 802.11ad MAC Protocol," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1513–1516, 2017.

[14] S. Chinchali, P. Hu, T. Chu, M. Sharma, M. Bansal, R. Misra, M. Pavone, and S. Katti, "Cellular Network Traffic Scheduling With Deep Reinforcement Learning," in *AAAI*, 2018.

[15] W. Li, F. Zhou, K. Chowdhury, and W. Meleis, "QTCP: Adaptive Congestion Control with Reinforcement Learning," *IEEE Transactions on Network Science and Engineering*, 2018.

[16] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "OpenAI Gym," *CoRR*, 2016.

[17] H. Assasa and J. Widmer, "Implementation and Evaluation of a WLAN IEEE 802.11ad Model in ns-3," in *Proceedings of the Workshop on Ns-3*, WNS3 '16, pp. 57–64, ACM, 2016.

[18] H. Assasa and J. Widmer, "Extending the IEEE 802.11Ad Model: Scheduled Access, Spatial Reuse, Clustering and Relaying," in *Proceedings of the Workshop on Ns-3*, WNS3 '17, pp. 39–46, ACM, 2017.

[19] H. Assasa, "GitHub Repository for the IEEE 802.11ad Model." `https://github.com/wigig-tools/ns3-802.11ad`, 2017. [Online; accessed 15-Nov-2018].

[20] Project-NS-3, "The ns-3 Wi-Fi Module Documentation," 2016.

[21] P. Gawlowicz and A. Zubow, "ns3-gym: Extending OpenAI Gym for Networking Research," *CoRR*, 2018.

[22] E. Perahia, "TGad Evaluation Methodology." `http://www.ieee802.org/11/Reports/tgad_update.htm`, 2009. [Online; accessed 16-Dec-2018].

Acronyms and Abbreviations

A-BFT Association Beamforming Training.

AC Access Category.

ADDTS Add Traffic Stream.

AID Associaton Identifier.

ANIPI Average Noise plus Interference Power Indicator.

AR Augmented Reality.

ASYN Asynchronous.

ATI Announcement Transmission Interval.

BER Bit Error Rate.

BF Beamforming.

BHI Beacon Header Interval.

BI Beacon Interval.

BRP Beam Refinement Protocol.

BTI Beacon Transmission Interval.

CBAP Contention-Based Access Period.

CSMA/CA Carrier Sense Multiple Access with Collision Avoidance.

CTS Clear To Send.

DCA Distributed Channel Access.

DCF Distributed Coordination Function.

DES Default Scheduler.

DFT Discrete Fourier Transform.

DMG Directional Multi-Gigabit.

DTI Data Transmission Interval.

E2E End-to-End.

EDCA Enhanced Distributed Channel Access.

ESE Extended Schedule Element.

IG Interference Graph.

ISM lndustrial, Scientific and Medical.

ISOC Isochronous.

LCV Lightly Compressed Video.

LFT Local File Transfer.

LoS Line of Sight.

MAC Medium Access Control.

MCS Modulation and Coding Scheme.

MDP Markov Decision Process.

ML Machine Learning.

NAV Network Allocation Vector.

NLoS Non Line of Sight.

ns-3 network simulator 3.

OFDM Orthogonal Frequency-Division Multiplexing.

PCP/AP Personal Basic Service Set (PBSS) Control Point/Access Point.

PER Packet Error Rate.

PHY Physical.

PLR Packet Loss Rate.

QoS Quality of Service.

RIMP Report Interference Measure Procedure.

RL Reinforcement Learning.

RSNI Received Signal-to-Noise Indicator.

RTS Request To Send.

RXSS Receive Sector Sweep.

SARSA State-Action-Reward-State-Action.

SC Single Carrier.

SINR Signal to Interference plus Noise Ratio.

SLS Sector-Level Sweep.

SNR Signal to Noise Ratio.

SP Service Period.

SPR Service Period Request.

SPS Spatial Sharing Scheduler.

SPSH Spatial Sharing.

SSW Sector Sweep.

STA Station.

TDMA Time Division Multiple Access.

TS Traffic Stream.

TU Time Unit.

TXOP Transmission Opportunity.

TXSS Transmit Sector Sweep.

UDP User Datagram Protocol.

UP User Priority.

VR Virtual Reality.

WLAN Wireless Local Area Network.

Acknowledgments

Innanzitutto, ci tengo a ringraziare i miei genitori per avermi sempre sostenuto durante il mio intero percorso accademico e, per avermi concesso l'opportunità di inseguire i miei sogni in qualsiasi momento.

Doveroso è il ringraziamento nei confronti del prof. Michele Zorzi, senza il quale, il mio periodo di ricerca negli Stati Uniti d'America non sarebbe stato realizzabile. Di conseguenza, ringrazio il capo della divisione di ricerca "Wireless Network Division" presso il NIST, Nada Golmie, per avermi concesso questa opportunità. Inoltre, ringrazio Tanguy Ropitault per il suo supporto tecnico durante il mio lavoro al NIST. Un grande abbraccio a tutte le persone conosciute negli States.

Ci tengo a ringraziare tutti i compagni di corso conosciuti durante questi anni all'università, perchè ognuno di loro ha contribuito alla mia crescita come studente e persona. In particolare, voglio ringraziare Matteo Drago, compagno di mille avventure e progetti, con il quale ho condiviso molte gioie, ma anche difficoltà, incontrate durante il mio percorso di studi.

Ringrazio il mio amico, Clemente Elia, per aver sempre creduto nelle mie capacità come studente in ingegneria.

Infine, ringazio Valeria Viganò perchè è la prima persona pronta a sostenermi in qualsiasi situazione e la prima persona su cui posso in contare in qualsiasi momento.